



**TECHNISCHE
UNIVERSITÄT
DRESDEN**

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

LTCS–Report

Adding Threshold Concepts to the Description Logic \mathcal{EL}

Franz Baader Gerhard Brewka Oliver Fernández Gil

LTCS-Report 15-09

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

Adding Threshold Concepts to the Description Logic \mathcal{EL}

Franz Baader¹, Gerhard Brewka², and Oliver Fernández Gil^{*2}

¹Theoretical Computer Science, TU Dresden, Germany

²Department of Computer Science, University of Leipzig, Germany

Abstract

We introduce an extension of the lightweight Description Logic \mathcal{EL} that allows us to define concepts in an approximate way. For this purpose, we use a graded membership function, which for each individual and concept yields a number in the interval $[0, 1]$ expressing the degree to which the individual belongs to the concept. Threshold concepts $C_{\sim t}$ for $\sim \in \{<, \leq, >, \geq\}$ then collect all the individuals that belong to C with degree $\sim t$. We generalize a well-known characterization of membership in \mathcal{EL} concepts to construct a specific graded membership function deg , and investigate the complexity of reasoning in the Description Logic $\tau\mathcal{EL}(deg)$, which extends \mathcal{EL} by threshold concepts defined using deg . We also compare the instance problem for threshold concepts of the form $C_{>t}$ in $\tau\mathcal{EL}(deg)$ with the relaxed instance queries of Ecke et al.

*Supported by DFG Graduiertenkolleg 1763 (QuantLA).

Contents

1	Introduction	3
2	The Description Logic \mathcal{EL}	6
3	The Logic $\tau\mathcal{EL}(m)$	10
3.1	Description graphs and homomorphisms in $\tau\mathcal{EL}(m)$	11
3.2	Deciding the existence of a τ -homomorphism	15
4	The membership function deg	18
4.1	Some properties of $\tau\mathcal{EL}(deg)$	24
5	Reasoning	26
5.1	Terminological reasoning	26
5.2	Assertional reasoning	33
6	Concept similarity and relaxed instance queries	42
6.1	Defining membership degree functions	42
6.2	Relation to the membership degree function deg	45
7	Conclusion	52
8	Appendix	55

1 Introduction

Description logics (DLs) [2] are a family of logic-based knowledge representation formalisms, which can be used to represent the conceptual knowledge of an application domain in a structured and formally well-understood way. They allow their users to define the important notions of the domain as concepts by stating necessary and sufficient conditions for an individual to belong to the concept. These conditions can be atomic properties required for the individual (expressed by concept names) or properties that refer to relationships with other individuals and their properties (expressed as role restrictions). The expressivity of a particular DL is determined by what sort of properties can be required and how they can be combined.

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} , even with respect to expressive terminological axioms [7]. On the other hand, though quite inexpressive, \mathcal{EL} can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT.¹ In \mathcal{EL} we can, for example, define the concept of a *happy man* as a male human that is healthy and handsome, has a rich and intelligent wife, a son and a daughter, and a friend:

$$\begin{aligned} & \text{Human} \sqcap \text{Male} \sqcap \text{Healthy} \sqcap \text{Handsome} \sqcap \\ & \exists \text{spouse} . (\text{Rich} \sqcap \text{Intelligent} \sqcap \text{Female}) \sqcap \\ & \exists \text{child} . \text{Male} \sqcap \exists \text{child} . \text{Female} \sqcap \exists \text{friend} . \top \end{aligned} \tag{1}$$

For an individual to belong to this concept, all the stated properties need to be satisfied. However, maybe we would still want to call a man happy if most, though not all, of the properties hold. It might be sufficient to have just a daughter without a son, or a wife that is only intelligent but not rich, or maybe an intelligent and rich spouse of a different gender. But still, not too many of the properties should be violated.

In this paper, we introduce a DL extending \mathcal{EL} that allows us to define concepts in such an approximate way. The main idea is to use a *graded membership function*, which instead of a Boolean membership value 0 or 1 yields a membership degree from the interval $[0, 1]$. We can then require a happy man to belong to the \mathcal{EL} concept (1) with degree at least .8. More generally, if C is an \mathcal{EL} concept, then the *threshold concept* $C_{\geq t}$ for $t \in [0, 1]$ collects all the individuals that belong to C with degree at least t . In addition to such upper threshold concepts, we will also consider lower threshold concepts $C_{\leq t}$ and allow the use of strict inequalities in both. For example, an unhappy man could be required to belong to the \mathcal{EL}

¹see <http://www.ihtsdo.org/snomed-ct/>

concept (1) with a degree less than .2.

The use of membership degree functions with values in the interval $[0, 1]$ may remind the reader of fuzzy logics. However, there is no strong relationship between this work and the work on fuzzy DLs [6] for two reasons. First, in fuzzy DLs the semantics is extended to fuzzy interpretations where concept and role names are interpreted as fuzzy sets and relations, respectively. The membership degree of an individual to belong to a complex concept is then computed using fuzzy interpretations of the concept constructors (e.g., conjunction is interpreted using an appropriate triangular norm). In our setting, we consider crisp interpretations of concept and role names, and directly define membership degrees for complex concepts based on them. Second, we use membership degrees to obtain new concept constructors, but the threshold concepts obtained by applying these constructors are again crisp rather than fuzzy.

In the next section, we will formally introduce the DL \mathcal{EL} , and then recall the well-known characterization of element-hood in \mathcal{EL} concepts via existence of homomorphisms between \mathcal{EL} description graphs (which can express both \mathcal{EL} concepts and interpretations in a graphical way). In Section 3, we then extend \mathcal{EL} by new threshold concept constructors, which are based on an arbitrary, but fixed graded membership function. We will impose some minimal requirements on such membership functions, and show the consequences that these conditions have for our threshold logic. In Section 4, we then introduce a specific graded membership function, which satisfies the requirements from the previous sections. Its definition is a natural extension of the homomorphism characterization of crisp membership in \mathcal{EL} . Basically, an individual is punished (in the sense that its membership degree is lowered) for each missing property in a uniform way. More sophisticated versions of this function, which weigh the absence of different properties in a different way, may be useful in practice. However, they are easy to define and considering them would only add clutter, but no new insights, to our investigation (in Section 5) of the computational properties of the threshold logic obtained by using this function.

In Section 6 we compare our graded membership function with similarity measures on \mathcal{EL} concepts. In fact, from a technical point of view, the graded membership function introduced in Section 4 is akin to the similarity measures for \mathcal{EL} concepts introduced in [16, 17], though only [17] directly draws its inspirations from the homomorphism characterization of subsumption in \mathcal{EL} . We show that a variant of the relaxed instance query approach of [9] can be used to turn a similarity measure into a graded membership function. It turns out that, applied to a simple instance \bowtie^1 of the framework for constructing similarity measures in [16], this approach actually yields our membership function *deg*. In addition, we can show that the relaxed instance queries of [16] can be expressed as instance queries w.r.t. threshold concepts of the form $C_{>t}$. However, the new DL introduced in this paper is considerably more expressive than just such threshold concepts since

we also allow the use of comparison operators other than $>$ in threshold concepts, and the threshold concepts can be embedded in complex \mathcal{EL} concepts.

2 The Description Logic \mathcal{EL}

We start by introducing the Description Logic \mathcal{EL} . Starting with finite sets of concept names \mathbf{N}_C and role names \mathbf{N}_R , the set $\mathcal{C}_{\mathcal{EL}}$ of \mathcal{EL} concept descriptions is obtained by using the concept constructors *conjunction* ($C \sqcap D$), *existential restriction* ($\exists r.C$) and *top* (\top), in the following way:

$$C ::= \top \mid A \mid C \sqcap C \mid \exists r.C$$

where $A \in \mathbf{N}_C$, $r \in \mathbf{N}_R$ and $C \in \mathcal{C}_{\mathcal{EL}}$.

We denote the set of all *sub-descriptions* of the concept description C as $\text{sub}(C)$. In addition, the *role-depth* $\text{rd}(C)$ of C is inductively defined as follows:

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

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 assigns subsets of $\Delta^{\mathcal{I}}$ to each concept name and binary relations over $\Delta^{\mathcal{I}}$ to each role name. The interpretation function $\cdot^{\mathcal{I}}$ is inductively extended to concept descriptions in the usual way:

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

Given $C, D \in \mathcal{C}_{\mathcal{EL}}$, we say that C is *subsumed by* D (denoted as $C \sqsubseteq D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for every interpretation \mathcal{I} . These two concept descriptions are *equivalent* (denoted as $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$. Finally, C is *satisfiable* iff $C^{\mathcal{I}} \neq \emptyset$ for some interpretation \mathcal{I} .

Given two interpretations \mathcal{I} and \mathcal{J} , we say that \mathcal{I} is *contained in* \mathcal{J} (denoted $\mathcal{I} \subseteq \mathcal{J}$) iff $\Delta^{\mathcal{I}} \subseteq \Delta^{\mathcal{J}}$ and $X^{\mathcal{I}} \subseteq X^{\mathcal{J}}$ for all $X \in (\mathbf{N}_C \cup \mathbf{N}_R)$.

Information about specific individuals can be expressed in an ABox. An *ABox* \mathcal{A} is a finite set of *assertions* of the form $C(a)$ or $r(a, b)$, where C is an \mathcal{EL} concept description, $r \in \mathbf{N}_R$, and a, b are individual names. In addition to concept and role names, an interpretation \mathcal{I} now assigns domain elements $a^{\mathcal{I}}$ to individual names a . The assertion $C(a)$ is satisfied by \mathcal{I} iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$, and $r(a, b)$ is satisfied by \mathcal{I} iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$. The interpretation \mathcal{I} is a *model* of \mathcal{A} iff \mathcal{I} satisfies all assertion in \mathcal{A} . The ABox \mathcal{A} is *consistent* iff it has a model, and the individual a is an *instance* of the concept C in \mathcal{A} iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ holds in all models of \mathcal{A} . We denote the set of individual names occurring in \mathcal{A} as $\text{Ind}(\mathcal{A})$.

Our definition of graded membership will be based on graphical representations of concepts and interpretations, and on homomorphisms between such representa-

tions. For this reason, we recall these notions together with the pertinent results. They are all taken from [4, 14, 1].

Definition 1 (\mathcal{EL} description graphs). An \mathcal{EL} description graph is a graph of the form $G = (V_G, E_G, \ell_G)$ where:

- V_G is a set of nodes.
- $E_G \subseteq V_G \times \mathbf{N}_R \times V_G$ is a set of edges labelled by role names,
- $\ell_G : V_G \rightarrow 2^{\mathbf{N}_c}$ is a function that labels nodes with sets of concept names.

The empty label corresponds to the *top-concept*. In particular, an \mathcal{EL} description tree T is a description graph that is a tree with a distinguished element v_0 representing its root. In [4], it was shown the correspondence that exists between \mathcal{EL} concept descriptions and \mathcal{EL} description trees, i.e., every \mathcal{EL} concept description C can be translated into a corresponding description tree T_C and vice versa. Furthermore, every interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ can be translated into an \mathcal{EL} description graph $G_{\mathcal{I}} = (V_{\mathcal{I}}, E_{\mathcal{I}}, \ell_{\mathcal{I}})$ in the following way [1]:

- $V_{\mathcal{I}} = \Delta^{\mathcal{I}}$,
- $E_{\mathcal{I}} = \{(vrw) \mid (v, w) \in r^{\mathcal{I}}\}$,
- $\ell_{\mathcal{I}}(v) = \{A \mid v \in A^{\mathcal{I}}\}$ for all $v \in V_{\mathcal{I}}$.

The following example illustrates the relation between concept descriptions and description trees, and interpretations and description graphs.

Example 2. The \mathcal{EL} concept description

$$C := A \sqcap \exists r.(A \sqcap B \sqcap \exists r.\top) \sqcap \exists r.A$$

yields the \mathcal{EL} description tree T_C depicted on the left-hand side in Figure 1. The description graph on the right-hand side corresponds to the following interpretation:

- $\Delta^{\mathcal{I}} := \{a_1, a_2, a_3\}$,
- $A^{\mathcal{I}} := \{a_1, a_2\}$ and $B^{\mathcal{I}} := \{a_2, a_3\}$,
- $r^{\mathcal{I}} := \{(a_1, a_2), (a_2, a_3), (a_3, a_1)\}$.

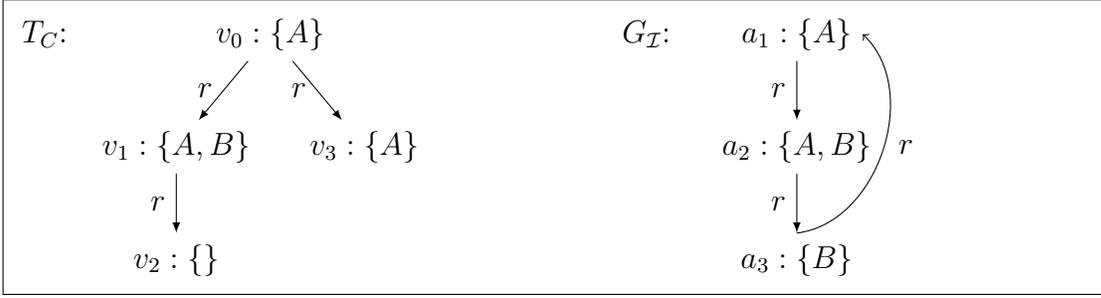


Figure 1: \mathcal{EL} -description graphs.

Next, we generalize homomorphisms between \mathcal{EL} description trees [4] to arbitrary graphs.

Definition 3 (Homomorphisms on \mathcal{EL} description graphs). Let $G = (V_G, E_G, \ell_G)$ and $H = (V_H, E_H, \ell_H)$ be two \mathcal{EL} description graphs. A mapping $\varphi : V_G \rightarrow V_H$ is a *homomorphism* from G to H iff the following conditions are satisfied:

1. $\ell_G(v) \subseteq \ell_H(\varphi(v))$ for all $v \in V_G$, and
2. $vrw \in E_G$ implies $\varphi(v)r\varphi(w) \in E_H$.

This homomorphism is an *isomorphism* iff it is bijective, equality instead of just inclusion holds in 1., and biimplication instead of just implication holds in 2.

In Example 2, the mapping φ with $\varphi(v_i) = a_{i+1}$ for $i = 0, 1, 2$ and $\varphi(v_3) = a_2$ is a homomorphism. Homomorphisms between \mathcal{EL} description trees can be used to characterize subsumption in \mathcal{EL} .

Theorem 4 ([4]). *Let C, D be \mathcal{EL} concept descriptions and T_C, T_D the corresponding \mathcal{EL} description trees. Then $C \sqsubseteq D$ iff there exists a homomorphism from T_D to T_C that maps the root of T_D to the root of T_C .*

The proof of this result can be easily adapted to obtain a similar characterization of *element-hood* in \mathcal{EL} , i.e., whether $d \in C^{\mathcal{I}}$ for some $d \in \Delta^{\mathcal{I}}$.

Theorem 5. *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 $\varphi(v_0) = d$.*

In Example 2, the existence of the homomorphism φ defined above thus shows that $a_1 \in C^{\mathcal{I}}$. Equivalence of \mathcal{EL} concept descriptions can be characterized via the existence of isomorphisms, but for this the concept descriptions first need to be normalized by removing redundant existential restrictions. To be more precise, the *reduced form* of an \mathcal{EL} concept description is obtained by applying the

rewrite rule $\exists r.C \sqcap \exists r.D \longrightarrow \exists r.C$ if $C \sqsubseteq D$ as long as possible. This rule is applied modulo associativity and commutativity of \sqcap , and not only on the top-level conjunction of the description, but also under the scope of existential restrictions. Since every application of the rule decreases the size of the description, it is easy to see that the reduced form can be computed in polynomial time. We say that an \mathcal{EL} concept description is *reduced* iff this rule does not apply to it. In our Example 2, the reduced form of C is the reduced description $A \sqcap \exists r.(A \sqcap B \sqcap \exists r.\top)$.

Theorem 6 ([14]). *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} .*

3 The Logic $\tau\mathcal{EL}(m)$

Our new logic will allow us to take an arbitrary \mathcal{EL} concept C and turn it into a threshold concept. To this end we introduce a family of constructors that are based on the membership degree of individuals in C . For instance, the threshold concept $C_{>.8}$ represents the individuals that belong to C with degree $>.8$. The semantics of the new threshold concepts depends on a (graded) membership function m . Given an interpretation \mathcal{I} , this function takes a domain element $d \in \Delta^{\mathcal{I}}$ and an \mathcal{EL} concept C as input, and returns a value between 0 and 1, representing the extent to which d belongs to C in \mathcal{I} .

The choice of an appropriate membership function m is obviously crucial. In Section 4 we will propose one specific such function deg , but we do not claim this is the only reasonable way to define such a function. Rather, the membership function is a parameter in defining the logic. To highlight this dependency, we call the logic $\tau\mathcal{EL}(m)$.

Nevertheless, membership functions are not arbitrary. There are two properties we require such functions to satisfy:

Definition 7. A graded membership function m is a family of functions that contains for every interpretation \mathcal{I} a function $m^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \rightarrow [0, 1]$ satisfying the following conditions (for $C, D \in \mathcal{C}_{\mathcal{EL}}$):

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

Property $M1$ requires that the value 1 is a distinguished value reserved for proper containment in a concept. Property $M2$ requires equivalence invariance. It expresses the intuition that the membership value should not depend on the syntactic form of a concept, but only on its semantics. Note that the right to left implication in $M2$ is already a consequence of $M1$: if $C \not\equiv D$, this would imply that for some interpretation \mathcal{I} and $d \in \Delta^{\mathcal{I}}$, $d \in C^{\mathcal{I}}$ and $d \notin D^{\mathcal{I}}$ (or the opposite). Then, by $M1$ we would have $m^{\mathcal{I}}(d, C) = 1$ and consequently $m^{\mathcal{I}}(d, D) = 1$, which is a contradiction with $d \notin D^{\mathcal{I}}$ and $M1$.

We now turn to the syntax of $\tau\mathcal{EL}(m)$. Given finite sets of concept names $\mathbf{N}_{\mathcal{C}}$ and role names $\mathbf{N}_{\mathcal{R}}$, $\tau\mathcal{EL}(m)$ *concept descriptions* are defined as follows:

$$\widehat{C} ::= \top \mid A \mid \widehat{C} \sqcap \widehat{C} \mid \exists r.\widehat{C} \mid E_{\sim q}$$

where $A \in \mathbf{N}_{\mathcal{C}}$, $r \in \mathbf{N}_{\mathcal{R}}$, $\sim \in \{<, \leq, >, \geq\}$, $q \in [0, 1] \cap \mathbb{Q}$, E is an \mathcal{EL} concept description and \widehat{C} is a $\tau\mathcal{EL}(m)$ concept descriptions. Concepts of the form $E_{\sim q}$ are called *threshold concepts* and we denote as $\widehat{\mathbf{N}}_{\mathcal{E}}$ the set of all threshold concepts.

A $\tau\mathcal{EL}(m)$ ABox is an \mathcal{EL} ABox that, in addition, allows assertions of the form $\widehat{C}(a)$. The definition of *role-depth*, given in Section 2 for \mathcal{EL} concept descrip-

tions, extends to $\tau\mathcal{EL}(m)$ concept descriptions by defining $\text{rd}(E_{\sim q}) := 0$ for each threshold concept $E_{\sim q} \in \widehat{\mathbf{N}}_{\mathbf{E}}$.

The semantics of the new threshold concepts is defined in the following way:

$$[E_{\sim q}]^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid m^{\mathcal{I}}(d, E) \sim q\}.$$

The extension of $\cdot^{\mathcal{I}}$ to more complex concepts is defined as in \mathcal{EL} by additionally considering the underlying semantics of the newly introduced threshold concepts.

Requiring property *M1* has the following consequences for the semantics of threshold concepts.

Proposition 8. *For every \mathcal{EL} concept description E we have*

$$E_{\geq 1} \equiv E \quad \text{and} \quad E_{< 1} \equiv \neg E,$$

where the semantics of negation is defined as usual, i.e., $[\neg E]^{\mathcal{I}} := \Delta^{\mathcal{I}} \setminus E^{\mathcal{I}}$.

The second equivalence basically says that $\tau\mathcal{EL}(m)$ can express negation of \mathcal{EL} concept descriptions. This does not imply that $\tau\mathcal{EL}(m)$ is closed under negation since the threshold constructors can only be applied to \mathcal{EL} concept descriptions. Thus, negation cannot be nested using these constructors. A formal proof that $\tau\mathcal{EL}(deg)$ for the membership function *deg* introduced in the next section cannot express full negation can be found in Section 4.1. However, atomic negation (i.e., negation applied to concept names) can obviously be expressed. Consequently, unlike \mathcal{EL} concept descriptions, not all $\tau\mathcal{EL}(m)$ concept descriptions are satisfiable (i.e., can be interpreted by a non-empty set). A simple example is the concept description $A_{\geq 1} \sqcap A_{< 1}$, which is equivalent to $A \sqcap \neg A$.

3.1 Description graphs and homomorphisms in $\tau\mathcal{EL}(m)$

In this section we show that the characterization of membership in \mathcal{EL} presented in Section 2 can be extended to $\tau\mathcal{EL}(m)$. In addition, we will show that given an ABox \mathcal{A} and an interpretation \mathcal{I} , the question “is \mathcal{I} a model of \mathcal{A} ?” can also be characterized by the existence of homomorphisms. For this, we first extend \mathcal{EL} description graphs to $\tau\mathcal{EL}(m)$ description graphs. This is done by allowing the node labelling function to assign, in addition, threshold concepts as labels.

Definition 9 ($\tau\mathcal{EL}(m)$ description graph). A $\tau\mathcal{EL}(m)$ description graph is a graph of the form $\widehat{G} = (V_G, E_G, \widehat{\ell}_G)$ where:

- V_G is a set of nodes,
- $E_G \subseteq V_G \times \mathbf{N}_{\mathbf{R}} \times V_G$ is a set of edges labelled by role names, and

- $\widehat{\ell}_G : V_G \rightarrow 2^{\mathbf{N}_C \cup \widehat{\mathbf{N}}_E}$ is a function that labels nodes with subsets of $\mathbf{N}_C \cup \widehat{\mathbf{N}}_E$.

Like in \mathcal{EL} (see Definition 1), a $\tau\mathcal{EL}(m)$ description tree \widehat{T} is a $\tau\mathcal{EL}(m)$ description graph that is a tree with a distinguished element v_0 representing its root. Then we can establish a similar relationship between concept descriptions and description trees in $\tau\mathcal{EL}(m)$, i.e., every $\tau\mathcal{EL}(m)$ concept description \widehat{C} can be translated into a $\tau\mathcal{EL}(m)$ description tree $T_{\widehat{C}}$ and vice versa. The translation is the same as in \mathcal{EL} by considering, in addition, the concepts of the form $E_{\sim q}$. The following example illustrates such a relationship.

Example 10. Let E be an \mathcal{EL} concept description. The $\tau\mathcal{EL}(m)$ concept description

$$\widehat{C} := A \sqcap E_{>0.8} \sqcap \exists r.(A \sqcap B \sqcap E_{\leq 0.5} \sqcap \exists r.E_{<1}) \sqcap \exists r.A$$

yields the $\tau\mathcal{EL}(m)$ description tree depicted on the left-hand side of Figure 2. The \mathcal{EL} description tree T_C depicted in the right-hand side of Figure 2 represents the \mathcal{EL} description tree that results by ignoring the threshold concepts in the labels of $T_{\widehat{C}}$.

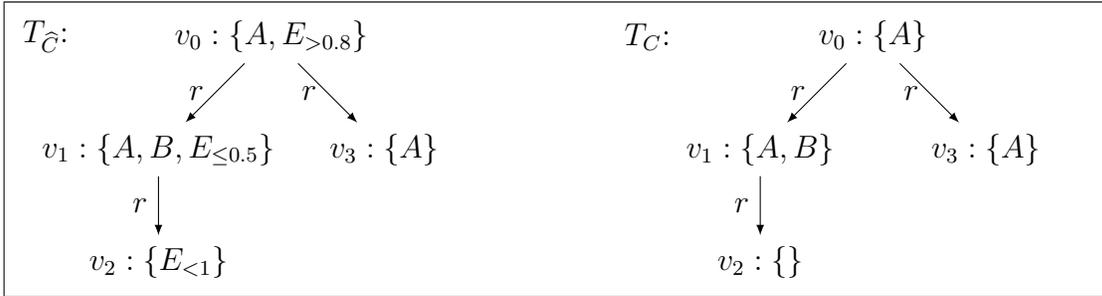


Figure 2: $\tau\mathcal{EL}(m)$ description trees.

If we consider an ABox the use of individual names and role assertions excludes the possibility of representing it as a concept description in \mathcal{EL} . Individuals in the ABox may have no relation at all or it could also be the case that role assertions enforce the existence of a cycle involving some of them. In fact, the translation of concept descriptions into description trees in \mathcal{EL} is adapted in [15] for an ABox \mathcal{A} into a description graph $G(\mathcal{A})$.

We now lift the very same translation (see Section 3 in [15]) to ABoxes and description graphs in $\tau\mathcal{EL}(m)$. Some of the notation used in [15] is slightly changed for the sake of readability within this document.

Definition 11 (ABoxes and $\tau\mathcal{EL}(m)$ description graphs). Let \mathcal{A} be a $\tau\mathcal{EL}(m)$ ABox. \mathcal{A} is translated into a $\tau\mathcal{EL}(m)$ description graph $\widehat{G}(\mathcal{A})$ in the following way:

- For each $a \in \text{Ind}(\mathcal{A})$ the $\tau\mathcal{EL}(m)$ description concept \widehat{C}_a is defined as:

$$\widehat{C}_a := \prod_{\widehat{D}(a) \in \mathcal{A}} \widehat{D}$$

If there exists no assertion of the form $\widehat{D}(a)$ in \mathcal{A} , then $\widehat{C}_a := \top$.

- For each $a \in \text{Ind}(\mathcal{A})$, let $\widehat{T}(a) = (V_a, E_a, a, \widehat{\ell}_a)$ be the $\tau\mathcal{EL}(m)$ description tree corresponding to the concept \widehat{C}_a where a itself represents its root. Without loss of generality let the sets V_a with $a \in \text{Ind}(\mathcal{A})$ be pairwise disjoint. Then, $\widehat{G}(\mathcal{A}) = (V_{\mathcal{A}}, E_{\mathcal{A}}, \widehat{\ell}_{\mathcal{A}})$ is defined as:
 - $V_{\mathcal{A}} := \bigcup_{a \in \text{Ind}(\mathcal{A})} V_a$,
 - $E_{\mathcal{A}} := \bigcup_{a \in \text{Ind}(\mathcal{A})} E_a \cup \{arb \mid r(a, b) \in \mathcal{A}\}$, and
 - $\widehat{\ell}_{\mathcal{A}}(v) := \widehat{\ell}_a(v)$ for $v \in V_a$.

The following example illustrates the previous definition.

Example 12. Let E be an \mathcal{EL} concept description and \mathcal{A} be the following ABox:

$$\mathcal{A} := \{A(a), B(a), E_{<1}(b), (\exists r.A)(d), r(a, b), r(b, c), s(c, a)\}$$

the corresponding $\tau\mathcal{EL}(m)$ description graph $\widehat{G}(\mathcal{A})$ is depicted in Figure 3.

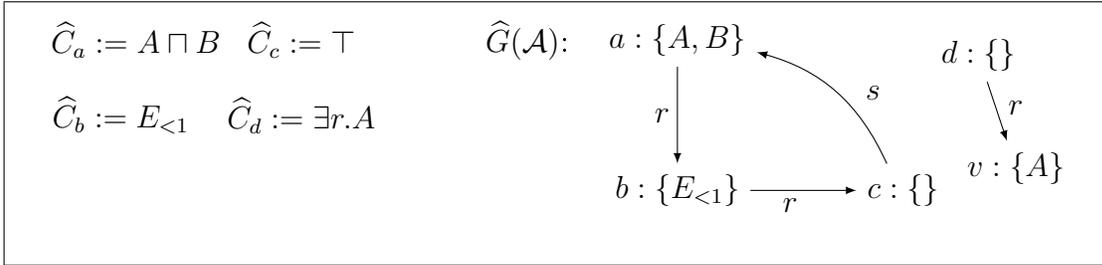


Figure 3: $\tau\mathcal{EL}(m)$ description graph associated to an ABox.

Using the notion of $\tau\mathcal{EL}(m)$ description graphs, we define homomorphisms from $\tau\mathcal{EL}(m)$ description graphs to the associated \mathcal{EL} description graph of an interpretation \mathcal{I} . To differentiate these kinds of homomorphisms from the classical ones, we name them τ -homomorphisms and use the greek letter ϕ (possibly with subscripts) to denote them.

Definition 13. Let $\widehat{H} = (V_H, E_H, \widehat{\ell}_H)$ be a $\tau\mathcal{EL}(deg)$ description graph and \mathcal{I} an interpretation. The mapping $\phi : V_H \rightarrow V_{\mathcal{I}}$ is a τ -homomorphism from \widehat{H} to $G_{\mathcal{I}}$ iff

1. ϕ is a homomorphism from \widehat{H} to $G_{\mathcal{I}}$ according to Definition 3, where threshold concepts in labels are ignored, and
2. for all $v \in V_H$: if $E_{\sim q} \in \widehat{\ell}_H(v)$, then $\phi(v) \in [E_{\sim q}]^{\mathcal{I}}$.

We denote by $\text{dom}(\phi)$ the domain of ϕ , i.e., $\text{dom}(\phi) := V_H$. Furthermore, we denote by $\text{img}(\phi)$ the image of the mapping ϕ , i.e., $\text{img}(\phi) := \{\phi(v) \mid v \in V_H\}$. Note that τ -homomorphisms are a generalization of the classical notion of homomorphism in \mathcal{EL} from Definition 3.

We now provide a characterization of *element-hood* for $\tau\mathcal{EL}(m)$ concept descriptions. This characterization is based on the existence of a τ -homomorphism and generalizes Lemma 5 from \mathcal{EL} to $\tau\mathcal{EL}(m)$.

Theorem 14. *Let \widehat{C} be a $\tau\mathcal{EL}(m)$ concept description and $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ an interpretation. The following statements are equivalent for all $d \in \Delta^{\mathcal{I}}$:*

1. $d \in \widehat{C}^{\mathcal{I}}$.
2. there exists a τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = d$.

Proof. The 1) \rightarrow 2) direction is shown by induction on the role-depth of \widehat{C} , while the other direction is proved by induction on the number of nodes in $T_{\widehat{C}}$. The details of the proof are deferred to the Appendix. \square

Using this lemma we give a similar characterization for ABoxes and interpretations in $\tau\mathcal{EL}(m)$.

Theorem 15. *Let \mathcal{A} be a $\tau\mathcal{EL}(m)$ ABox and $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation. The following statements are equivalent:*

1. \mathcal{I} is a model of \mathcal{A} .
2. there exists a τ -homomorphism ϕ from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}}$ such that $\phi(a) = a^{\mathcal{I}}$ for all $a \in \text{Ind}(\mathcal{A})$.

Proof. 1) \rightarrow 2). Assume that \mathcal{I} is a model of \mathcal{A} . Then, it holds that $a^{\mathcal{I}} \in \widehat{D}^{\mathcal{I}}$ for each assertion of the form $\widehat{D}(a) \in \mathcal{A}$. Therefore, by definition of \widehat{C}_a we obtain that $a^{\mathcal{I}} \in [\widehat{C}_a]^{\mathcal{I}}$. Now, by Theorem 14 and since a is the root of $\widehat{T}(a)$ we know that there exists a τ -homomorphism ϕ_a from $\widehat{T}(a)$ to $G_{\mathcal{I}}$ with $\phi_a(a) = a^{\mathcal{I}}$. Further, we have $a^{\mathcal{I}}rb^{\mathcal{I}} \in E_{\mathcal{I}}$ for all $r(a, b) \in \mathcal{A}$. Hence, since all the sets V_a used to build $\widehat{G}(\mathcal{A})$ are pairwise disjoint, it is easy to verify that $\phi := \bigcup_{a \in \text{Ind}(\mathcal{A})} \phi_a$ is a τ -homomorphism from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}}$ such that $\phi(a) = a^{\mathcal{I}}$ for all $a \in \text{Ind}(\mathcal{A})$.

2) \rightarrow 1) Assume that statement 2) holds. We show that \mathcal{I} is a model of \mathcal{A} :

- $r(a, b) \in \mathcal{A}$. By construction of $\widehat{G}(\mathcal{A})$ we know that $arb \in E_{\mathcal{A}}$ and since ϕ is a homomorphism from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}}$, we also have $\phi(a)r\phi(b) \in E_{\mathcal{I}}$. Then, from $\phi(a) = a^{\mathcal{I}}$ we obtain $a^{\mathcal{I}}rb^{\mathcal{I}} \in E_{\mathcal{I}}$ and therefore: $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$.
- $\widehat{D}(a) \in \mathcal{A}$. By construction of $\widehat{G}(\mathcal{A})$, one can see that the description graph $\widehat{T}(a)$ is a sub-graph of the description graph $\widehat{G}(\mathcal{A})$. Then, it is not difficult to see that ϕ is also a τ -homomorphism from $\widehat{T}(a)$ to $G_{\mathcal{I}}$ with $\phi(a) = a^{\mathcal{I}}$. Then, the application of Theorem 14 yields $a^{\mathcal{I}} \in [\widehat{C}_a]^{\mathcal{I}}$. Since \widehat{D} is one of the conjuncts in the definition of \widehat{C}_a , it follows that $a^{\mathcal{I}} \in \widehat{D}^{\mathcal{I}}$.

□

3.2 Deciding the existence of a τ -homomorphism

If the interpretation \mathcal{I} is finite and the degree membership function m is computable, then the existence of a τ -homomorphism can be decided. We provide two algorithms which decide, under the previous conditions, the existence of a τ -homomorphism according with the characterizations given in Theorems 14 and 15.

Our 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.

- 1: Let $T_1 = (V_1, E_1, v_0, \ell_1)$ and $T_2 = (V_2, E_2, w_0, \ell_2)$. Further, let $\{v_1, \dots, v_n\}$ be a post-order sequence of V_1 , i.e., v_1 is a leaf and $v_n = v_0$.
 - 2: Define a labelling $\delta : V_2 \rightarrow \mathcal{P}(V_1)$ as follows.
 - 3: Initialize δ by $\delta(w) := \emptyset$ for all $w \in V_2$.
 - 4: **for all** $1 \leq i \leq n$ **do**
 - 5: **for all** $w \in V_2$ **do**
 - 6: If $\ell_1(v_i) \subseteq \ell_2(w)$ **and**
 - 7: for all $v_i r v \in E_1$ there is $w' \in V_2$ such that
 - 8: $v \in \delta(w')$ **and** $w r w' \in E_2$
 - 9: then $\delta(w) := \delta(w) \cup \{v_i\}$
 - 10: **end for**
 - 11: **end for**
 - 12: If $v_0 \in \delta(w_0)$, then return “yes”, else return “no”.
-

In Theorem 14, membership in $\tau\mathcal{EL}(m)$ concept descriptions is characterized by the existence of a τ -homomorphism from a $\tau\mathcal{EL}(m)$ description tree $T_{\widehat{C}}$ to $G_{\mathcal{I}}$. If

\mathcal{I} is finite, then Algorithm 1 can be used to decide whether there exists a mapping satisfying Condition 1 in Definition 13. We only need to replace the last line by $v_0 \in \delta(w)$ for some $w \in V_{\mathcal{I}}$, since now T_2 becomes $G_{\mathcal{I}}$ which has no root. In order to verify Condition 2 in Definition 13, we then extend the test in line 6 to check whether $m^{\mathcal{I}}(w, E) \sim q$ for each $E_{\sim q} \in \widehat{\ell}_{T_{\widehat{C}}}(v_i)$. Algorithm 2 below, implements this addition to Algorithm 1. Note that a simple modification in line 12, namely $v_0 \in \delta(d)$, adapts the algorithm to answer the question of whether $d \in \widehat{C}^{\mathcal{I}}$ for a specific $d \in \Delta^{\mathcal{I}}$.

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 $\{v_1, \dots, v_n\}$ be a post-order sequence of V_T , i.e., v_1 is a leaf and $v_n = v_0$.
 - 2: Define a labelling $\delta : V_{\mathcal{I}} \rightarrow \mathcal{P}(V_T)$ as follows.
 - 3: Initialize δ by $\delta(w) := \emptyset$ for all $w \in V_{\mathcal{I}}$.
 - 4: **for all** $1 \leq i \leq n$ **do**
 - 5: **for all** $w \in V_{\mathcal{I}}$ **do**
 - 6: If $\ell_T(v_i) \subseteq \ell_{\mathcal{I}}(w)$ **and** $[E_{\sim q} \in \widehat{\ell}_T(v_i) \Rightarrow m^{\mathcal{I}}(w, E) \sim q]$ **and**
 - 7: for all $v_i r v \in E_T$ there is $w' \in V_{\mathcal{I}}$ such that
 - 8: $v \in \delta(w')$ **and** $w r w' \in E_{\mathcal{I}}$
 - 9: then $\delta(w) := \delta(w) \cup \{v_i\}$
 - 10: **end for**
 - 11: **end for**
 - 12: If there exists $w \in V_{\mathcal{I}}$ such that $v_0 \in \delta(w)$, then return “yes”, else return “no”.
-

Regarding the complexity of Algorithm 2, the main difference with Algorithm 1 is the computation of $m^{\mathcal{I}}$. Therefore, its complexity depends on how difficult is to compute the chosen $m^{\mathcal{I}}$. For instance, if $m^{\mathcal{I}}$ can be computed in polynomial time as for the membership function *deg* introduced in the next section, Algorithm 2 will run in polynomial time.

Algorithm 3 decides, given a finite interpretation \mathcal{I} and an ABox \mathcal{A} , whether \mathcal{I} is a model of \mathcal{A} . It is based on the characterization in Theorem 15 and uses Algorithm 2. Note that the description graph $\widehat{G}(\mathcal{A})$ associated to an ABox \mathcal{A} is not necessarily a tree. Therefore, finding a τ -homomorphism from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}}$ includes finding a homomorphism between two graphs, which in general is an NP-complete problem [12]. However, by Definition 11, it can be seen that $\widehat{G}(\mathcal{A})$ has a particular form where cycles only involve nodes and edges in the graph corresponding to the individual elements and role assertions, respectively, occurring in \mathcal{A} . Since Theorem 15 requires $\phi(a) = a^{\mathcal{I}}$ for all $a \in \text{Ind}(\mathcal{A})$, the wanted τ -homomorphism is partially fixed with respect to those elements. Then, it suffices to check whether the interpretation of the individual names satisfies

the role assertions in \mathcal{A} and $a \in [\widehat{C}_a]^{\mathcal{I}}$ (see Definition 11), for all $a \in \text{Ind}(\mathcal{A})$.

Algorithm 3 τ -homomorphisms for ABoxes and interpretations.

Input: An ABox \mathcal{A} and a finite interpretation \mathcal{I} .

Output: “yes”, if there exists a τ -homomorphism ϕ from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}}$ with $\phi(a) = a^{\mathcal{I}}$, “no”, otherwise.

- 1: Let $\widehat{G}(\mathcal{A})$ be as in Definition 11 and $G_{\mathcal{I}} = (V_{\mathcal{I}}, E_{\mathcal{I}}, \ell_{\mathcal{I}})$.
 - 2: **for all** $r(a, b) \in \mathcal{A}$ **do**
 - 3: If $a^{\mathcal{I}} r b^{\mathcal{I}} \notin E_{\mathcal{I}}$ then
 - 4: return “no”.
 - 5: **end for**
 - 6: **for all** $a \in \text{Ind}(\mathcal{A})$ **do**
 - 7: If $a^{\mathcal{I}} \notin [\widehat{C}_a]^{\mathcal{I}}$ then {this can be checked using Algorithm 2}
 - 8: return “no”.
 - 9: **end for**
 - 10: return “yes”.
-

4 The membership function deg

To make things more concrete, we now introduce a specific membership function, denoted deg . Given an interpretation \mathcal{I} , an element $d \in \Delta^{\mathcal{I}}$, and an \mathcal{EL} concept description C , this function is supposed to measure to which degree d satisfies the conditions for membership expressed by C . To come up with such a measure, we use the homomorphism characterization of membership (see Theorem 5) as starting point. Basically, we consider all partial mappings from T_C to $G_{\mathcal{I}}$ that map the root of T_C to d and respect the edge structure of T_C . For each of these mappings we then calculate to which degree it satisfies the homomorphism conditions, and take the degree of the best such mapping as the membership degree $deg^{\mathcal{I}}(d, C)$.

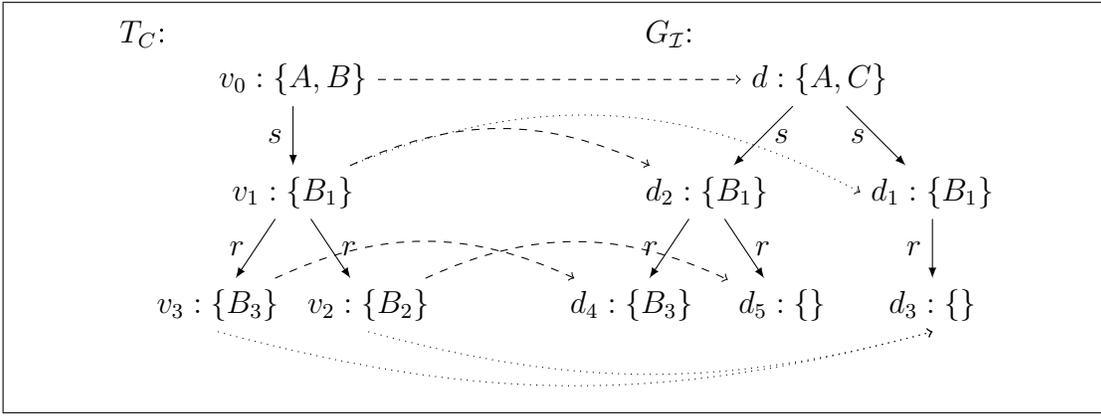


Figure 4

Example 16. Figure 4 shows the \mathcal{EL} description tree corresponding to the \mathcal{EL} concept description $C := A \sqcap B \sqcap \exists s.(B_1 \sqcap \exists r.B_3 \sqcap \exists r.B_2)$ and a fragment of an interpretation graph $G_{\mathcal{I}}$. In addition, it depicts two mappings from V_{T_C} to $V_{G_{\mathcal{I}}}$. The one represented by the *dashed* lines and a variation represented with the *dotted* lines. None of them are a homomorphism from T_C to $G_{\mathcal{I}}$ in the sense of Definition 3 and moreover, since obviously $d \notin C^{\mathcal{I}}$, by Theorem 5 there exists no such homomorphism.

To compute the membership value induced by an specific mapping, we count the number of properties of v_0 (say m), see how many of those does d in \mathcal{I} actually have (say n) and give $\frac{n}{m}$ as the membership degree value. In the example v_0 has three properties, e.g., A , B and the existence of an s -successor (represented by v_1) with certain properties. Interesting to see is that for both mappings, the selected s -successor of d does not satisfy all the properties of v_1 . Should we just assume that d does not have this last property and give $\frac{1}{3}$ as the membership degree value? Instead of that, we would like to compute a value that expresses to which degree the s -successor of d (to which v_1 is mapped to), satisfies the conditions for membership expressed by the subtree of T_C rooted at v_1 . This will be done using

the very same idea recursively.

As mentioned before, we consider partial mappings rather than total ones since one of the violations of properties demanded by C could be that a required role successor does not exist at all.

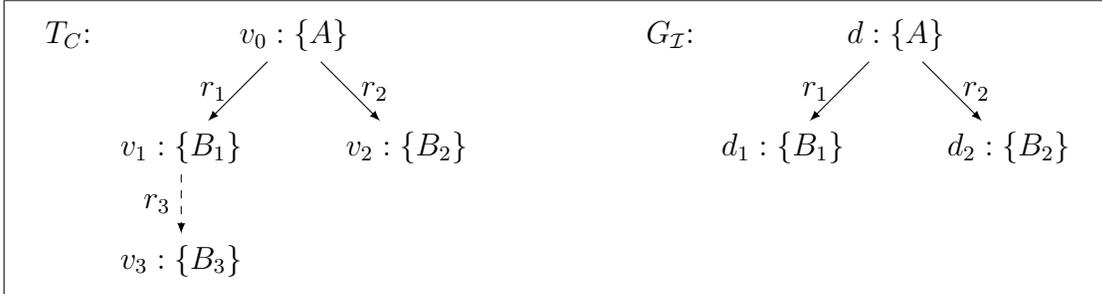


Figure 5

Example 17. Consider the description tree T_C and the fragment of \mathcal{I} depicted in Figure 5. Obviously, there exists no total mapping from T_C to G_I since neither d_1 nor d_2 have a successor. Thus, restricting to consider only total mappings would give *zero* as the membership degree value of d in C . This is not desired, since just like concept names may be missing and the membership value does not become zero, also role successors (required by C) may be missing and the membership degree not need to be zero.

To formalize this idea, we first define the notion of *partial tree-to-graph homomorphisms* from description trees to description graphs. In this definition, the node labels are ignored (they will be considered in the next step).

Definition 18 (Partial tree-to-graph homomorphisms). Let $T = (V_t, E_t, \ell_t, v_0)$ and $G = (V_g, E_g, \ell_g)$ be a description tree (with root v_0) and a description graph, respectively. A partial mapping $h : V_t \rightarrow V_g$ is a *partial tree-to-graph homomorphism* (*ptgh*) from T to G iff the following conditions are satisfied:

1. $\text{dom}(h)$ is a sub-tree of T with root v_0 , i.e., $v_0 \in \text{dom}(h)$ and if $(v, r, w) \in E_t$ and $w \in \text{dom}(h)$, then $v \in \text{dom}(h)$;
2. for all edges $(v, r, w) \in E_t$, $w \in \text{dom}(h)$ implies $(h(v), r, h(w)) \in E_g$.

To abbreviate, from now on we will write *ptgh* (*ptghs* for the plural) instead of *partial tree-to-graph homomorphism*.

In order to measure how far away from a homomorphism according to Definition 3 such a *ptgh* is, we define the notion of a *weighted* homomorphism between a finite \mathcal{EL} description tree and an \mathcal{EL} description graph.

Definition 19. Let T be a finite \mathcal{EL} description tree, G an \mathcal{EL} description graph and $h : V_T \rightarrow V_G$ a *ptgh* from T to G . We define the *weighted* homomorphism induced by h from T to G as a recursive function $h_w : \text{dom}(h) \rightarrow [0..1]$ as follows:

$$h_w(v) := \begin{cases} 1 & \text{if } |\ell_T(v)| + k^*(v) = 0 \\ \frac{|\ell_T(v) \cap \ell_G(h(v))| + \sum_{1 \leq i \leq k} h_w(v_i)}{|\ell_T(v)| + k^*(v)} & \text{otherwise.} \end{cases}$$

The elements used to define h_w have the following meaning. For a given $v \in \text{dom}(h)$, $k^*(v)$ denotes the number of successors of v in T , and v_1, \dots, v_k with $0 \leq k \leq k^*(v)$ are the children of v in T such that $v_i \in \text{dom}(h)$.

It is easy to see that h_w is well-defined. In fact, T is a finite tree, which ensures that the recursive definition of h_w is well-founded. In addition, the first case in the definition ensures that division by zero is avoided. Using value 1 in this case is justified since then no property is required. In the second case, missing concept names and missing successors decrease the weight of a node since then the required name or successor contributes to the denominator, but not to the numerator. Required successors that are there are only counted if they are successors for the correct role, and then they do not contribute with value 1 to the numerator, but only with their weight (i.e., the degree to which they match the requirements for this successor).

When defining the value of the membership function $\text{deg}^{\mathcal{I}}(d, C)$, we do not use the concept C directly, but rather its reduced form C^r . This will ensure that deg satisfies property *M2*.

Definition 20. Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation, d an element of $\Delta^{\mathcal{I}}$ and C an \mathcal{EL} concept description with reduced form C^r . In addition, let $\mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$ be the set of all *ptghs* from T_{C^r} to $G_{\mathcal{I}}$ with $h(v_0) = d$. The set $\mathcal{V}^{\mathcal{I}}(d, C^r)$ of all relevant values is defined as:

$$\mathcal{V}^{\mathcal{I}}(d, C^r) := \{q \mid h_w(v_0) = q \text{ and } h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)\}$$

Then we define $\text{deg}^{\mathcal{I}}(d, C) := \max \mathcal{V}^{\mathcal{I}}(d, C^r)$.

In case the interpretation \mathcal{I} is infinite, there may exist infinitely many *ptghs* from T_{C^r} to $G_{\mathcal{I}}$ with $h(v_0) = d$. Therefore, it is not immediately clear whether the maximum in the above definition actually exists, and thus whether $\text{deg}^{\mathcal{I}}(d, C)$ is well-defined. To prove that the maximum exists also for infinite interpretations, we show that the set $\mathcal{V}^{\mathcal{I}}(d, C^r)$ is actually a finite set. For this purpose, we introduce canonical interpretations induced by *ptghs*.

Definition 21 (Canonical interpretation). Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation, C an \mathcal{EL} concept description and h be a *ptgh* from T_{C^r} to $G_{\mathcal{I}}$. The *canonical*

interpretation \mathcal{I}_h induced by h is the one having the description tree $T_{\mathcal{I}_h} = (V_{\mathcal{I}_h}, E_{\mathcal{I}_h}, v_0, \ell_{\mathcal{I}_h})$ with

$$\begin{aligned} V_{\mathcal{I}_h} &:= \text{dom}(h), \\ E_{\mathcal{I}_h} &:= \{vrw \in E_{T_{C^r}} \mid v, w \in \text{dom}(h)\} \\ \ell_{\mathcal{I}_h}(v) &:= \ell_{T_{C^r}}(v) \cap \ell_{\mathcal{I}}(h(v)) \text{ for all } v \in \text{dom}(h). \end{aligned}$$

Remark 22. From the previous definition, one can see that $T_{\mathcal{I}_h}$ satisfies $V_{\mathcal{I}_h} \subseteq V_{T_{C^r}}$, $E_{\mathcal{I}_h} \subseteq E_{T_{C^r}}$, $\ell_{\mathcal{I}_h}(v) \subseteq \ell_{T_{C^r}}(v)$ and $\ell_{\mathcal{I}_h}(v) \subseteq \ell_{\mathcal{I}}(h(v))$ for all $v \in \text{dom}(h)$. Moreover, the construction of \mathcal{I}_h verifies that the mapping h is a homomorphism from $T_{\mathcal{I}_h}$ to $G_{\mathcal{I}}$.

Lemma 23. Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation, $d \in \Delta^{\mathcal{I}}$ and C an \mathcal{EL} concept description. The set $\mathcal{V}^{\mathcal{I}}(d, C^r)$ contains finitely many elements.

Proof. Let $\mathcal{I}_{\mathcal{H}}$ be the set of all canonical interpretations induced by each $h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$, i.e.,

$$\mathcal{I}_{\mathcal{H}} := \{\mathcal{I}_h \mid h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)\}$$

Consider now the following set $\{(V_{\mathcal{I}_h}, E_{\mathcal{I}_h}, v_0) \mid \mathcal{I}_h \in \mathcal{I}_{\mathcal{H}}\}$, i.e., the set of trees corresponding to interpretations in $\mathcal{I}_{\mathcal{H}}$ without labels. From Remark 22 we have that $V_{\mathcal{I}_h} \subseteq V_{T_{C^r}}$ and $|E_{\mathcal{I}_h}| \leq |E_{T_{C^r}}|$. Therefore, the previously defined set is finite. Since each interpretation \mathcal{I}_h corresponds to one of these trees with the addition of the labelling function and we assume that $\mathbf{N}_{\mathcal{C}}$ and $\mathbf{N}_{\mathcal{R}}$ are finite, it follows that $\mathcal{I}_{\mathcal{H}}$ must be a finite set. Hence, there are only finitely many different canonical interpretations induced by *ptghs* $h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$.

Now, consider any $h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$ and let $i^{\mathcal{I}_h} : \text{dom}(h) \rightarrow V_{\mathcal{I}_h}$ be a mapping such that $i^{\mathcal{I}_h}(v) = v$ for all $v \in \text{dom}(h)$. Note that $i^{\mathcal{I}_h}$ is well-defined by definition of \mathcal{I}_h and it is easy to see that it is a *ptgh* from T_{C^r} to $T_{\mathcal{I}_h}$. Furthermore, let $\mathcal{V}^{\mathcal{I}_{\mathcal{H}}}$ be the set:

$$\mathcal{V}^{\mathcal{I}_{\mathcal{H}}} := \{q \mid i_w^{\mathcal{I}_h}(v_0) = q \text{ for all } h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)\}$$

Since $\text{dom}(h) \subseteq V_{T_{C^r}}$, there are finitely many sets that could act as the source for a mapping $i^{\mathcal{I}_h}$. Moreover, $\mathcal{I}_{\mathcal{H}}$ is a finite set of finite interpretations. Hence, there can only be finitely many different mappings $i^{\mathcal{I}_h}$. Consequently, the set $\mathcal{V}^{\mathcal{I}_{\mathcal{H}}}$ must be finite. In addition, one can see that the following three properties hold:

- $\text{dom}(i^{\mathcal{I}_h}) = \text{dom}(h)$,
- $\ell_{\mathcal{I}_h}(i^{\mathcal{I}_h}(v)) = \ell_{T_{C^r}}(v) \cap \ell_{\mathcal{I}}(h(v))$ for all $v \in \text{dom}(h)$, and
- for all $v, w \in \text{dom}(h)$: if $vrw \in E_{T_{C^r}}$, then $h(v)rh(w) \in E_{\mathcal{I}}$ and $i^{\mathcal{I}_h}(v)ri^{\mathcal{I}_h}(w) \in E_{\mathcal{I}_h}$.

Then, from Definition 19 it follows that $h_w(v_0) = i_w^{\mathcal{I}^h}(v_0)$. This means that for all $h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$ it is the case that $h_w(v_0) \in \mathcal{V}^{\mathcal{I}^h}$. Hence, $\mathcal{V}^{\mathcal{I}}(d, C^r) \subseteq \mathcal{V}^{\mathcal{I}^h}$ and $\mathcal{V}^{\mathcal{I}}(d, C^r)$ is a finite set.

Thus, $\max \mathcal{V}^{\mathcal{I}}(d, C^r)$ exists and $\deg^{\mathcal{I}}(d, C)$ is well-defined. \square

Based on the proof of Lemma 23, we show that the value $\deg^{\mathcal{I}}(d, C)$ is preserved by the canonical interpretation corresponding to a *ptgh* h such that $h_w(v_0) = \deg^{\mathcal{I}}(d, C)$. To this end, we also use the following lemma (see the Appendix) which shows that \deg satisfies a monotonicity property with respect to two interpretations \mathcal{I} and \mathcal{J} which are related by the existence of a homomorphism.

Lemma 24. *Let \mathcal{I} and \mathcal{J} be two interpretations such that there exists a homomorphism φ from $G_{\mathcal{I}}$ to $G_{\mathcal{J}}$. Then, for any individual $d \in \Delta^{\mathcal{I}}$ and any \mathcal{EL} concept description C it holds: $\deg^{\mathcal{I}}(d, C) \leq \deg^{\mathcal{J}}(\varphi(d), C)$.*

Lemma 25. *Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation, d be an individual of $\Delta^{\mathcal{I}}$ and C an \mathcal{EL} concept description. Let h be a *ptgh* from T_{C^r} to $G_{\mathcal{I}}$ such that $h(v_0) = d$ and $h_w(v_0) = \deg^{\mathcal{I}}(d, C^r)$. In addition, let \mathcal{I}_h be the canonical interpretation induced by h . Then, $\deg^{\mathcal{I}_h}(v_0, C) = \deg^{\mathcal{I}}(d, C)$.*

Proof. Assume that $\deg^{\mathcal{I}}(d, C) = q$. From Definition 20 we have:

$$\deg^{\mathcal{I}}(d, C) = \max \mathcal{V}^{\mathcal{I}}(d, C^r) = \deg^{\mathcal{I}}(d, C^r) = q$$

In the proof of Lemma 23 we saw that $i^{\mathcal{I}_h}$ is a *ptgh* from T_{C^r} to $T_{\mathcal{I}_h}$ with $i^{\mathcal{I}_h}(v_0) = v_0$ and $h_w(v_0) = i_w^{\mathcal{I}_h}(v_0)$. Hence, $\deg^{\mathcal{I}_h}(v_0, C^r) \geq q$. Remark 22 tells us that h is a homomorphism from $T_{\mathcal{I}_h}$ to $G_{\mathcal{I}}$ with $h(v_0) = d$. Then, the application of Lemma 24 yields:

$$\deg^{\mathcal{I}_h}(v_0, C^r) \leq \deg^{\mathcal{I}}(d, C^r)$$

Therefore, $\deg^{\mathcal{I}_h}(v_0, C^r) = q$ and consequently, Definition 20 yields $\deg^{\mathcal{I}_h}(v_0, C) = q = \deg^{\mathcal{I}}(d, C)$. \square

If the interpretation \mathcal{I} is finite, $\deg^{\mathcal{I}}(d, C)$ for $d \in \Delta^{\mathcal{I}}$ and an \mathcal{EL} concept description C can actually be computed in polynomial time. The polynomial time algorithm described below is inspired by the polynomial time algorithm for checking the existence of a homomorphism between \mathcal{EL} description trees [11, 4], and similar to the algorithm for computing the similarity degree between \mathcal{EL} concept descriptions introduced in [17].

Algorithm 4 Computation of $deg^{\mathcal{I}}$.

Input: An \mathcal{EL} concept description C , a finite interpretation \mathcal{I} and $d \in \Delta^{\mathcal{I}}$.

Output: $deg^{\mathcal{I}}(d, C)$.

- 1: Let C^r be the reduced form of C , $G_{\mathcal{I}} = (V_{\mathcal{I}}, E_{\mathcal{I}}, \ell_{\mathcal{I}})$ and $\{v_1, \dots, v_n\}$ be a post-order sequence of $V_{T_{C^r}}$ where $v_n = v_0$.
 - 2: The assignment $S : V_{T_{C^r}} \times V_{\mathcal{I}} \rightarrow [0..1]$ is computed as follows:
 - 3: **for all** $1 \leq i \leq n$ **do**
 - 4: **if** $|\ell_{T_{C^r}}(v_i)| + k^*(v_i) = 0$ **then**
 - 5: $S(v_i, e) := 1$ for all $e \in \Delta^{\mathcal{I}}$
 - 6: **else**
 - 7: **for all** $e \in V_{\mathcal{I}}$ **do**
 - 8: $c := |\ell_{T_{C^r}}(v_i) \cap \ell_{\mathcal{I}}(e)|$
 - 9: **for all** $v_i r v \in E_{T_{C^r}}$ **do**
 - 10: $c := c + \max\{S(v, e') \mid e r e' \in E_{\mathcal{I}}\}$
 - 11: **end for**
 - 12: $S(v_i, e) := \frac{c}{|\ell_{T_{C^r}}(v_i)| + k^*(v_i)}$
 - 13: **end for**
 - 14: **end if**
 - 15: **end for**
 - 16: **return** $S(v_0, d)$.
-

Since the algorithm considers each pair of nodes (v, e) with $v \in V_{T_{C^r}}$ and $e \in V_{\mathcal{I}}$ only once, it is easy to see that it runs in polynomial time in the size of C and \mathcal{I} . The following lemma shows that Algorithm 4 computes the value of $deg^{\mathcal{I}}$, i.e., $deg^{\mathcal{I}}(d, C) = S(v_0, d)$ (see the Appendix).

Lemma 26. *Let C be an \mathcal{EL} concept description, \mathcal{I} a finite interpretation and $d \in \Delta^{\mathcal{I}}$. Then, Algorithm 4 terminates on input (C, \mathcal{I}, d) and outputs $deg^{\mathcal{I}}(d, C)$, i.e., $S(v_0, d) = deg^{\mathcal{I}}(d, C^r)$.*

Finally, it remains to show that deg satisfies the properties required for a membership function.

Proposition 27. *The function deg satisfies the properties M1 and M2.*

Proof. We first show that M1 is satisfied by deg . Assume that $d \in C^{\mathcal{I}}$. Since C is equivalent to its reduced form, we also have $d \in [C^r]^{\mathcal{I}}$. The application of Theorem 5 yields that there exists a homomorphism φ from T_{C^r} to $G_{\mathcal{I}}$ with $\varphi(v_0) = d$. Then it is easy to verify from Definition 19 that $\varphi_w(v_0) = 1$ and hence, $deg^{\mathcal{I}}(d, C^r) = 1$. Thus, $deg^{\mathcal{I}}(d, C) = 1$. Conversely, assume that $deg^{\mathcal{I}}(d, C) = 1$. This means that there exists a *ptgh* h from T_{C^r} to $G_{\mathcal{I}}$ with $h(v_0) = d$ and $h_w(v_0) = 1$. Similar as before, it is easy to see that h is homomorphism according

to Definition 3. The application of Theorem 5 yields $d \in [C^r]^{\mathcal{I}}$ and consequently, $d \in C^{\mathcal{I}}$.

We now turn into *M2*. As mentioned in Section 3, the right to left implication is already a consequence of *M1*, which we just proved to be satisfied by deg . Assume that $C \equiv D$, then by Theorem 6 there exists an isomorphism ψ between T_{C^r} and T_{D^r} . Consider an arbitrary interpretation \mathcal{I} and any element d of $\Delta^{\mathcal{I}}$. We show that $deg^{\mathcal{I}}(d, C^r) = deg^{\mathcal{I}}(d, D^r)$, which obviously implies $deg^{\mathcal{I}}(d, C) = deg^{\mathcal{I}}(d, D)$ (see Definition 20).

Let h be a *ptgh* h from T_{C^r} to $G_{\mathcal{I}}$ with $h(v_0) = d$ and $h_w(v_0) = \max \mathcal{V}^{\mathcal{I}}(d, C^r)$. Since ψ is an isomorphism, it is not hard to see that the composition $h \circ \psi$ is a *ptgh* from T_{D^r} to $G_{\mathcal{I}}$, with $(h \circ \psi)(v_0) = d$ and $(h \circ \psi)_w(v_0) = h_w(v_0)$. This means that $deg^{\mathcal{I}}(d, C^r) \leq deg^{\mathcal{I}}(d, D^r)$. Since the same reasoning applies starting with T_{D^r} , we thus have shown $deg^{\mathcal{I}}(d, C^r) = deg^{\mathcal{I}}(d, D^r)$. \square

Note that the proof of *M2* is based on the fact that we use the reduced form of a concept description rather than the description itself. Otherwise, *M2* would not hold. For example, consider the concept description $C := \exists r.A \sqcap \exists r.(A \sqcap B)$, which is equivalent to its reduced form $C^r = \exists r.(A \sqcap B)$. Let d be an individual that has a single r -successor belonging to A , but not to B . Then using C instead of C^r would yield membership degree $\frac{3}{4}$, whereas the use of C^r yields the degree $\frac{1}{2}$.

4.1 Some properties of $\tau\mathcal{EL}(deg)$

We mentioned in Section 3 that, although $\tau\mathcal{EL}(m)$ can express negation of \mathcal{EL} concept descriptions, negation cannot be nested using the constructors of $\tau\mathcal{EL}(m)$. We now formally prove that deg cannot express full negation by showing that it cannot express the concept constructor $\forall r.A$. In addition, we highlight that negated threshold concepts can be equivalently expressed by threshold concepts.

Let us start with the concept constructor $\forall r.C$, whose semantics is defined as follows:

$$[\forall r.C]^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \forall e \in \Delta^{\mathcal{I}}. ((d, e) \in r^{\mathcal{I}} \Rightarrow e \in C^{\mathcal{I}})\}$$

We show, that there exists no $\tau\mathcal{EL}(deg)$ concept description equivalent to the very simple concept description $\forall r.A$ where $A \in \mathbf{N}_{\mathbf{C}}$.

Lemma 28. *In $\tau\mathcal{EL}(deg)$, there is no concept description \widehat{C} such that $\forall r.A \equiv \widehat{C}$, where $A \in \mathbf{N}_{\mathbf{C}}$.*

Proof. Suppose we could find a $\tau\mathcal{EL}(deg)$ concept description \widehat{C} such that $\forall r.A \equiv \widehat{C}$. Then for each interpretation \mathcal{I} we have $[\forall r.A]^{\mathcal{I}} = \widehat{C}^{\mathcal{I}}$. Consider now the interpretation $\mathcal{I}_0 = (\{d\}, \cdot^{\mathcal{I}_0})$ such that $X^{\mathcal{I}_0} = \emptyset$ for all $X \in \mathbf{N}_{\mathbf{C}} \cup \mathbf{N}_{\mathbf{R}}$. It is obvious that $d \in [\forall r.A]^{\mathcal{I}_0}$ and by assumption $d \in \widehat{C}^{\mathcal{I}_0}$.

By Theorem 14 there exists a τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}_0}$ with $\phi(v_0) = d$. Since d has no r -successors in $\Delta^{\mathcal{I}_0}$ nor it is an instance of any concept name, this means that \widehat{C} must be of the following form:

$$E^1_{\sim q_1} \sqcap \dots \sqcap E^k_{\sim q_k}$$

where each E^i is an \mathcal{EL} concept description. Note that, without loss of generality, one can forget conjuncts of the form \top . In addition, $\widehat{C} = \top$ is not possible since $\forall r.A \not\equiv \top$.

Let us now consider the interpretations \mathcal{I}_1 and \mathcal{I}_2 which have the description graphs shown below.

$$\begin{array}{ccc} \mathcal{I}_0 : d : \{\} & \mathcal{I}_1 : d_1 : \{\} & \mathcal{I}_2 : d_2 : \{\} \\ & \downarrow r & \downarrow r \\ & d_3 : \{\} & d_4 : \{A\} \end{array}$$

Then, in addition to $d \in [\forall r.A]^{\mathcal{I}_0}$, it is also the case that $d_2 \in [\forall r.A]^{\mathcal{I}_2}$. Since $\widehat{C} \equiv \forall r.A$, we also have $d_2 \in \widehat{C}^{\mathcal{I}_2}$. This implies that $d \in [E^i_{\sim q_i}]^{\mathcal{I}_0}$ and $d_2 \in [E^i_{\sim q_i}]^{\mathcal{I}_2}$ for each conjunct $E^i_{\sim q_i}$ in \widehat{C} . Furthermore, it is easy to see that Lemma 24 can be applied to obtain for all E^i :

$$\deg^{\mathcal{I}_0}(d, E^i) \leq \deg^{\mathcal{I}_1}(d_1, E^i) \leq \deg^{\mathcal{I}_2}(d_2, E^i)$$

Hence, it is immediate that $d_1 \in [E^i_{\sim q_i}]^{\mathcal{I}_1}$ for all $E^i_{\sim q_i}$ in \widehat{C} , and consequently $d_1 \in \widehat{C}^{\mathcal{I}_1}$. But then, since we assumed $\forall r.A \equiv \widehat{C}$, we have $d_1 \in [\forall r.A]^{\mathcal{I}_1}$ which is a contradiction since obviously this is not the case.

Thus, we have shown that there is no $\tau\mathcal{EL}(deg)$ concept description \widehat{C} such that $\widehat{C} \not\equiv \forall r.A$. \square

This lemma immediately implies that full negation of arbitrary concept descriptions cannot be expressed in $\tau\mathcal{EL}(deg)$. Otherwise, since $\forall r.A \equiv \neg\exists r.\neg A$, there would be a $\tau\mathcal{EL}(deg)$ concept description \widehat{D} such that $\widehat{D} \equiv \neg\exists r.\neg A$ contradicting Lemma 28. Furthermore, since $\exists r.\neg A \equiv \exists r.A_{<1}$, this implies that neither negation of $\tau\mathcal{EL}(deg)$ concept descriptions can be expressed.

Last, for any threshold concept $E_{\sim q}$, it is easy to verify that its negation $\neg E_{\sim q}$ is equivalent to the threshold concept $E_{\gamma(\sim)q}$, where γ is the following mapping:

$$\gamma(<) := \geq \quad \gamma(\leq) := > \quad \gamma(>) := \leq \quad \gamma(\geq) := <$$

5 Reasoning

In this section we investigate the complexity of reasoning problems in $\tau\mathcal{EL}(deg)$. We start with investigating the complexity of terminological reasoning (satisfiability, subsumption), and then turn to assertional reasoning (consistency, instance checking.)

5.1 Terminological reasoning

We start by recalling the two decision problems we will look at:

- *Concept satisfiability*: Let \widehat{C} be a $\tau\mathcal{EL}(deg)$ concept \widehat{C} . The concept \widehat{C} is satisfiable iff there exists an interpretation \mathcal{I} such that $\widehat{C}^{\mathcal{I}} \neq \emptyset$.
- *Subsumption*: Let \widehat{C} and \widehat{D} be two $\tau\mathcal{EL}(deg)$ concept descriptions. \widehat{C} is subsumed by \widehat{D} iff $\widehat{C}^{\mathcal{I}} \subseteq \widehat{D}^{\mathcal{I}}$ for every interpretation \mathcal{I} .

The size $s(\widehat{C})$ of a $\tau\mathcal{EL}(deg)$ concept description \widehat{C} is the number of occurrences of symbols needed to write \widehat{C} .

In contrast to \mathcal{EL} , where every concept description is satisfiable, we have seen in Section 3 that there are unsatisfiable $\tau\mathcal{EL}(deg)$ concept descriptions, such as $A_{\geq 1} \sqcap A_{< 1}$. Thus, the satisfiability problem is non-trivial in $\tau\mathcal{EL}(deg)$. In fact, by a simple reduction from the well-known NP-complete problem ALL-POS ONE-IN-THREE 3SAT (see [12], page 259), we can show that testing $\tau\mathcal{EL}(deg)$ concept descriptions for satisfiability is actually NP-hard.

Definition 29 (ALL-POS ONE-IN-THREE 3SAT). Let U be a set of propositional variables and $\mathcal{C} = \{C_1, \dots, C_n\}$ be a finite set of propositional clauses over U such that:

- Each clause in \mathcal{C} is a set of literals over U which, in addition, contains exactly three literals,
- Each clause in \mathcal{C} is *positive*, i.e., no clause contains a negative 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.

Now, we show how to build a concept description $\widehat{C}_{\mathcal{C}}$ in $\tau\mathcal{EL}(deg)$ such that U has a truth assignment where exactly one literal per clause in \mathcal{C} is true **if, and only if**, $\widehat{C}_{\mathcal{C}}$ is satisfiable. For each propositional variable $u \in U$ we use a concept

name A_u . In addition, to each clause $C_i = (u_{i1}, u_{i2}, u_{i3})$ in \mathcal{C} we associate an \mathcal{EL} -concept description D_i of the form $A_{u_{i1}} \sqcap A_{u_{i2}} \sqcap A_{u_{i3}}$. Then the concept $\widehat{C}_{\mathcal{C}}$ is defined as follows:

$$\widehat{C}_{\mathcal{C}} := \prod_{i=1}^n (D_{i \leq \frac{1}{3}} \sqcap D_{i \geq \frac{1}{3}})$$

The main idea underlying this reduction is that, for any three distinct concept names A_i, A_j, A_k , an individual belongs to $(A_i \sqcap A_j \sqcap A_k)_{\leq 1/3} \sqcap (A_i \sqcap A_j \sqcap A_k)_{\geq 1/3}$ iff it belongs to exactly one of these three concepts.

Lemma 30. *$\widehat{C}_{\mathcal{C}}$ is satisfiable iff there exists a truth assignment to the variables in U such that each clause in \mathcal{C} has exactly one true literal.*

Proof. (\Rightarrow) Assume that $\widehat{C}_{\mathcal{C}}$ is satisfiable. Then, there exists an interpretation \mathcal{I} such that $[\widehat{C}_{\mathcal{C}}]^{\mathcal{I}} \neq \emptyset$ and for some $d \in \Delta^{\mathcal{I}}$ it holds $d \in [\widehat{C}_{\mathcal{C}}]^{\mathcal{I}}$. We construct an assignment for U in the following way:

$$u = \top \text{ iff } d \in [A_u]^{\mathcal{I}}, \text{ for all } u \in U$$

Now, let $C_i = (u_{i1}, u_{i2}, u_{i3})$ be any clause in \mathcal{C} . Since $d \in [\widehat{C}_{\mathcal{C}}]^{\mathcal{I}}$, then $d \in [D_{i \leq \frac{1}{3}}]^{\mathcal{I}}$ and $d \in [D_{i \geq \frac{1}{3}}]^{\mathcal{I}}$. This means that $\text{deg}^{\mathcal{I}}(d, D_i) = \frac{1}{3}$ and by definition of $\text{deg}^{\mathcal{I}}$, d is an instance of exactly one of the concepts $A_{u_{i1}}, A_{u_{i2}}, A_{u_{i3}}$. Thus, only one literal in C_i is assigned to \top .

(\Leftarrow) We assume that there exists a truth assignment to the variables in U such that exactly one literal is true for each clause in \mathcal{C} . Then, we build a *single-pointed* interpretation $\mathcal{I} = (\{d\}, \cdot^{\mathcal{I}})$ such that:

$$d \in [A_u]^{\mathcal{I}} \text{ iff } u = \top, \text{ for all } u \in U$$

By the properties of U with respect to \mathcal{C} , a similar reasoning as before yields $\text{deg}^{\mathcal{I}}(d, D_i) = \frac{1}{3}$ for all i ($1 \leq i \leq n$). Consequently, $d \in [\widehat{C}_{\mathcal{C}}]^{\mathcal{I}}$ and \mathcal{I} satisfies $\widehat{C}_{\mathcal{C}}$. \square

This also yields coNP-hardness for subsumption in $\tau\mathcal{EL}(\text{deg})$ since unsatisfiability can be reduced to subsumption: \widehat{C} is not satisfiable iff $\widehat{C} \sqsubseteq A_{\geq 1} \sqcap A_{< 1}$.

Lemma 31. *In $\tau\mathcal{EL}(\text{deg})$, satisfiability is NP-hard and subsumption is coNP-hard.*

To show an NP upper bound for satisfiability, we use the τ -homomorphism characterization of membership in a $\tau\mathcal{EL}(\text{deg})$ concept description shown in Section 3. Using Theorem 14 we prove a bounded model property for $\tau\mathcal{EL}(\text{deg})$ concept descriptions.

Lemma 32. *Let \widehat{C} be a $\tau\mathcal{EL}(\text{deg})$ concept description of size $\mathfrak{s}(\widehat{C})$. If \widehat{C} is satisfiable, then there exists an interpretation \mathcal{J} such that $\widehat{C}^{\mathcal{J}} \neq \emptyset$ and $|\Delta^{\mathcal{J}}| \leq \mathfrak{s}(\widehat{C})$.*

Proof. Since \widehat{C} is satisfiable, there exists an interpretation \mathcal{I} such that $d \in \widehat{C}^{\mathcal{I}}$ for some $d \in \Delta^{\mathcal{I}}$. Therefore, there exists a τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = d$ (Theorem 14). The idea is to use ϕ and *small* fragments of \mathcal{I} to build \mathcal{J} and a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$, and then apply Theorem 14 to \widehat{C} and \mathcal{J} .

The interpretation \mathcal{J} is built in two steps. We first use as base interpretation \mathcal{I}_0 , the one associated to the description tree $T_{\widehat{C}}$, where we ignore the labels of the form $E_{\sim q}$. It is easy to see that the identity mapping ϕ_{id} is a homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}_0}$. However, this interpretation and homomorphism need not satisfy Condition 2 of Definition 13, i.e., $v \notin [E_{\sim q}]^{\mathcal{I}_0}$ for some $v \in \Delta^{\mathcal{I}_0}$ with $E_{\sim q} \in \widehat{\ell}_{T_{\widehat{C}}}(v)$. To repair this we extend \mathcal{I}_0 to \mathcal{J} by adding appropriate fragments of \mathcal{I} .

More precisley, for a given node v in \mathcal{I}_0 such that $E_{\sim q} \in \widehat{\ell}_{T_{\widehat{C}}}(v)$ we know that $\phi(v) \in [E_{\sim q}]^{\mathcal{I}}$, i.e., $\text{deg}^{\mathcal{I}}(\phi(v), E) \sim q$. By Lemma 25 we do not need all of \mathcal{I} to obtain $\text{deg}^{\mathcal{I}}(\phi(v), E)$ for v in \mathcal{J} . It is sufficient to use the canonical interpretation \mathcal{I}_h induced by a *ptgh* h from T_{E^r} to $G_{\mathcal{I}}$ such that $h(w_0) = \phi(v)$ and $\text{deg}^{\mathcal{I}}(\phi(v), E) = h_w(w_0)$. Here, w_0 is the root of T_{E^r} and we rename it as v for the rest of the proof.

We denote \mathcal{I}_h as \mathcal{I}_v^E and the *ptgh* h which induces \mathcal{I}_h as h_v^E . Let \mathfrak{J} be the family of all interpretations \mathcal{I}_v^E needed to repair the inconsistencies in \mathcal{I}_0 , i.e.,

$$\mathfrak{J} := \{\mathcal{I}_v^E \mid v \in \Delta^{\mathcal{I}_0}, E_{\sim q} \in \widehat{\ell}_{T_{\widehat{C}}}(v) \text{ and } v \notin [E_{\sim q}]^{\mathcal{I}_0}\}$$

We assume each pair $\Delta^{\mathcal{I}_v^{E_1}}$ and $\Delta^{\mathcal{I}_w^{E_2}}$, for $\mathcal{I}_v^{E_1}, \mathcal{I}_w^{E_2} \in \mathfrak{J}$, to be pairwise disjoint in the following sense: if $v \neq w$ they do not have any common element and only share v if $v = w$. In addition, each set $\Delta^{\mathcal{I}_v^E}$ shares only the distinguished element v with $\Delta^{\mathcal{I}_0}$. Then, \mathcal{J} is built in the following way:

- $\Delta^{\mathcal{J}} := \Delta^{\mathcal{I}_0} \cup \bigcup_{\mathcal{K} \in \mathfrak{J}} \Delta^{\mathcal{K}}$,
- $X^{\mathcal{J}} := X^{\mathcal{I}_0} \cup \bigcup_{\mathcal{K} \in \mathfrak{J}} X^{\mathcal{K}}$ for all $X \in (\mathbf{N}_C \cup \mathbf{N}_R)$.

We now show that Condition 2 of Definition 13 is satisfied by ϕ_{id} and \mathcal{J} . Let $E_{\sim q} \in \widehat{\ell}_{T_{\widehat{C}}}(v)$ for some $v \in V_{T_{\widehat{C}}}$. We distinguish two cases:

- $\sim \in \{>, \geq\}$. Suppose that $v \in [E_{\sim q}]^{\mathcal{I}_0}$. Since $\mathcal{I}_0 \subseteq \mathcal{J}$, this makes Lemma 24 to be applicable to $\mathcal{I}_0, \mathcal{J}$ and v . Then we have $\text{deg}^{\mathcal{I}_0}(v, E) \leq \text{deg}^{\mathcal{J}}(v, E)$ and obviously $v \in [E_{\sim q}]^{\mathcal{J}}$. Conversely, suppose that $v \notin [E_{\sim q}]^{\mathcal{I}_0}$. The selection of \mathcal{I}_v^E to build \mathcal{J} and the application of Lemma 25 yields:

$$\text{deg}^{\mathcal{I}_v^E}(v, E) = \text{deg}^{\mathcal{I}}(\phi(v), E)$$

This means that $v \in [E_{\sim q}]^{\mathcal{I}_v^E}$, since $\phi(v) \in [E_{\sim q}]^{\mathcal{I}}$. Note, in addition, that $\mathcal{I}_v^E \subseteq \mathcal{J}$. Then a second application of Lemma 24 yields $v \in [E_{\sim q}]^{\mathcal{J}}$.

- $\sim \in \{<, \leq\}$. Since $\phi(v) \in [E_{\sim q}]^{\mathcal{I}}$, we intend to use again Lemma 24 with respect to \mathcal{J} and \mathcal{I} . For this, we build a mapping φ from $V_{\mathcal{J}}$ to $V_{\mathcal{I}}$ such that $\varphi(v) = \phi(v)$ for all $v \in \Delta^{\mathcal{I}_0}$, and show that it is a homomorphism from $G_{\mathcal{J}}$ to $G_{\mathcal{I}}$.

$$\varphi := \phi \cup \bigcup_{\mathcal{I}_v^E \in \mathfrak{J}} h_v^E$$

One can see that the mapping is defined for all $d \in \Delta^{\mathcal{J}}$ and since ϕ and each h_v^E have $V_{\mathcal{I}}$ as their image, φ is a mapping from $V_{\mathcal{J}}$ to $V_{\mathcal{I}}$. In addition, by the disjointness assumptions made for the construction of $\Delta^{\mathcal{J}}$ and the fact that $\phi(v) = h_v^E(v)$, we have that φ is unambiguous and $\varphi(v) = \phi(v)$ for all $v \in \Delta^{\mathcal{I}_0}$. We now show that it is a homomorphism in the sense of Definition 3:

- Let $w \in \Delta^{\mathcal{J}}$, we either have $w \in \Delta^{\mathcal{I}_0}$ and

$$\ell_{\mathcal{J}}(w) = \ell_{\mathcal{I}_0}(w) \cup \bigcup_{\mathcal{I}_w^E \in \mathfrak{J}} \ell_{\mathcal{I}_w^E}(w)$$

or $w \in \Delta^{\mathcal{I}_v^E}$ for some $\mathcal{I}_v^E \in \mathfrak{J}$ and $\ell_{\mathcal{J}}(w) = \ell_{\mathcal{I}_v^E}(w)$. Since ϕ is a homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$, this means that $\ell_{\mathcal{I}_0}(w) \subseteq \ell_{\mathcal{I}}(\phi(w))$. In addition, by Remark 22 we know that each h_v^E is also a homomorphism from $T_{\mathcal{I}_v^E}$ to $G_{\mathcal{I}}$. Therefore, $\ell_{\mathcal{I}_v^E}(w) \subseteq \ell_{\mathcal{I}}(h_v^E(w))$ for all $\mathcal{I}_v^E \in \mathfrak{J}$ and $w \in \Delta^{\mathcal{I}_v^E}$. Hence, the way φ is defined implies that $\ell_{\mathcal{J}}(w) \subseteq \ell_{\mathcal{I}}(\varphi(w))$ for all $w \in \Delta^{\mathcal{J}}$.

- $v_1 r v_2 \in E_{\mathcal{J}}$. If $v_1, v_2 \in \Delta^{\mathcal{I}_0}$, then ϕ implies that $\varphi(v_1) r \varphi(v_2) \in E_{\mathcal{I}}$. Otherwise, v_1, v_2 are part of the same interpretation \mathcal{I}_v^E and the homomorphism h_v^E implies that $\varphi(v_1) r \varphi(v_2) \in E_{\mathcal{I}}$.

Consequently, φ is a homomorphism from $G_{\mathcal{J}}$ to $G_{\mathcal{I}}$ with $\varphi(v) = \phi(v)$ for all $v \in \Delta^{\mathcal{I}_0}$. Then a further application of Lemma 24 with respect to \mathcal{I} , \mathcal{J} and v yields $v \in [E_{\sim q}]^{\mathcal{J}}$.

Thus, we have shown that ϕ_{id} is τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$. Since $\phi_{id}(v_0) = v_0$, the application of Theorem 14 yields $v_0 \in \widehat{C}^{\mathcal{J}}$.

To conclude, we look at the size of \mathcal{J} . By construction of \mathcal{J} we have:

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

It is not hard to see that the size of \mathcal{I}_0 is bounded by the size of \widehat{C} (without counting the threshold concepts). In addition, each occurrence of a threshold

concept $E_{\sim q}$ in \widehat{C} is considered at most once to build \mathcal{J} . Finally, since the size of the canonical interpretation \mathcal{I}_E^v is bounded by the size of E^r (see Definition 21) and the size of E^r is obviously smaller than the size of E , we have $|\Delta^{\mathcal{I}_E^v}| \leq \mathfrak{s}(E_{\sim q})$. Thus, it is clear that $|\Delta^{\mathcal{J}}| \leq \mathfrak{s}(\widehat{C})$. \square

This lemma yields a standard *guess-and-check* NP-algorithm to decide satisfiability of a concept \widehat{C} : the algorithm first guesses an interpretation \mathcal{J} of size at most $\mathfrak{s}(\widehat{C})$, and then it checks whether there exists a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$. To verify the existence of a τ -homomorphism it uses Algorithm 2 in Section 3.2. Since *deg* can be computed in polynomial time (Section 4), Algorithm 2 checks whether there exists a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$ in polynomial time with respect to *deg*.

Remark 33. We would like to point out that the interpretation \mathcal{J} constructed in the previous lemma is *tree-shaped*, i.e., $G_{\mathcal{J}}$ is a tree. Since $T_{\widehat{C}}$ is a tree, it is also the case for $G_{\mathcal{I}_0}$. In addition, by the disjointness assumptions applied on the canonical interpretations which are used to extend \mathcal{I}_0 into \mathcal{J} and since those are also tree-shaped, it is clear that $G_{\mathcal{J}}$ is a tree.

A co-NP upper bound for subsumption cannot directly be obtained from the fact that satisfiability is in NP. In fact, though we have $\widehat{C} \sqsubseteq \widehat{D}$ iff $\widehat{C} \sqcap \neg \widehat{D}$ is unsatisfiable, this equivalence cannot be used directly since $\neg \widehat{D}$ need not be a $\tau\mathcal{EL}(\text{deg})$ concept description, as shown in Section 4.1. Nevertheless, we can extend the ideas used in the proof of Lemma 32 to obtain a bounded model property for satisfiability of concepts of the form $\widehat{C} \sqcap \neg \widehat{D}$.

Lemma 34. *Let \widehat{C} and \widehat{D} be $\tau\mathcal{EL}(\text{deg})$ concept descriptions of respective sizes $\mathfrak{s}(\widehat{C})$ and $\mathfrak{s}(\widehat{D})$. If $\widehat{C} \sqcap \neg \widehat{D}$ is satisfiable, then there exists an interpretation \mathcal{J} such that $\widehat{C}^{\mathcal{J}} \setminus \widehat{D}^{\mathcal{J}} \neq \emptyset$ and $|\Delta^{\mathcal{J}}| \leq \mathfrak{s}(\widehat{C}) \times \mathfrak{s}(\widehat{D})$.*

Proof. Since $\widehat{C} \sqcap \neg \widehat{D}$ is satisfiable there exists an interpretation \mathcal{I} such that $d \in \widehat{C}^{\mathcal{I}}$ and $d \notin \widehat{D}^{\mathcal{I}}$ for some $d \in \Delta^{\mathcal{I}}$. We first apply the construction used in Lemma 32 to construct, with respect to \mathcal{I} , an interpretation \mathcal{J}_0 such that $\widehat{C}^{\mathcal{J}_0} \neq \emptyset$ and $|\Delta^{\mathcal{J}_0}| \leq \mathfrak{s}(\widehat{C})$. From Lemma 32 we know:

- $G_{\mathcal{J}_0}$ is a tree and $v_0 \in \widehat{C}^{\mathcal{J}_0}$.
- ϕ is a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = d$.
- φ is a homomorphism from $G_{\mathcal{J}_0}$ to $G_{\mathcal{I}}$ with $\varphi(v) = \phi(v)$ for all $v \in V_{T_{\widehat{C}}}$.

Since $\varphi(v_0) \notin \widehat{D}^{\mathcal{I}}$, the idea is to use φ to extract from \mathcal{I} the necessary information to extend \mathcal{J}_0 into an interpretation \mathcal{J} that falsifies \widehat{D} in v_0 , while keeping $v_0 \in \widehat{C}^{\mathcal{J}}$.

For this we consider all the nodes in $\Delta^{\mathcal{J}_0}$ in a top-down manner, starting with the root v_0 .

We compute a sequence of pairs $(\mathcal{J}_0, S_0)(\mathcal{J}_1, S_1) \dots$ where each \mathcal{J}_i is an interpretation and each S_i is a set of pairs of the form (v, \widehat{F}) where $v \in \Delta^{\mathcal{J}_0}$ and \widehat{F} is a $\tau\mathcal{EL}(\text{deg})$ concept description. The initial pair (\mathcal{J}_0, S_0) is defined as $(\mathcal{J}_0, \{(v_0, \widehat{D})\})$. We make this construction such that $\varphi(v) \notin \widehat{F}^{\mathcal{I}}$ represents an invariant for each pair $(v, \widehat{F}) \in S_i$. This will then be used to show that $v \notin \widehat{F}^{\mathcal{J}}$ and hence, $v_0 \notin \widehat{D}^{\mathcal{J}}$.

Each pair (\mathcal{J}_i, S_i) (for $i > 0$) is computed from the pair $(\mathcal{J}_{i-1}, S_{i-1})$ as follows:

- We first compute an auxiliary set S_i^* to decompose concepts \widehat{F} of the form $\widehat{F}_1 \sqcap \dots \sqcap \widehat{F}_n$. For each such concept \widehat{F} we choose one concept \widehat{F}' such that $\widehat{F}' = \widehat{F}_j$ for some j ($1 \leq j \leq n$) and $\varphi(v) \notin [\widehat{F}_j]^{\mathcal{I}}$. The set S_i^* is defined as follows:

$$S_i^* := S_{i-1} \cup \bigcup_{\substack{(v, \widehat{F}) \in S_{i-1} \\ \widehat{F} = \widehat{F}_1 \sqcap \dots \sqcap \widehat{F}_n}} \{(v, \widehat{F}')\}$$

- Then S_i is obtained from S_i^* as:

$$S_i := \{(w, \widehat{F}) \mid (v, \exists r.\widehat{F}) \in S_i^*, (v, w) \in r^{\mathcal{J}_0} \text{ and } w \in \widehat{F}^{\mathcal{J}_0}\}$$

- For each $(v, E_{\sim q}) \in S_i^*$ such that $v \in [E_{\sim q}]^{\mathcal{J}_0}$ we consider the interpretation \mathcal{I}_v^E (see the proof of Lemma 32) with $h_v^E(w_0) = \varphi(v)$. Let \mathfrak{J}_i be the following set:

$$\mathfrak{J}_i := \{\mathcal{I}_v^E \mid (v, E_{\sim q}) \in S_i^* \text{ and } v \in [E_{\sim q}]^{\mathcal{J}_0}\}$$

Using the same disjointness assumptions as in Lemma 32, we build \mathcal{J}_i as follows:

- $\Delta^{\mathcal{J}_i} := \Delta^{\mathcal{J}_{i-1}} \cup \bigcup_{\mathcal{K} \in \mathfrak{J}_i} \Delta^{\mathcal{K}}$,
- $X^{\mathcal{J}_i} := X^{\mathcal{J}_{i-1}} \cup \bigcup_{\mathcal{K} \in \mathfrak{J}_i} X^{\mathcal{K}}$ for all $X \in (\mathbf{N}_C \cup \mathbf{N}_R)$.

Since $G_{\mathcal{J}_0}$ is a tree, this construction considers every node in $\Delta^{\mathcal{J}_0}$ only once in the following sense. A node v does not occur in more than one set S_i ($i \geq 0$). In addition, if $(v, \widehat{F}) \in S_i$, there is no other pair $(v, _)$ occurring in S_i . This implies that the construction terminates for some p where $S_p = \emptyset$. Moreover, one can see that for each $(v, \widehat{F}) \in S_i$ the concept description \widehat{F} is a sub-description of \widehat{D} . Since $|\Delta^{\mathcal{J}_0}|$ is bounded by $\mathfrak{s}(\widehat{C})$ and since at most one canonical interpretation is added for each $v \in \Delta^{\mathcal{J}_0}$ (whose size is bounded by $\mathfrak{s}(\widehat{D})$), we have that $|\Delta^{\mathcal{J}_p}| \leq \mathfrak{s}(\widehat{C}) \times \mathfrak{s}(\widehat{D})$.

We now show that $v_0 \in \widehat{C}^{\mathcal{J}_p}$ and $v_0 \notin \widehat{D}^{\mathcal{J}_p}$. Let us start with $v_0 \in \widehat{C}^{\mathcal{J}_p}$. Consider the mapping φ^* from $V_{\mathcal{J}_p}$ to $V_{\mathcal{I}}$:

$$\varphi^* := \varphi \cup \bigcup_{i=1}^p \bigcup_{\mathcal{I}_v^E \in \mathcal{J}_i} h_v^E$$

One can show that φ^* is homomorphism from $G_{\mathcal{J}_p}$ to $G_{\mathcal{I}}$ with $\varphi^*(v) = \phi(v)$ for all $v \in V_{T_{\widehat{C}}}$. The proof uses the same arguments in Lemma 32 which show that φ is a homomorphism from $G_{\mathcal{J}}$ to $G_{\mathcal{I}}$. The rest is to show that ϕ_{id} is a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}_p}$. Since $\mathcal{J}_0 \subseteq \mathcal{J}_p$ and we have the homomorphism φ^* , we use the same idea in Lemma 32, that shows that ϕ_{id} is τ -homomorphism from $T_{\widehat{C}}$ to \mathcal{J}_0 . Finally, since ϕ_{id} is the identity mapping, we thus have $v_0 \in \widehat{C}^{\mathcal{J}_p}$.

Before going into the main details of why $v_0 \notin \widehat{D}^{\mathcal{J}_p}$, we make clear why the invariant mentioned before is satisfied along the construction of \mathcal{J}_p :

$$(v, \widehat{F}) \in S_i^* \Rightarrow \varphi(v) \notin \widehat{F}^{\mathcal{I}}$$

Note that the initial pair (v_0, \widehat{D}) satisfies it, since $\varphi(v_0) = d$ and $d \notin \widehat{D}^{\mathcal{I}}$. The new pairs added when computing S_i^* clearly satisfy the property. Finally, if $\varphi(v) \notin [\exists r. \widehat{F}]^{\mathcal{I}}$, for any r -successor w of v the homomorphism makes $(\varphi(v), \varphi(w)) \in r^{\mathcal{I}}$ and consequently, $\varphi(w) \notin \widehat{F}^{\mathcal{I}}$.

Now, to prove that $v_0 \notin \widehat{D}^{\mathcal{J}_p}$, we show the following more general claim.

$$\textit{Claim: If } (v, \widehat{F}) \in S_i^* \text{ (} 0 < i \leq p \text{), then } v \notin \widehat{F}^{\mathcal{J}_p}$$

We show the claim by induction on the *role-depth* of \widehat{F} . We will not explicitly consider pairs where \widehat{F} is of the form $\widehat{F}_1 \sqcap \dots \sqcap \widehat{F}_n$. Note that the computation of S_i^* from S_{i-1} and the fact that $\varphi(v) \notin \widehat{F}^{\mathcal{I}}$, ensure that there exists always a pair $(v, \widehat{F}_i) \in S_i^*$. In addition, \widehat{F}_i will fit into one of the two cases treated below, and $v \notin [\widehat{F}_i]^{\mathcal{J}_p}$ implies $v \notin \widehat{F}^{\mathcal{J}_p}$.

Let $(v, \widehat{F}) \in S_i^*$ for some i ($0 < i \leq p$):

- $\text{rd}(\widehat{F}) = 0$. Then \widehat{F} could be of the form \top , $A \in \mathbf{N}_{\mathcal{C}}$ or $E_{\sim q}$. Since $\varphi(v) \notin \widehat{F}^{\mathcal{I}}$ the case $\widehat{F} = \top$ could not happen, and since φ^* is a homomorphism from $G_{\mathcal{J}_p}$ to $G_{\mathcal{I}}$ with $\varphi^*(v) = \varphi(v)$ for all $v \in \Delta^{\mathcal{J}_0}$, we have $v \notin A^{\mathcal{J}_p}$.

If \widehat{F} is of the form $E_{\sim q}$, since $\neg E_{\sim q} \equiv E_{\gamma(\sim)q}$ (see Section 4.1) we know that $\varphi(v) \in [E_{\gamma(\sim)q}]^{\mathcal{I}}$. We do a similar case distinction, as in the proof of Lemma 32, with respect to whether $\gamma(\sim)$ is in $\{>, \geq\}$ or in $\{<, \leq\}$. Note that we have $\mathcal{J}_0 \subseteq \mathcal{J}_p$ and the homomorphism φ^* from $G_{\mathcal{J}_p}$ to $G_{\mathcal{I}}$ with $\varphi^*(v) = \varphi(v)$ for all $v \in \Delta^{\mathcal{J}_0}$. Moreover, whenever $v \in [E_{\sim q}]^{\mathcal{J}_0}$ (equivalently $v \notin [E_{\gamma(\sim)q}]^{\mathcal{J}_0}$), the construction adds an interpretation \mathcal{I}_v^E such that $\text{deg}^{\mathcal{I}_v^E}(v, E) = \text{deg}^{\mathcal{I}}(\varphi(v), E)$. Therefore, we will obtain $v \in [E_{\gamma(\sim)q}]^{\mathcal{J}_k}$ and consequently $v \notin [E_{\sim q}]^{\mathcal{J}_k}$.

- $\text{rd}(\widehat{F}) > 0$. Then \widehat{F} is of the form $\exists r.\widehat{F}'$. Since each node is considered only once in the construction, one can see that each direct r -successor of v in $G_{\mathcal{J}_p}$ is a node in $\Delta^{\mathcal{J}_0}$. Let $w \in \Delta^{\mathcal{J}_0}$ such that $(v, w) \in r^{\mathcal{J}_0}$. We distinguish two cases. If $w \in [\widehat{F}']^{\mathcal{J}_0}$, then $(w, \widehat{F}') \in S_i$ and consequently $(w, \widehat{F}') \in S_{i+1}^*$. Since $\text{rd}(\widehat{F}') < \text{rd}(\exists r.\widehat{F}')$, the application of induction hypothesis yields $w \notin [\widehat{F}']^{\mathcal{J}_p}$.

Suppose now that $w \notin [\widehat{F}']^{\mathcal{J}_0}$. Since $G_{\mathcal{J}_0}$ is a tree, this means that neither w nor any of its successors in $\Delta^{\mathcal{J}_0}$ is considered in the construction of \mathcal{J}_p . Therefore, the reachable elements from w in \mathcal{J}_p are exactly the same as in \mathcal{J}_0 . Assume that $w \in [\widehat{F}']^{\mathcal{J}_p}$, then by Theorem 14 there exists a τ -homomorphism ϕ' from $T_{\widehat{F}}$ to $G_{\mathcal{J}_p}$ with $\phi'(w_0) = w$ (w_0 is the root of $T_{\widehat{F}}$). But then, it would also be a τ -homomorphism from $T_{\widehat{F}}$ to $G_{\mathcal{J}_0}$ contradicting $w \notin [\widehat{F}']^{\mathcal{J}_0}$. Consequently $w \notin [\widehat{F}']^{\mathcal{J}_p}$.

In conclusion, we have that for any $w \in \Delta^{\mathcal{J}_p}$ such that $(v, w) \in r^{\mathcal{J}_p}$ it is the case that $w \notin [\widehat{F}']^{\mathcal{J}_p}$. Hence, $v \notin [\exists r.\widehat{F}']^{\mathcal{J}_p}$.

Since $(v_0, \widehat{D}) \in S_1^*$, we have shown that $v_0 \notin \widehat{D}^{\mathcal{J}_p}$. Overall we have $[\widehat{C} \sqcap \neg \widehat{D}]^{\mathcal{J}_p} \neq \emptyset$ and $|\Delta|^{\mathcal{J}_p} \leq \mathfrak{s}(\widehat{C}) \times \mathfrak{s}(\widehat{D})$. Thus, \mathcal{J}_p is the interpretation \mathcal{J} that satisfies our main claim. \square

The lemma yields an obvious guess-and-check NP-algorithm for non-subsumption, which shows that subsumption is in co-NP. Like for the satisfiability problem, the algorithm guesses an interpretation \mathcal{J} of size $\mathfrak{s}(\widehat{C}) \times \mathfrak{s}(\widehat{D})$, and then checks if $d \in \widehat{C}^{\mathcal{J}}$ and $d \notin \widehat{D}^{\mathcal{J}}$ for some element $d \in \Delta^{\mathcal{J}}$. This can obviously be done, in polynomial time, by using Algorithm 2.

Overall, we thus have shown:

Theorem 35. *In $\tau\mathcal{EL}(\text{deg})$, satisfiability is NP-complete and subsumption is coNP-complete.*

5.2 Assertional reasoning

In this section we consider the following two decision problems.

- *ABox consistency:* Let \mathcal{A} be a $\tau\mathcal{EL}(\text{deg})$ ABox. The ABox \mathcal{A} is consistent iff there exists an interpretation \mathcal{I} which is a model of \mathcal{A} (denoted $\mathcal{I} \models \mathcal{A}$).
- *Instance checking:* Let \mathcal{A} be $\tau\mathcal{EL}(\text{deg})$ ABox, \widehat{C} a $\tau\mathcal{EL}(\text{deg})$ concept description and a an individual. The individual a is an instance of \widehat{C} in \mathcal{A} (denoted $\mathcal{A} \models \widehat{C}(a)$) iff $a^{\mathcal{I}} \in \widehat{C}^{\mathcal{I}}$ holds in all models of \mathcal{A} .

We define the size $s(\mathcal{A})$ of an ABox \mathcal{A} as:

$$s(\mathcal{A}) := \sum_{\substack{\widehat{C}(a) \in \mathcal{A} \\ a \in \text{Ind}(\mathcal{A})}} s(\widehat{C}) + \sum_{\substack{r(a,b) \in \mathcal{A} \\ a,b \in \text{Ind}(\mathcal{A})}} 1$$

Since satisfiability can obviously be reduced to consistency (\widehat{C} is satisfiable iff $\{\widehat{C}(a)\}$ is consistent), and subsumption to the instance problem ($\widehat{C} \sqsubseteq \widehat{D}$ iff $\{\widehat{C}(a)\} \models \widehat{D}(a)$), the lower bounds shown above also hold for assertional reasoning.

Lemma 36. *In $\tau\mathcal{EL}(\text{deg})$, ABox consistency is NP-hard and instance checking is coNP-hard.*

Regarding upper bounds, we proceed in the same way as for concept satisfiability and subsumption. We first show a bounded model property for consistent ABoxes, which yields an NP-upper bound for ABox consistency. Then, similar to our treatment of subsumption, this bounded model can then be used to obtain a bounded model property for the complement of the instance problem (a is not an instance of \widehat{C} in \mathcal{A}). However, as we will show, the bound of the model has the size of \widehat{C} in the exponent. For this reason, we obtain a coNP upper bound for the instance problem only if we consider *data complexity* [8], where the size of the query concept \widehat{C} is assumed to be constant.

The consistency problem can be tackled in a similar way as the satisfiability problem. As we have shown in Section 3, based on the translation given in [15], $\tau\mathcal{EL}(\text{deg})$ ABoxes can be translated into $\tau\mathcal{EL}(\text{deg})$ description graphs and consistency can be characterized using τ -homomorphisms (see Theorem 15). We use this characterization to prove an appropriate bounded model property with a polynomial bound, in a similar way as the satisfiability problem.

Lemma 37. *Let \mathcal{A} be an ABox in $\tau\mathcal{EL}(\text{deg})$ of size $s(\mathcal{A})$. If \mathcal{A} is consistent, then there exists an interpretation \mathcal{I} such that $\mathcal{I} \models \mathcal{A}$ and $|\Delta^{\mathcal{I}}| \leq s(\mathcal{A})$.*

Proof. Assume that \mathcal{A} is consistent, then there exists an interpretation \mathcal{I} such that $\mathcal{I} \models \mathcal{A}$. Therefore, there exists a τ -homomorphism ϕ from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}}$ such that $\phi(a) = a^{\mathcal{I}}$ for all $a \in \text{Ind}(\mathcal{A})$ (Theorem 15).

We proceed in the same way as in Lemma 32. The base interpretation \mathcal{I}_0 is the one having the description graph $\widehat{G}(\mathcal{A})$, where we ignore the labels of the form $E_{\sim q}$. Again, the identity mapping ϕ_{id} is a homomorphism from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}_0}$, but need not satisfy Condition 2 of Definition 13. Then, we extend \mathcal{I}_0 into \mathcal{I} using the same procedure in Lemma 32. In addition, we also have a homomorphism φ from $G_{\mathcal{I}}$ to $G_{\mathcal{I}}$ with $\varphi(v) = \phi(v)$ for all $v \in V_{\mathcal{A}}$.

Hence, the same arguments can be used to show that ϕ_{id} is a τ -homomorphism from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{J}}$. If, in addition, we make $a^{\mathcal{J}} = a$ we then have $\phi_{id}(a) = a^{\mathcal{J}}$ for all $a \in \text{Ind}(\mathcal{A})$, and the application of Theorem 15 yields $\mathcal{J} \models \mathcal{A}$.

From the construction of $\widehat{G}(\mathcal{A})$ it is not difficult to see that, like in Lemma 32, the size of \mathcal{I}_0 is bounded by the size of \mathcal{A} without counting the threshold concepts. Furthermore, each threshold concept $E_{\sim q}$ occurring in a concept assertion in \mathcal{A} is also used at most once to build \mathcal{J} . Therefore, the same arguments given in Lemma 32 yield $|\Delta^{\mathcal{J}}| \leq s(\mathcal{A})$. \square

Similar as for concept satisfiability, this lemma yields an NP-algorithm to decide the consistency problem: first it guesses an interpretation \mathcal{J} of size at most $s(\mathcal{A})$. Then, it checks (in polynomial time) using Algorithm 3 whether there exists a τ -homomorphism ϕ from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{J}}$ with $\phi(a) = a^{\mathcal{J}}$ for all $a \in \text{Ind}(\mathcal{A})$.

As mentioned before, we use this bounded model property to obtain a bounded model property for non-instance, i.e., a is not an instance of \widehat{C} in \mathcal{A} iff $\mathcal{A} \cup \{\neg\widehat{C}(a)\}$ is consistent. However, different from the interpretation \mathcal{J}_0 used in the construction of Lemma 34, the bounded model of an ABox obtained in Lemma 37 does not necessarily have a tree shape. This means that the procedure described in Lemma 34 to construct \mathcal{J} would consider nodes from $\Delta^{\mathcal{J}_0}$ more than one time.

Example 38. Let E be the \mathcal{EL} concept description $\exists r.A \sqcap \exists r.B$. Consider the following ABox \mathcal{A} and $\tau\mathcal{EL}(deg)$ concept description \widehat{C} :

$$\mathcal{A} := \{r(a, a)\} \text{ and } \widehat{C} := \underbrace{\exists r \dots r}_p.E_{<1}$$

Obviously a is not an instance of \widehat{C} in \mathcal{A} . The *single-pointed* interpretation $\mathcal{I} = (\{d\}, \cdot^{\mathcal{I}})$ with $a^{\mathcal{I}} = d$, $(d, d) \in r^{\mathcal{I}}$ and $d \in A^{\mathcal{I}} \cap B^{\mathcal{I}}$, is a model of \mathcal{A} which does not satisfy $\widehat{C}(a)$.

Now, if we try to adapt the construction in Lemma 34, it would start with \mathcal{J}_0 as the bounded model of \mathcal{A} obtained in Lemma 37, which is actually \mathcal{I} with $d \notin A^{\mathcal{I}} \cup B^{\mathcal{I}}$. The difference is that since $G_{\mathcal{J}_0}$ is not a tree, although the same procedure will still terminate, it will generate a sequence of sets S_0, \dots, S_{p+1} with $S_i = \{(d, \underbrace{\exists r \dots r}_{p-i}.E_{<1})\}$ for all $i \leq p$ and $S_{p+1} = \emptyset$.

Since $d \in [E_{<1}]^{\mathcal{J}_0}$ and $\varphi(d) \in [E_{\geq 1}]^{\mathcal{I}}$, \mathcal{J}_0 is extended by adding a canonical interpretation which has the same description tree as E . This will ensure that $d \notin [E_{<1}]^{\mathcal{J}_p}$, however, \mathcal{J}_p introduces two new r -successors of d and as “asserted” in S_{p-1} it must also hold $d \notin [\exists r.E_{<1}]^{\mathcal{J}_p}$, which is clearly not the case. To repair this S_{p-1} has to be analysed, in addition, with respect to the newly added elements. Note that after fixing the problem for S_{p-1} , the same issue will arise with respect to $(d, \exists rr.E_{<1}) \in S_{p-2}$ and so on. Therefore, whenever a node v requires the addition of a canonical interpretation and has additional constraints

(as just explained), the same idea needs to be recursively applied with respect to its new successors and those constraints.

Finally, one can see in our example, that this recursive application of the procedure leads to a final model whose size is exponential in the size of \widehat{C} . This, however, does not necessarily imply that this is the best bound we can hope for. In fact, as mentioned above, \mathcal{I} is a very small model satisfying $\mathcal{A} \cup \{\neg\widehat{C}(a)\}$, but the procedure does not realize that d can be an instance of A and B in \mathcal{J}_0 without contradicting $\mathcal{J}_0 \models \mathcal{A}$. We do not yet know whether there is a better bound which applies to all possible combinations of \mathcal{A} and \widehat{C} .

We use the intuition given in Example 38, to extend the idea in Lemma 34 to consistency of an ABox $\mathcal{A} \cup \{\neg\widehat{C}(a)\}$. From now on, we will use ABoxes that besides assertions of the form $\widehat{C}(a)$ may also contain assertions of the form $\neg\widehat{C}(a)$. In case we want to refer to an ABox strictly in $\tau\mathcal{EL}(deg)$ we will mention it explicitly.

We now introduce a set of rules to transform an ABox $\mathcal{A} \cup \{\neg\widehat{C}(a)\}$ into an equivalent ABox \mathcal{A}' , with the purpose to unfold the necessary information to falsify $\widehat{C}(a)$ in a model of \mathcal{A} , as it is done for $\widehat{C} \sqcap \neg\widehat{D}$ in the proof of Lemma 34. They are, in addition, similar some of the *pre-processing* rules defined in [3, 13], with the addition of specific rules to deal with the negation of threshold concepts.

Definition 39 (pre-processing rules). Let \mathcal{A} be an ABox. We define the following pre-processing rules:

- $\mathcal{A} \rightarrow_{\neg\sqcap} \mathcal{A} \cup \{\neg\widehat{D}(a)\}$
if $\neg\widehat{C}(a) \in \mathcal{A}$ with \widehat{C} of the form $\widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_n$, $\neg\widehat{C}_i(a) \notin \mathcal{A}$ for all $i \in \{1 \dots n\}$
and $\widehat{D} = \widehat{C}_i$ for some $i \in \{1 \dots n\}$.
- $\mathcal{A} \rightarrow_{\neg\exists} \mathcal{A} \cup \{\neg\widehat{D}(b)\}$
if $(\neg\exists r.\widehat{D})(a) \in \mathcal{A}$, $r(a, b) \in \mathcal{A}$ and $\neg\widehat{D}(b) \notin \mathcal{A}$.
- $\mathcal{A} \rightarrow_{\neg\sim} \mathcal{A} \cup \{E_{\gamma(\sim)q}(a)\}$
if $\neg E_{\sim q}(a) \in \mathcal{A}$ and $E_{\gamma(\sim)q}(a) \notin \mathcal{A}$.
- $\mathcal{A} \rightarrow_{\neg A} \mathcal{A} \cup \{A_{<1}(a)\}$
if $A \in \mathbf{N}_{\mathbf{C}}$, $\neg A(a) \in \mathcal{A}$ and $A_{<1}(a) \notin \mathcal{A}$.

A *pre-processing* of \mathcal{A} is an ABox \mathcal{A}' obtained by a sequence of rule applications such that no further rule application is possible over \mathcal{A}' . Note that if \mathcal{A} is a $\tau\mathcal{EL}(deg)$ ABox, the unique *pre-processing* of \mathcal{A} is \mathcal{A} . The last two rules are motivated by the fact that $\neg E_{\sim q} \equiv E_{\gamma(\sim)q}$ and $\neg A \equiv A_{<1}$ (see Section 4.1 and Section 3).

One can see that no rule application introduces either a new individual or a new role assertion. Therefore, we have the same set of individuals and role assertions

in \mathcal{A} and \mathcal{A}' . In addition, each application of a rule introduces either an assertion of the form $\neg\widehat{C}(a)$, $E_{\gamma(\sim)_q}(a)$ or $A_{<1}(a)$. In the first case, \widehat{C} is a *sub-description* of some concept \widehat{D} such that $\neg\widehat{D}(b)$ is an assertion initially in \mathcal{A} , whereas no further rule is applicable to the other two cases. Hence, since \mathcal{A} is finite, we will never have an infinite sequence of rule applications. Finally, we can prove the following proposition (see the Appendix).

Proposition 40. *Let \mathcal{A} be an ABox. Then, \mathcal{A} is consistent iff there exists a consistent pre-processing \mathcal{A}' of \mathcal{A} .*

As a direct consequence from the proof of the previous proposition we have the following remark.

Remark 41. Let \mathcal{A} be an ABox and \mathcal{I} an interpretation. If $\mathcal{I} \models \mathcal{A}$, then there exists a pre-processing \mathcal{A}' of \mathcal{A} such that $\mathcal{I} \models \mathcal{A}'$.

We are now ready to show a bounded model property for consistent ABoxes of the form $\mathcal{A} \cup \{\neg\widehat{C}(a)\}$. Before going into the details it would be useful to introduce some notation. An ABox which contains only one individual name and no role assertions is called a *single-element* ABox. In addition, given an ABox \mathcal{A} , the ABox $\mathcal{A}(a)$ consists of all the concept assertions $\widehat{C}(a)$ or $\neg\widehat{C}(a)$ occurring in \mathcal{A} . Furthermore, \mathcal{A}^+ is defined as:

$$\mathcal{A}^+ := \bigcup_{\substack{\widehat{C}(a) \in \mathcal{A} \\ a \in \text{Ind}(\mathcal{A})}} \{\widehat{C}(a)\} \cup \bigcup_{\substack{r(a,b) \in \mathcal{A} \\ a,b \in \text{Ind}(\mathcal{A})}} \{r(a,b)\}$$

and \mathcal{A}^- is defined as:

$$\mathcal{A}^- := \bigcup_{\substack{\neg\widehat{C}(a) \in \mathcal{A} \\ a \in \text{Ind}(\mathcal{A})}} \{\neg\widehat{C}(a)\}$$

The idea is the following, for a consistent ABox \mathcal{A} and an arbitrary interpretation \mathcal{I} satisfying \mathcal{A} , we first consider a *pre-processing* \mathcal{A}' of \mathcal{A} such that $\mathcal{I} \models \mathcal{A}'$. In particular, we look at each of the ABoxes $\mathcal{A}'(a)$ for each $a \in \text{Ind}(\mathcal{A})$. Even when those contain assertions over negated concepts they are, nevertheless, simpler than \mathcal{A} in the sense that only contain one individual name and do not have role assertions. We then show how to provide a model of bounded size for this particular kind of ABox using the following lemma, whose proof is deferred to the Appendix.

Lemma 42. *Let \mathcal{A} be a single-element ABox and \mathcal{I} an interpretation such that $\mathcal{I} \models \mathcal{A}$. In addition, let \mathcal{J} be the bounded model of \mathcal{A}^+ obtained in Lemma 37. Then, there exists a tree-shaped interpretation \mathcal{K} such that:*

1. $\mathcal{K} \models \mathcal{A}$,

2. there exists a homomorphism φ from $G_{\mathcal{K}}$ to $G_{\mathcal{I}}$ such that $\varphi(a^{\mathcal{K}}) = a^{\mathcal{I}}$, and
3. $|\Delta^{\mathcal{K}}| \leq |\Delta^{\mathcal{J}}| \times n$, where:

$$n := \begin{cases} 1, & \text{if } \mathcal{A}^- = \emptyset \\ \prod_{-\widehat{C}(a) \in \mathcal{A}^-} s(\widehat{C}), & \text{otherwise.} \end{cases}$$

The rest is to combine all those models into a model of \mathcal{A}' of bounded size. More precisely we show that the disjoint union of all those models combined with the role assertions in \mathcal{A} gives the wanted model. This is formalize in the following lemma (see the Appendix for its proof).

Lemma 43. *Let \mathcal{A} be an ABox and \mathcal{I} an interpretation such that $\mathcal{I} \models \mathcal{A}$. In addition, let \mathcal{A}' be a pre-processing of \mathcal{A} such that $\mathcal{I} \models \mathcal{A}'$. For each $a \in \text{Ind}(\mathcal{A})$, let \mathcal{I}_a be a tree-shaped interpretation such that:*

- $\mathcal{I}_a \models \mathcal{A}'(a)$,
- there exists a homomorphism φ_a from $G_{\mathcal{I}_a}$ to $G_{\mathcal{I}}$ with $\varphi_a(a^{\mathcal{I}_a}) = a^{\mathcal{I}}$.

Last, let \mathcal{J} be the following interpretation:

- $\Delta^{\mathcal{J}} := \bigcup_{a \in \text{Ind}(\mathcal{A})} \Delta^{\mathcal{I}_a}$,
- $A^{\mathcal{J}} := \bigcup_{a \in \text{Ind}(\mathcal{A})} A^{\mathcal{I}_a}$ for all $A \in \mathbf{N}_{\mathbf{C}}$,
- $r^{\mathcal{J}} := \{a^{\mathcal{I}_a} r b^{\mathcal{I}_b} \mid r(a, b) \in \mathcal{A}\} \cup \bigcup_{a \in \text{Ind}(\mathcal{A})} r^{\mathcal{I}_a}$ for all $r \in \mathbf{N}_{\mathbf{R}}$, and
- $a^{\mathcal{J}} := a^{\mathcal{I}_a}$, for all $a \in \text{Ind}(\mathcal{A})$.

where the sets $\Delta^{\mathcal{I}_a}$ are pair-wise disjoint. Then, $\mathcal{J} \models \mathcal{A}$.

Using these two lemmas we obtain the bounded model property for consistent ABoxes of the form $\mathcal{A} \cup \{\neg \widehat{C}(a)\}$. Recall from Section 2 that $\text{sub}(\widehat{C})$ denotes the set of sub-descriptions of a concept description \widehat{C} .

Lemma 44. *Let \mathcal{A} be an ABox in $\tau\mathcal{EL}(\text{deg})$ of size $s(\mathcal{A})$, \widehat{C} a $\tau\mathcal{EL}(\text{deg})$ concept description of size $s(\widehat{C})$ and $a \in \mathbf{N}_{\mathbf{I}}$. If $\mathcal{A} \cup \{\neg \widehat{C}(a)\}$ is consistent, then there exists an interpretation \mathcal{J} such that:*

1. $\mathcal{J} \models \mathcal{A} \cup \{\neg \widehat{C}(a)\}$, and
2. $|\Delta^{\mathcal{J}}| \leq s(\mathcal{A}) \times [s(\widehat{C})]^u$, where $u = |\text{sub}(\widehat{C})|$.

Proof. Let \mathcal{I} be an interpretation such that $\mathcal{I} \models \mathcal{A} \cup \{-\widehat{C}(a)\}$. By Remark 41 there exists a *pre-processing* \mathcal{A}' of $\mathcal{A} \cup \{-\widehat{C}(a)\}$ such that $\mathcal{I} \models \mathcal{A}'$. For each $a \in \text{Ind}(\mathcal{A})$ we apply Lemma 42 with respect to $\mathcal{A}'(a)$, to obtain a *tree-shaped* interpretation \mathcal{I}_a such that:

- $\mathcal{I}_a \models \mathcal{A}'(a)$,
- there exists a homomorphism φ_a from $G_{\mathcal{I}_a}$ to $G_{\mathcal{I}}$ with $\varphi_a(a^{\mathcal{I}_a}) = a^{\mathcal{I}}$.

Then, we can apply Lemma 43 to obtain an interpretation \mathcal{J} such that:

$$\mathcal{J} \models \mathcal{A} \text{ and } \Delta^{\mathcal{J}} = \bigcup_{a \in \text{Ind}(\mathcal{A})} \Delta^{\mathcal{I}_a}$$

We now look at the size of \mathcal{J} . For each $a \in \text{Ind}(\mathcal{A})$, let \mathcal{J}_a be the bounded model of $\mathcal{A}'^+(a)$ obtained in Lemma 37. The construction of \mathcal{I}_a in Lemma 42 yields:

$$|\Delta^{\mathcal{I}_a}| \leq |\Delta^{\mathcal{J}_a}| \times \prod_{\neg \widehat{D}(a) \in \mathcal{A}'^-(a)} s(\widehat{D}) \quad (1)$$

One can see that each assertion in $\mathcal{A}'^+(a)$ is either of the form $\widehat{C}(a) \in \mathcal{A}(a)$ or $E_{\gamma(\sim)q}$, the latter case obtained by applying the rule $\rightarrow_{\neg \sim}$ or the rule $\rightarrow_{\neg A}$. For the rule $\rightarrow_{\neg A}$, we consider $A_{<1}$ as $E_{\gamma(\sim)q}$, since it is obtained from $\neg A$ and $A \equiv A_{\geq 1}$. In Lemma 36, the interpretation \mathcal{J}_a is built starting with the interpretation \mathcal{I}_0 which have the description graph $\widehat{G}(\mathcal{A}'^+(a))$ (without threshold concepts), and it is then extended by considering each occurrence of a threshold concept $E_{\sim q}$ in $\widehat{G}(\mathcal{A}'^+(a))$. Since the only assertions of the form $\widehat{C}(a)$ in $\mathcal{A}'^+(a)$ are from $\mathcal{A}(a)$, one can easily see that $|V_{\mathcal{A}(a)}| = |V_{\mathcal{A}'^+(a)}|$. Thus, we obtain:

$$|\Delta^{\mathcal{J}_a}| \leq |V_{\mathcal{A}(a)}| + \sum_{E_{\sim q} \in \widehat{G}(\mathcal{A}(a))} s(E_{\sim q}) + \sum_{E_{\gamma(\sim)q}(a) \notin \widehat{G}(\mathcal{A}(a))} s(E_{\gamma(\sim)q})$$

Note that the partial sum of the first two elements in the right-hand side of the inequality is actually bounded by the size of $\mathcal{A}(a)$. In addition, since $s(E_{\gamma(\sim)q}) > 1$ we further have:

$$|\Delta^{\mathcal{J}_a}| \leq s(\mathcal{A}(a)) \times \prod_{E_{\gamma(\sim)q}(a) \notin \widehat{G}(\mathcal{A}(a))} s(E_{\gamma(\sim)q}) \quad (2)$$

It is not hard to see that for each concept assertion $\neg \widehat{D}(a) \in \mathcal{A}'^-(a)$, the concept \widehat{D} is a sub-description of \widehat{C} . In addition, for each $E_{\gamma(\sim)q}(a) \notin \widehat{G}(\mathcal{A}(a))$, the threshold concept assertion $E_{\gamma(\sim)q}(a)$ is obtained after applying the rule $\rightarrow_{\neg \sim}$ ($\rightarrow_{\neg A}$), to an assertion of the form $\neg E_{\sim q}(a)$ ($\neg A(a)$), with $E_{\sim q}(A)$ being also a sub-description of \widehat{C} . Thus, we combine inequalities (1) and (2) to obtain:

$$|\Delta^{\mathcal{I}_a}| \leq s(\mathcal{A}(a)) \times [s(\widehat{C})]^u$$

Thus, since $\sum_{a \in \text{Ind}(\mathcal{A})} \mathfrak{s}(\mathcal{A}(a))$ is obviously bounded by $\mathfrak{s}(\mathcal{A})$, we finally have:

$$|\Delta^{\mathcal{J}}| = \sum_{a \in \text{Ind}(\mathcal{A})} |\Delta^{\mathcal{I}_a}| \leq \left(\sum_{a \in \text{Ind}(\mathcal{A})} \mathfrak{s}(\mathcal{A}(a)) \right) \times [\mathfrak{s}(\widehat{C})]^u \leq \mathfrak{s}(\mathcal{A}) \times [\mathfrak{s}(\widehat{C})]^u$$

□

Using this bounded model property, we can obtain a non-deterministic procedure to decide consistency of an ABox of the form $\mathcal{A} \cup \{\neg \widehat{C}(a)\}$:

1. Guess an interpretation \mathcal{J} of size at most $\mathfrak{s}(\mathcal{A}) \times [\mathfrak{s}(\widehat{C})]^u$.
2. Check whether $\mathcal{J} \models \mathcal{A}$. Like for the consistency problem, this can be done in polynomial time using Algorithm 3. If it is not the case, then the algorithm answers “no”. Otherwise, it remains to verify whether $a^{\mathcal{J}} \notin \widehat{C}^{\mathcal{J}}$.
3. To verify $a^{\mathcal{J}} \notin \widehat{C}^{\mathcal{J}}$, by Theorem 14 it is enough to check that there exists no τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$ with $\phi(v_0) = a^{\mathcal{J}}$. This can also be checked in polynomial time by using Algorithm 2 with respect to *deg*. If there is no such τ -homomorphism the algorithm answers “yes”, and “no” otherwise.

Now, if the size of \widehat{C} is considered as a constant, this algorithm becomes an NP-procedure for consistency of $\mathcal{A} \cup \{\neg \widehat{C}(a)\}$ and consequently, a coNP-procedure to decide instance checking with respect to data complexity. Altogether, we thus have shown:

Theorem 45. *In $\tau\mathcal{EL}(\text{deg})$, consistency is NP-complete, and instance checking is coNP-complete w.r.t. data complexity.*

The instance problem becomes simpler if we consider only \mathcal{EL} ABoxes and *positive $\tau\mathcal{EL}(\text{deg})$ concept descriptions*, i.e., concept descriptions \widehat{C} that only contain threshold concepts of the form $E_{\geq t}$ or $E_{> t}$. Basically, given an \mathcal{EL} ABox, a positive $\tau\mathcal{EL}(\text{deg})$ concept description \widehat{C} , and an individual a , one considers the interpretation \mathcal{I} corresponding to the description graph of \mathcal{A} , and then checks whether there is a τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = a$. The following lemma supports the previous idea.

Lemma 46. *Let \mathcal{A} be an \mathcal{EL} ABox, $a \in \text{Ind}(\mathcal{A})$ and \widehat{C} a positive $\tau\mathcal{EL}(\text{deg})$ concept description. Additionally, let $\mathcal{I}_{\mathcal{A}}$ be the interpretation corresponding to the description graph $G(\mathcal{A})$ with $a^{\mathcal{I}_{\mathcal{A}}} = a$ for all $a \in \text{Ind}(\mathcal{A})$. Then, the following statements are equivalent:*

1. $\mathcal{A} \models \widehat{C}(a)$, and

2. $a \in [\widehat{C}]^{\mathcal{I}_A}$.

Proof. (1 \Rightarrow 2) Assume that $\mathcal{A} \models \widehat{C}(a)$. Then, for every model \mathcal{I} of \mathcal{A} we have $a^{\mathcal{I}} \in [\widehat{C}]^{\mathcal{I}}$. Since \mathcal{I}_A is obviously a model of \mathcal{A} and $a^{\mathcal{I}_A} = a$, this means that $a \in [\widehat{C}]^{\mathcal{I}_A}$.

(2 \Rightarrow 1) Assume that $a \in [\widehat{C}]^{\mathcal{I}_A}$. The characterization for membership in $\tau\mathcal{EL}(deg)$, given in Theorem 14, yields a τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G(\mathcal{A})$ with $\phi(v_0) = a$. Now, consider any model \mathcal{I} of \mathcal{A} . The application of Theorem 15 yields the existence of a τ -homomorphism φ from $G(\mathcal{A})$ to $G_{\mathcal{I}}$ such that $\varphi(a) = a^{\mathcal{I}}$ for all $a \in \text{Ind}(\mathcal{A})$. We then show that the mapping $\varphi \circ \phi$ is a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$:

- From ϕ we know that $\ell_{T_{\widehat{C}}}(v) \subseteq \ell_{\mathcal{A}}(\phi(v))$, for all $v \in V_{T_{\widehat{C}}}$. Similarly, φ implies that $\ell_{\mathcal{A}}(a) \subseteq \ell_{\mathcal{I}}(\varphi(a))$, for all $a \in V_{\mathcal{A}}$. Hence, $\ell_{T_{\widehat{C}}}(v) \subseteq \ell_{\mathcal{I}}((\varphi \circ \phi)(v))$, for all $v \in V_{T_{\widehat{C}}}$. The edge preserving relation can be verified in a similar way.
- Let $v \in V_{T_{\widehat{C}}}$ and $E_{\sim_q} \in \widehat{\ell}_{T_{\widehat{C}}}(v)$. Since ϕ is a τ -homomorphism, this means that $\phi(v) \in [E_{\sim_q}]^{\mathcal{A}}$. Furthermore, the application of Lemma 24 to \mathcal{I}_A , \mathcal{J} and φ yields:

$$deg^{\mathcal{I}_A}(\phi(v), E) \leq deg^{\mathcal{I}}(\varphi(\phi(v)), E)$$

Then, since \widehat{C} is positive, this means that \sim is either $>$ or \geq . Consequently, $(\varphi \circ \phi)(v) \in [E_{\sim_q}]^{\mathcal{I}}$.

Hence, $\varphi \circ \phi$ is a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$. Since $(\varphi \circ \phi)(v_0) = a^{\mathcal{I}}$, then $a^{\mathcal{I}} \in [\widehat{C}]^{\mathcal{I}}$.

Overall, we have shown $a^{\mathcal{I}} \in [\widehat{C}]^{\mathcal{I}}$ for all models \mathcal{I} of \mathcal{A} . Thus, $\mathcal{A} \models \widehat{C}(a)$. \square

Then, since \mathcal{I}_A is linear on the size of \mathcal{A} , checking whether $a \in [\widehat{C}]^{\mathcal{I}_A}$ can be done in polynomial time in the size of \mathcal{A} and \widehat{C} by using Algorithm 2. Therefore, we obtain the following proposition.

Proposition 47. *For positive $\tau\mathcal{EL}(deg)$ concept descriptions and \mathcal{EL} ABoxes, the instance checking problem can be decided in polynomial time.*

6 Concept similarity and relaxed instance queries

In this section we compare our graded membership function with similarity measures on \mathcal{EL} concept descriptions.

We first show how to use the relaxed instance query approach from [9] to turn a concept similarity measure (CSM) \bowtie into a membership degree function m_{\bowtie} . Such a membership degree function, however, need not be well-defined. We present two properties that when satisfied by \bowtie , are sufficient to obtain well-definedness for m_{\bowtie} . Additionally, we show that the relaxed instance queries from [9] can be expressed as instance queries w.r.t. threshold concepts of the form $C_{>t}$.

Next, we present the framework *simi* introduced in [16], which can be used to define a variety of CSMs. We further show that a particular instance \bowtie^1 of this framework turns out to be equivalent to our membership degree function *deg*.

6.1 Defining membership degree functions

In its most general form, a concept similarity measure (CSM) \bowtie is a function that maps each pair of concepts C, D (of a given DL) to a value $C \bowtie D \in [0, 1]$ such that $C \bowtie C = 1$. Intuitively, the higher the value of $C \bowtie D$ is, the more similar the two concepts are supposed to be. Such measures can in principle be defined for arbitrary DLs, but here we restrict the attention to CSMs between \mathcal{EL} concepts, i.e., a CSM is a mapping $\bowtie : \mathcal{C}_{\mathcal{EL}} \times \mathcal{C}_{\mathcal{EL}} \rightarrow [0, 1]$.

Ecke et al. [9, 10] use CSMs to relax instance queries, i.e., instead of requiring that an individual is an instance of the query concept, they only require that it is an instance of a concept that is “similar enough” to the query concept.

Definition 48 ([9, 10]). Let \bowtie be a CSM, \mathcal{A} an \mathcal{EL} ABox, and $t \in [0, 1)$. The individual $a \in \mathbb{N}_I$ is a *relaxed instance* of the \mathcal{EL} query concept Q w.r.t. \mathcal{A} , \bowtie , and the threshold t iff there exists an \mathcal{EL} concept description X such that $Q \bowtie X > t$ and $\mathcal{A} \models X(a)$. The set of all individuals occurring in \mathcal{A} that are relaxed instances of Q w.r.t. \mathcal{A} , \bowtie , and t is denoted by $\text{Relax}_{\bowtie}^t(Q, \mathcal{A})$.

We apply the same idea on the semantic level of an interpretation rather than the ABox level to obtain graded membership functions from similarity measures.

Definition 49. Let \bowtie be a CSM. Then, for each interpretation \mathcal{I} , we define the function $m_{\bowtie}^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \rightarrow [0, 1]$ as

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

For an arbitrary CSM \bowtie , the maximum in this definition need not exist since D ranges over infinitely many concept descriptions. However, two properties that

are satisfied by many similarity measures considered in the literature are sufficient to obtain well-definedness for m_{\bowtie} . The first is equivalence invariance:

- The CSM \bowtie is *equivalence invariant* iff $C \equiv C'$ and $D \equiv D'$ implies $C \bowtie D = C' \bowtie D'$ for all $C, C', D, D' \in \mathcal{C}_{\mathcal{EL}}$.

To formulate the second property, we need to recall that the *role depth* of an \mathcal{EL} concept description C is the maximal nesting of existential restrictions in C (see Section 2 for the formal definition); equivalently, it is the height of the description tree T_C . The *restriction C_k of C to role depth k* is the concept description whose description tree is obtained from T_C by removing all the nodes (and edges leading to them) whose distance from the root is larger than k . More formally,

$$\begin{aligned} C_k &:= C_k && \text{if } C_k \in \mathbf{N}_C \text{ or } C_k = \top, \\ C_k &:= [C_1]_k \sqcap \dots \sqcap [C_n]_k && \text{if } C = C_1 \sqcap \dots \sqcap C_n, \\ [\exists r.C]_k &:= \begin{cases} \top & \text{if } k = 0, \\ \exists r.[C]_{k-1} & \text{otherwise.} \end{cases} \end{aligned}$$

- The CSM \bowtie is *role-depth bounded* iff $C \bowtie D = C_k \bowtie D_k$ for all $C, D \in \mathcal{C}_{\mathcal{EL}}$ and any k that is larger than the minimal role depth of C, D .

Role-depth boundedness implies that, in Definition 49, we can restrict the maximum computation to concepts D whose role depth is at most $d+1$, where d is the role depth of C . Since it is well-known that, up to equivalence, $\mathcal{C}_{\mathcal{EL}}$ contains only finitely many concept descriptions of any fixed role depth (see Proposition 13 in [5]), these two properties yield well-definedness for m_{\bowtie} . For m_{\bowtie} to be a graded membership function, it also needs to satisfy the properties *M1* and *M2*. To obtain these two properties for m_{\bowtie} , we must require that \bowtie satisfies the following additional property:

- The CSM \bowtie is *equivalence closed* iff the following equivalence holds:
 $C \equiv D$ iff $C \bowtie D = 1$.

Proposition 50. *Let \bowtie be an equivalence invariant, role-depth bounded, and equivalence closed CSM. Then m_{\bowtie} is a well-defined graded membership function.*

Proof. Let \mathcal{I} be an interpretation, $d \in \Delta^{\mathcal{I}}$ and C an \mathcal{EL} concept description of role-depth k . Since \bowtie is role-depth bounded, this means that $m_{\bowtie}^{\mathcal{I}}(d, C)$ can be equivalently expressed as:

$$\max\{C \bowtie D \mid D \in \mathcal{C}_{\mathcal{EL}}, d \in D^{\mathcal{I}} \text{ and } \text{rd}(D) \leq k + 1\} \quad (3)$$

Now, let D_1 be an \mathcal{EL} concept description such that $d \in [D_1]^{\mathcal{I}}$. Since \bowtie is equivalence invariant, this means that for any other \mathcal{EL} concept D_2 such that

$D_1 \equiv D_2$, the values $C \bowtie D_1$ and $C \bowtie D_2$ are the same. Therefore, since there are finitely many concepts in $\mathcal{C}_{\mathcal{EL}}$ of depth at most $k + 1$ (up to equivalence), it follows that the maximum always exists.

Since \bowtie is equivalence closed, it easily follows that m_{\bowtie} satisfies property *M1*. As mentioned in Section 3, the right to left implication in *M2* already follows from *M1*. The left to right direction is a consequence of the definition of m_{\bowtie} and the fact that \bowtie is equivalence invariant. Hence, m_{\bowtie} satisfies property *M2*.

Thus, m_{\bowtie} is a well-defined graded membership function. \square

Consequently, an equivalence invariant, role-depth bounded, and equivalence closed CSM \bowtie induces a DL $\tau\mathcal{EL}(m_{\bowtie})$. Moreover, as we show in the following, computing instances of threshold concepts of the form $Q_{>t}$ in this logic corresponds to answering relaxed instance queries w.r.t. \bowtie .

Proposition 51. *Let \bowtie be an equivalence invariant, role-depth bounded, and equivalence closed CSM, \mathcal{A} an \mathcal{EL} ABox, and $t \in [0, 1)$. Then*

$$\text{Relax}_t^{\bowtie}(Q, \mathcal{A}) = \{a \mid \mathcal{A} \models Q_{>t}(a) \text{ and } a \text{ occurs in } \mathcal{A}\},$$

where the semantics of the threshold concept $Q_{>t}$ is defined as in $\tau\mathcal{EL}(m_{\bowtie})$.

Proof. (\Rightarrow) Let $a \in \text{Ind}(\mathcal{A})$ such that $a \in \text{Relax}_t^{\bowtie}(Q, \mathcal{A})$. Then, there exists an \mathcal{EL} concept description X such that $\mathcal{A} \models X(a)$ and $Q \bowtie X > t$. Since $\mathcal{A} \models X(a)$, this means that for each interpretation \mathcal{J} such that $\mathcal{J} \models \mathcal{A}$, it happens that $a^{\mathcal{J}} \in X^{\mathcal{J}}$. Hence, by definition of m_{\bowtie} we have $m_{\bowtie}^{\mathcal{J}}(d, Q) > t$ for all models of \mathcal{A} . Thus, $\mathcal{A} \models Q_{>t}(a)$.

(\Leftarrow) Conversely, assume that $\mathcal{A} \models Q_{>t}(a)$. By definition of m_{\bowtie} , we know that for each model \mathcal{J} of \mathcal{A} there exists $X_{\mathcal{J}}$ such that $a^{\mathcal{J}} \in [X_{\mathcal{J}}]^{\mathcal{J}}$ and $Q \bowtie X_{\mathcal{J}} > t$. However, to guarantee that $a \in \text{Relax}_t^{\bowtie}(Q, \mathcal{A})$, we need to show that there exists one such concept which is common for all models of \mathcal{A} .

To this end, consider the description graph $G(\mathcal{A})$ induced by \mathcal{A} . Additionally, let $\mathcal{I}_{\mathcal{A}}$ denote the interpretation corresponding to $G(\mathcal{A})$ such that $a^{\mathcal{I}_{\mathcal{A}}} = a$ for all $a \in \text{Ind}(\mathcal{A})$. The following facts are easy consequences of Theorem 15:

- $\mathcal{I}_{\mathcal{A}} \models \mathcal{A}$, and
- for each \mathcal{J} such that $\mathcal{J} \models \mathcal{A}$, there exists a homomorphism $\varphi_{\mathcal{J}}$ from $G(\mathcal{A})$ to $G_{\mathcal{J}}$ with $\varphi(a) = a^{\mathcal{J}}$ for all $a \in \text{Ind}(\mathcal{A})$.

Since $\mathcal{I}_{\mathcal{A}} \models \mathcal{A}$, this means that there exists an \mathcal{EL} concept description X such that $Q \bowtie X > t$ and $a^{\mathcal{I}_{\mathcal{A}}} \in X^{\mathcal{I}_{\mathcal{A}}}$. The membership characterization via homomorphism in Theorem 5, yields the existence of a homomorphism φ_1 from T_X to $G(\mathcal{A})$ with $\varphi_1(v_0) = a$. Then, the composition $\varphi_{\mathcal{J}} \circ \varphi_1$ yields a similar homomorphism to each model \mathcal{J} of \mathcal{A} , which implies $a^{\mathcal{J}} \in X^{\mathcal{J}}$. Therefore, $\mathcal{A} \models X(a)$ and thus, $a \in \text{Relax}_t^{\bowtie}(Q, \mathcal{A})$. \square

6.2 Relation to the membership degree function *deg*

Lehman and Turhan [16] introduced a framework (called *simi framework*) that can be used to define a variety of similarity measures between \mathcal{EL} concepts satisfying the properties required by our Propositions 50 and 51. They first define a directional measure $simi_d$ and then use a fuzzy connector \otimes to combine the values obtained by comparing the concepts in both directions with this directional measure. Given two \mathcal{EL} concepts C and D , one could say that *simi* uses $simi_d$ to measure how many properties of C are present in D and vice versa. Then, the bidirectional similarity measure *simi* is defined as:

$$simi(C, D) := simi_d(C^r, D^r) \otimes simi_d(D^r, C^r)$$

Regarding the fuzzy connector, it is an operator $\otimes : [0, 1] \times [0, 1] \rightarrow [0, 1]$ satisfying certain properties (see [16]). Examples are the *average operator* and *t-norms*. The general definition of *simi* can be found in [16]. Here, we consider only one instance of this framework and show that the similarity measure obtained this way induces our graded membership function *deg*.

Definition 52 ([16]). Let C, D be two \mathcal{EL} concept descriptions. If one of these two concepts is equivalent to \top , then we define $simi_d(\top, D) := 1$ for all D and $simi_d(D, \top) := 0$ for $D \not\equiv \top$. Otherwise, let $top(C), top(D)$ respectively be the set of concept names and existential restrictions in the top-level conjunction of C, D . We define

$$simi_d(C, D) := \frac{\sum_{C' \in top(C)} \max\{simi_a(C', D') \mid D' \in top(D)\}}{|top(C)|}, \text{ where}$$

$$\begin{aligned} simi_a(A, A) &:= 1, \quad simi_a(A, B) := 0 \text{ for } A, B \in \mathbf{N}_C, A \neq B, \\ simi_a(\exists r.E, A) &:= simi_a(A, \exists r.E) := 0 \text{ for } A \in \mathbf{N}_C, \\ simi_a(\exists r.E, \exists r.F) &:= simi_d(E, F), \quad simi_a(\exists r.E, \exists s.F) := 0 \text{ for } r, s \in \mathbf{N}_R, r \neq s. \end{aligned}$$

The bidirectional similarity measure \bowtie^1 is then defined as

$$C \bowtie^1 D := \min\{simi_d(C^r, D^r), simi_d(D^r, C^r)\}.$$

In [16], several properties for $simi_d$ are shown. Among them, the following will be useful later on to obtain our results.

$$simi_d(C, D) = 1 \text{ iff } D \sqsubseteq C \tag{4}$$

It is easy to show that \bowtie^1 is equivalence invariant, role-depth bounded, and equivalence closed. Equivalence closed follows from [16], since it is shown that this is the case for any instance of *simi*. Looking at the definition of $simi_d$ it is

not hard to see that \bowtie^1 is also role-depth bounded. Finally, note that equivalence invariance depends on the fact that we apply $simi_d$ to the reduced forms of C, D . Since \bowtie^1 satisfies the properties required by Proposition 50, it induces a graded membership function m_{\bowtie^1} . From now on we will refer to $simi_d$, in the context of m_{\bowtie^1} , as \bowtie_d^1 .

To show that m_{\bowtie^1} coincides with deg , some intermediate results need to be shown. We start by showing the following relation between the directional measure \bowtie_d^1 and deg .

Lemma 53. *Let X be an \mathcal{EL} concept description and \mathcal{I}_X be the interpretation corresponding to the \mathcal{EL} description tree T_X . Then, for each \mathcal{EL} concept description C , it holds:*

$$C^r \bowtie_d^1 X = deg^{\mathcal{I}_X}(d_0, C)$$

where d_0 is the domain element corresponding to the root of T_X .

Proof. We prove the claim by induction on the structure of C .

Induction Base. $C \in \mathbf{N}_C$ or $C = \top$. Then, $C = C^r$. If C^r is of the form A , then $A \bowtie_d^1 X = 1$ when $A \in top(X)$ and 0 otherwise. A similar relationship holds for $deg^{\mathcal{I}_X}(d_0, A)$, but with respect to whether $d_0 \in A^{\mathcal{I}_X}$. Since $A \in top(X)$ iff $d_0 \in A^{\mathcal{I}_X}$, this means that $A \bowtie_d^1 X = deg^{\mathcal{I}_X}(d_0, A)$. The case for \top is trivial, since $\top \bowtie_d^1 X = deg^{\mathcal{I}_X}(d_0, \top) = 1$.

Induction Step. We distinguish two cases:

- C is of the form $\exists r.D$. Then, C^r is of the form $\exists r.D^r$. By definition of \bowtie_d^1 and deg , it is easy to see that whenever X does not have a top-level atom of the form $\exists r.X'$, it is the case that:

$$\exists r.D^r \bowtie_d^1 X = deg^{\mathcal{I}_X}(d_0, \exists r.D) = 0$$

Hence, without loss of generality, we focus on the cases where there exists at least one top-level atom in X of the form $\exists r.X'$. Consequently, since $|top(\exists r.D^r)| = 1$, we have:

$$\exists r.D^r \bowtie_d^1 X = \max\{D^r \bowtie_d^1 X' \mid \exists r.X' \in top(X)\} \quad (5)$$

Since \mathcal{I}_X is induced by T_X , then for each atom $\exists r.X' \in top(X)$ there exists a corresponding domain element $e \in \Delta^{\mathcal{I}_X}$ such that $(d_0, e) \in r^{\mathcal{I}_X}$. This correspondence also holds in the opposite direction. Moreover, it is easy to see that the tree rooted at e in T_X corresponds to the \mathcal{EL} description tree $T_{X'}$. Hence, the application of induction hypothesis to D yields:

$$D^r \bowtie_d^1 X' = deg^{\mathcal{I}_X}(e, D), \text{ for all } \exists r.X' \in top(X)$$

Therefore, it follows from Equation 5:

$$\exists r.D^r \bowtie_d^1 X = \max\{deg^{\mathcal{I}X}(e, D) \mid (d_0, e) \in r^{\mathcal{I}X}\} \quad (6)$$

Now, let $T_{\exists r.D^r}$ be the corresponding \mathcal{EL} description tree of $\exists r.D^r$ and v_0 its root. Obviously, there exists exactly one r -successor v_1 of v_0 in $T_{\exists r.D^r}$ and moreover, the subtree of $T_{\exists r.D^r}$ rooted at v_1 is exactly the \mathcal{EL} description tree T_{D^r} associated to D^r . Consider, then, the set $\mathcal{H}(T_{\exists r.D^r}, G_{\mathcal{I}X}, d_0)$. By Definition 20 we have:

$$deg^{\mathcal{I}X}(d_0, \exists r.D) = \max\{h_w(v_0) \mid h \in \mathcal{H}(T_{\exists r.D^r}, G_{\mathcal{I}X}, d_0)\} \quad (7)$$

Now, let h be any *ptgh* in $\mathcal{H}(T_{\exists r.D^r}, G_{\mathcal{I}X}, d_0)$ with $h(v_1) = e$, for some $e \in \Delta^{\mathcal{I}X}$ such that $(d_0, e) \in r^{\mathcal{I}X}$. We know that there exists at least one and any *ptgh* h' of a different form will not be interesting, since $h'_w(v_0) = 0$. By definition of h_w (Definition 19), it follows that $h_w(v_0) = h_w(v_1)$. Additionally, for any *ptgh* $h \in \mathcal{H}(T_{\exists r.D^r}, G_{\mathcal{I}X}, d_0)$ with $h(v_1) = e$, its restriction to $(V_{T_{\exists r.D^r}} \setminus \{v_0\})$ is a *ptgh* in $\mathcal{H}(T_{D^r}, G_{\mathcal{I}X}, e)$. Conversely, any *ptgh* g in $\mathcal{H}(T_{D^r}, G_{\mathcal{I}X}, e)$ can be extended to a *ptgh* in $\mathcal{H}(T_{\exists r.D^r}, G_{\mathcal{I}X}, d_0)$, by defining $g(v_0) = d_0$. Hence, Equation 7 can be transformed into:

$$deg^{\mathcal{I}X}(d_0, \exists r.D) = \max_{(d_0, e) \in r^{\mathcal{I}X}} \{g_w(v_1) \mid g \in \mathcal{H}(T_{D^r}, G_{\mathcal{I}X}, e)\}$$

Finally, since for each $e \in \Delta^{\mathcal{I}X}$ there exists a *ptgh* $g \in \mathcal{H}(T_{D^r}, G_{\mathcal{I}X}, e)$ such that $deg^{\mathcal{I}X}(e, D) = g_w(v_1)$ and $g_w(v_1)$ gives the maximum value, we further obtain the following equation:

$$deg^{\mathcal{I}X}(d_0, \exists r.D) = \max\{deg^{\mathcal{I}X}(e, D) \mid (d_0, e) \in r^{\mathcal{I}X}\} \quad (8)$$

Thus, the combination of Equations 6 and 8 yields

$$\exists r.D^r \bowtie_d^1 X = deg^{\mathcal{I}X}(d_0, \exists r.D)$$

- C is of the form $C_1 \sqcap \dots \sqcap C_k$. Then, its reduced form C^r is of the form $D_1 \sqcap \dots \sqcap D_n$, where $1 \leq n \leq k$ and each D_j is the reduced form $[C_i]^r$ of some conjunct C_i . Now, it is easy to see from the definition of \bowtie_d^1 , that $C^r \bowtie_d^1 X$ can be equivalently expressed as:

$$C^r \bowtie_d^1 X = \frac{\sum_{j=1}^n (D_j \bowtie_d^1 X)}{n} \quad (9)$$

Furthermore, though more involved, it is not hard to see from the definitions of deg and h_w , that a similar situation occurs with respect to deg :

$$deg^{\mathcal{I}X}(d_0, C^r) = \frac{\sum_{j=1}^n deg^{\mathcal{I}X}(d_0, D_j)}{n} \quad (10)$$

Then, for each D_j one can apply induction hypothesis to the atom C_i that has $[C_i]^r = D_j$ to obtain $D_j \bowtie_d^1 X = \text{deg}^{\mathcal{I}X}(d_0, C_i)$. Since deg is equivalence invariant (in the sense of property $M2$), we have $D_j \bowtie_d^1 X = \text{deg}^{\mathcal{I}X}(d_0, C_i) = \text{deg}^{\mathcal{I}X}(d_0, D_j)$. Hence, the combination of Equations 9 and 10 yields $C^r \bowtie_d^1 X = \text{deg}^{\mathcal{I}X}(d_0, C)$.

□

The following lemma is an easy consequence of Lemma 53 and the properties of deg . It tells us that for \bowtie_d^1 it does not make a difference to consider the concept on the *right hand side* in reduced form.

Lemma 54. *Let C, X_1 and X_2 be \mathcal{EL} concept descriptions such that $X_1 \equiv X_2$. Then,*

$$C^r \bowtie_d^1 X_1 = C^r \bowtie_d^1 X_2$$

Proof. From Lemma 53 we have:

$$C^r \bowtie_d^1 X_1 = \text{deg}^{\mathcal{I}X_1}(d_{01}, C) \quad \text{and} \quad C^r \bowtie_d^1 X_2 = \text{deg}^{\mathcal{I}X_2}(d_{02}, C)$$

Since $X_1 \equiv X_2$, this means that $X_1 \sqsubseteq X_2$ and $X_2 \sqsubseteq X_1$. Therefore, by the characterization of subsumption from Theorem 4, there are homomorphisms φ_1 and φ_2 from T_{X_1} to T_{X_2} and vice versa, such that $\varphi_1(d_{01}) = d_{02}$ and $\varphi_2(d_{02}) = d_{01}$, respectively. Based on this, a double application of Lemma 24 yields:

$$\text{deg}^{\mathcal{I}X_1}(d_{01}, C) \leq \text{deg}^{\mathcal{I}X_2}(d_{02}, C) \leq \text{deg}^{\mathcal{I}X_1}(d_{01}, C)$$

Thus, $C^r \bowtie_d^1 X_1 = C^r \bowtie_d^1 X_2$. □

Lemma 53 tells us in a certain way, that given an interpretation \mathcal{I} the maximal value for $C^r \bowtie_d^1 X$ among those X such that $d \in X^{\mathcal{I}}$, is the same as $\text{deg}^{\mathcal{I}}(d, C)$. However, to show that $m_{\bowtie_d^1}$ is equivalent to deg , one must not forget that \bowtie_d^1 is used in both directions to compute $m_{\bowtie_d^1}^{\mathcal{I}}(d, C)$. The following example gives the intuition of why the value $X^r \bowtie_d^1 C^r$ can be ignored.

Example 55. Consider the \mathcal{EL} concept description $C := A_1 \sqcap A_2 \sqcap \exists r.A_1$ and the interpretation \mathcal{I} :

- $\Delta^{\mathcal{I}} := \{d_0, d_1, d_2\}$,
- $A_1^{\mathcal{I}} := \{d_0, d_1\}$, $A_2^{\mathcal{I}} := \{d_2\}$ and $A_3^{\mathcal{I}} := \{d_0\}$,
- $r^{\mathcal{I}} := \{(d_0, d_1)\}$ and $s^{\mathcal{I}} := \{(d_0, d_2)\}$.

One can see that T_C consists of two nodes v_0 and v_1 and the edge v_0rv_1 . Then, regarding the computation of $deg^{\mathcal{I}}(d_0, C)$, the best mapping φ from T_C to $G_{\mathcal{I}}$ with $\varphi(v_0) = d_0$ is the one having $\varphi(v_1) = d_1$. This yields $deg^{\mathcal{I}}(d_0, C) = 2/3$.

Now, let us see what happens with m_{\bowtie^1} . Clearly, for the concept description $X := A_1 \sqcap A_2 \sqcap \exists r.A_1 \sqcap \exists s.A_2$, it holds $d_0 \in X^{\mathcal{I}}$. Note that $\mathcal{I}_X = \mathcal{I}$, therefore as a consequence of Lemmas 53 and 54 we have $C^r \bowtie_d^1 X^r = C^r \bowtie_d^1 X = deg^{\mathcal{I}}(d_0, C)$. However, to compute $m_{\bowtie^1}^{\mathcal{I}}(d_0, C)$, the opposite direction when using \bowtie_d^1 also counts. In particular, $X^r \bowtie_d^1 C^r = 1/2$ and then, $C \bowtie^1 X = X^r \bowtie_d^1 C^r = 1/2$.

In contrast, the concept $Y := A_1 \sqcap \exists r.A_1$ also has d_0 as an instance under \mathcal{I} , it keeps the value $C^r \bowtie_d^1 Y = 2/3$ and more important: since $Y \sqsubseteq C$, by property 4 it holds that $Y \bowtie_d^1 C = 1$. This implies that $C \bowtie^1 Y = C^r \bowtie_d^1 Y^r = 2/3$. Hence, by definition of m_{\bowtie^1} , Y is preferred over X to obtain the value of $m_{\bowtie^1}^{\mathcal{I}}(d, C)$ and the value $C^r \bowtie_d^1 Y^r$ does not have any influence on the final value.

We now show that the intuition illustrated in the previous example is always satisfied by \bowtie^1 .

Lemma 56. *For every pair of \mathcal{EL} concept descriptions C and X , there exists an \mathcal{EL} concept description Y such that:*

1. $X \sqsubseteq Y$ and $C \sqsubseteq Y$.
2. $C^r \bowtie_d^1 X = C^r \bowtie_d^1 Y$.

Proof. The application of Lemma 53 yields:

$$C^r \bowtie_d^1 X = deg^{\mathcal{I}_X}(d_0, C)$$

Once deg is introduced, we know that there exists a *ptgh* h from T_{C^r} to \mathcal{I}_X such that $h(v_0) = d_0$ and $h_w(v_0) = deg^{\mathcal{I}_X}(d_0, C)$. Let \mathcal{I}_h be the canonical interpretation induced by h (Definition 21). Since \mathcal{I}_h is a tree we can speak of its corresponding \mathcal{EL} concept description and denote it as Y .

By Remark 22, there exists a homomorphism from $T_{\mathcal{I}_h}$ to \mathcal{I}_X . Furthermore, by construction of \mathcal{I}_h , it is easy to see that there is also a homomorphism from $T_{\mathcal{I}_h}$ to T_{C^r} . Since $C \equiv C^r$, the application of Theorem 4 yields $X \sqsubseteq Y$ and $C \sqsubseteq Y$.

Similarly as for X , since \mathcal{I}_h corresponds to T_Y , we have:

$$C^r \bowtie_d^1 Y = deg^{\mathcal{I}_h}(v_0, C)$$

In addition, the application of Lemma 25 yields $deg^{\mathcal{I}_h}(v_0, C) = deg^{\mathcal{I}_X}(d_0, C)$. Thus, we can conclude that $C^r \bowtie_d^1 X = C^r \bowtie_d^1 Y$. \square

As already mentioned in Example 55, since \bowtie^1 is an instance of *simi*, it then satisfies property 4. Since $C \equiv C^r$ and $Y \equiv Y^r$, this means that $C^r \sqsubseteq Y^r$ and therefore, $Y^r \bowtie_d^1 C^r = 1$. In addition, the operator \min satisfies $\min\{x, 1\} = x$ and it is monotonic, i.e., $\min\{x, y\} \leq \min\{x, z\}$ whenever $y \leq z$. This actually implies that the value $Y^r \bowtie_d^1 C^r$ becomes irrelevant when computing m_{\bowtie^1} . We formally prove this in the following lemma.

Lemma 57. *For all interpretations \mathcal{I} , $d \in \Delta^{\mathcal{I}}$, and \mathcal{EL} concept descriptions C we have:*

$$m_{\bowtie^1}^{\mathcal{I}}(d, C) = \max\{C^r \bowtie_d^1 D \mid D \in \mathcal{C}_{\mathcal{EL}} \text{ and } d \in D^{\mathcal{I}}\}$$

Proof. By Definition 49

$$m_{\bowtie^1}^{\mathcal{I}}(d, C) = \max\{C \bowtie^1 D \mid D \in \mathcal{C}_{\mathcal{EL}} \text{ and } d \in D^{\mathcal{I}}\}$$

Let X be a concept description such that $m_{\bowtie^1}^{\mathcal{I}}(d, C) = C \bowtie^1 X$ and $d \in X^{\mathcal{I}}$. By definition of \bowtie^1 we know that:

$$C \bowtie^1 X = \min\{C^r \bowtie_d^1 X^r, X^r \bowtie_d^1 C^r\}$$

One can assume without loss of generality that X satisfies $X^r \bowtie_d^1 C^r = 1$. If that were not the case for any maximal X , the application of Lemma 56 to C and X^r yields a concept Y such that:

$$C^r \bowtie_d^1 X^r = C^r \bowtie_d^1 Y, \quad Y \bowtie_d^1 C^r = 1, \quad X^r \sqsubseteq Y$$

This means that $d \in Y^{\mathcal{I}}$ and since $Y \equiv Y^r$, the application of Lemma 54 yields $C^r \bowtie_d^1 Y = C^r \bowtie_d^1 Y^r$. Then,

$$C \bowtie^1 Y = \min\{C^r \bowtie_d^1 Y, 1\}$$

The monotonicity of \min implies $C \bowtie^1 X \leq C \bowtie^1 Y$. Consequently, Y must be a maximal concept in the definition of $m_{\bowtie^1}^{\mathcal{I}}(d, C)$, which would give a contradiction. Therefore, it is safe to assume $X^r \bowtie_d^1 C^r = 1$. Thus,

$$m_{\bowtie^1}^{\mathcal{I}}(d, C) = C \bowtie^1 X = C^r \bowtie_d^1 X^r \tag{11}$$

Now, suppose for a contradiction that there exists a concept description Y such that $d \in Y^{\mathcal{I}}$ and $C^r \bowtie_d^1 X^r < C^r \bowtie_d^1 Y$. Similarly as before, the application of Lemmas 54 and 56 to C and Y yields a concept Z such that:

$$C^r \bowtie_d^1 Y = C^r \bowtie_d^1 Z^r, \quad Z^r \bowtie_d^1 C^r = 1, \quad Y \sqsubseteq Z$$

Hence, $C \bowtie^1 Z = C^r \bowtie_d^1 Z^r$ and $C^r \bowtie_d^1 X^r < C^r \bowtie_d^1 Z^r$. Then, using Equation 11 we have $C \bowtie^1 X < C \bowtie^1 Z$. Since $Y \sqsubseteq Z$ implies $d \in Z^{\mathcal{I}}$, we obtain a contradiction w.r.t. the maximality of X in the definition of $m_{\bowtie^1}^{\mathcal{I}}$.

Thus, there is no Y such that $C^r \bowtie_d^1 X^r < C^r \bowtie_d^1 Y$ and our claim follows from Equation 11. \square

Next, we finally show the equivalence between m_{\bowtie^1} and deg .

Theorem 58. *For all interpretations \mathcal{I} , $d \in \Delta^{\mathcal{I}}$, and \mathcal{EL} concept descriptions Q we have $m_{\bowtie^1}^{\mathcal{I}}(d, Q) = deg^{\mathcal{I}}(d, Q)$.*

Proof. (\Rightarrow) From Lemma 57, we know that there exists an \mathcal{EL} concept description X such that $m_{\bowtie^1}^{\mathcal{I}}(d, Q) = Q^r \bowtie_d^1 X$ and $d \in X^{\mathcal{I}}$. The application of Lemma 53 to Q and X yields:

$$Q^r \bowtie_d^1 X = deg^{\mathcal{I}_X}(d_0, Q)$$

The characterization of crisp membership in \mathcal{EL} yields the existence of a homomorphism φ from $G_{\mathcal{I}_X}$ (or T_X) to $G_{\mathcal{I}}$ with $\varphi(d_0) = d$. Hence, the application of Lemma 24 yields $deg^{\mathcal{I}_X}(d_0, Q) \leq deg^{\mathcal{I}}(d, Q)$. Therefore, we obtain:

$$m_{\bowtie^1}^{\mathcal{I}}(d, Q) \leq deg^{\mathcal{I}}(d, Q) \tag{12}$$

(\Leftarrow) Consider a *ptgh* $h \in \mathcal{H}(T_{Q^r}, G_{\mathcal{I}}, d)$ such that $h_w(v_0) = deg^{\mathcal{I}}(d, Q)$. Let \mathcal{I}_h be the canonical interpretation induced by h . Since $T_{\mathcal{I}_h}$ is a tree, we can speak of its corresponding \mathcal{EL} concept description $Q_{\mathcal{I}_h}$. Then, we obtain the following equalities:

$$Q^r \bowtie_d^1 Q_{\mathcal{I}_h} = deg^{\mathcal{I}_h}(v_0, Q) \tag{Lemma 53}$$

$$= deg^{\mathcal{I}}(d, Q) \tag{Lemma 25}$$

Furthermore, it is easy to see that by definition of \mathcal{I}_h , it holds that $d \in [Q_{\mathcal{I}_h}]^{\mathcal{I}}$. Hence, Lemma 57 implies that $Q^r \bowtie_d^1 Q_{\mathcal{I}_h} \leq m_{\bowtie^1}^{\mathcal{I}}(d, Q)$ and consequently:

$$deg^{\mathcal{I}}(d, Q) \leq m_{\bowtie^1}^{\mathcal{I}}(d, Q) \tag{13}$$

Thus, our claim follows from the combination of inequations 12 and 13. \square

Proposition 51 thus implies that answering of relaxed instance queries w.r.t. \bowtie^1 is the same as computing instances for threshold concepts of the form $Q_{>t}$ in $\tau\mathcal{EL}(deg)$. Since such concepts are positive, Proposition 47 yields the following corollary.

Corollary 59. *Let \mathcal{A} be an \mathcal{EL} ABox, Q an \mathcal{EL} query concept, a an individual name, and $t \in [0, 1)$. Then it can be decided in polynomial time whether $a \in \text{Relax}_{t}^{\bowtie^1}(Q, \mathcal{A})$ or not.*

Note that Ecke et al. [9, 10] show only an NP upper bound w.r.t. data complexity for this problem, albeit for a larger class of instances of the simi framework.

7 Conclusion

We have introduced a family of DLs $\tau\mathcal{EL}(m)$ parameterized with a graded membership function m , which extends the popular lightweight DL \mathcal{EL} by threshold concepts that can be used to approximate classical concepts. Inspired by the homomorphism characterization of membership in \mathcal{EL} concepts, we have defined a particular membership function deg and have investigated the complexity of reasoning in $\tau\mathcal{EL}(deg)$. It turns out that the higher expressiveness takes its toll: whereas reasoning in \mathcal{EL} can be done in polynomial time, it is NP- or coNP-complete in $\tau\mathcal{EL}(deg)$, depending on which inference problem is considered. applications since they provide a flexible and formally well-founded way to define concepts by approximation. We have also shown that concept similarity measures satisfying certain properties can be used to define graded membership functions. In particular, the function deg can be constructed in this way from a particular instance of the *simi* framework of Lehmann and Turhan [16]. Nevertheless, our direct definition of deg based on homomorphisms is important since the partial tree-to-graph homomorphisms used there are the main technical tool for showing our decidability and complexity results.

While introduced as formalism for defining concepts by approximation, a possible use-case for $\tau\mathcal{EL}(deg)$ is relaxation of instance queries, as motivated and investigated in [9, 10]. Compared to the setting considered in [9, 10], $\tau\mathcal{EL}(deg)$ yields a considerably more expressive query language since we can combine threshold concepts using the constructors of \mathcal{EL} and can also forbid that thresholds are reached. Restricted to the setting of relaxed instance queries, our approach actually allows relaxed instance checking in polynomial time. On the other hand, [9, 10] can also deal with other instances of the *simi* framework.

An important topic for future research is to consider graded membership functions m_{\bowtie} that are induced by other instances of *simi*. We conjecture that these instances can also be defined directly by an appropriate adaptation of our homomorphism-based definition. The hope is then that our decidability and complexity results can be generalized to these functions. Another important topic for future research is to add TBoxes. While acyclic TBoxes can already be handled by our approach through unfolding, we would like to treat them directly by an adaptation of the homomorphism-based approach to avoid a possible exponential blowup due to unfolding. For cyclic and general TBoxes, homomorphisms probably need to be replaced by simulations [1, 10].

References

- [1] Franz Baader. Terminological cycles in a description logic with existential restrictions. In Georg Gottlob and Toby Walsh, editors, *IJCAI*, pages 325–330. Morgan Kaufmann, 2003.
- [2] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA, 2003.
- [3] Franz Baader and Bernhard Hollunder. A terminological knowledge representation system with complete inference algorithms. In Harold Boley and Michael M. Richter, editors, *Processing Declarative Knowledge, International Workshop PDK'91, Kaiserslautern, Germany, July 1-3, 1991, Proceedings*, volume 567 of *Lecture Notes in Computer Science*, pages 67–86. Springer, 1991.
- [4] Franz Baader, Ralf Küsters, and Ralf Molitor. Computing least common subsumers in description logics with existential restrictions. In Thomas Dean, editor, *IJCAI*, pages 96–103. Morgan Kaufmann, 1999.
- [5] Franz Baader, Baris Sertkaya, and Anni-Yasmin Turhan. Computing the least common subsumer w.r.t. a background terminology. *J. Applied Logic*, 5(3):392–420, 2007.
- [6] Stefan Borgwardt, Felix Distel, and Rafael Peñaloza. The limits of decidability in fuzzy description logics with general concept inclusions. *Artificial Intelligence*, 218:23–55, 2015.
- [7] Sebastian Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else? In Ramon López de Mántaras and Lorenza Saitta, editors, *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, Valencia, Spain, August 22-27, 2004*, pages 298–302. IOS Press, 2004.
- [8] Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Andrea Schaerf. Deduction in concept languages: From subsumption to instance checking. *J. Log. Comput.*, 4(4):423–452, 1994.
- [9] Andreas Ecke, Rafael Peñaloza, and Anni-Yasmin Turhan. Answering instance queries relaxed by concept similarity. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014.

- [10] Peñaloza R. Turhan A.Y. Ecke, A. Similarity-based relaxed instance queries. In press. *Journal of Applied Logic*, 2015.
- [11] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. LTCS-Report LTCS-98-09, LuFG Theoretical Computer Science, RWTH Aachen, Germany, 1998. See <http://lat.inf.tu-dresden.de/research/reports.html>.
- [12] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [13] Bernhard Hollunder. Consistency checking reduced to satisfiability of concepts in terminological systems. *Ann. Math. Artif. Intell.*, 18(2-4):133–157, 1996.
- [14] Ralf Küsters. *Non-Standard Inferences in Description Logics*, volume 2100 of *Lecture Notes in Computer Science*. Springer, 2001.
- [15] Ralf Küsters and Ralf Molitor. Approximating most specific concepts in description logics with existential restrictions. *AI Commun.*, 15(1):47–59, 2002.
- [16] Karsten Lehmann and Anni-Yasmin Turhan. A framework for semantic-based similarity measures for \mathcal{ELH} -concepts. In Luis Fariñas del Cerro, Andreas Herzig, and Jérôme Mengin, editors, *Logics in Artificial Intelligence - 13th European Conference, JELIA 2012, Toulouse, France, September 26-28, 2012. Proceedings*, volume 7519 of *Lecture Notes in Computer Science*, pages 307–319. Springer, 2012.
- [17] Boontawee Suntisrivaraporn. A similarity measure for the description logic EL with unfoldable terminologies. In *2013 5th International Conference on Intelligent Networking and Collaborative Systems, Xi'an city, Shaanxi province, China, September 9-11, 2013*, pages 408–413. IEEE, 2013.

8 Appendix

Missing proofs of Section 3

Theorem 14. Let \widehat{C} be a $\tau\mathcal{EL}(m)$ concept description and $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ an interpretation. The following statements are equivalent for all $d \in \Delta^{\mathcal{I}}$:

1. $d \in \widehat{C}^{\mathcal{I}}$.
2. there exists a τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = d$.

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 $\widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_m \sqcap \exists r_1. \widehat{D}_1 \sqcap \dots \sqcap \exists r_n. \widehat{D}_n$, where each \widehat{C}_i is of the form $A \in \mathbf{N}_C$ or $E_{\sim q} \in \widehat{\mathbf{N}}_E$.

(\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 $1 \leq i \leq m$ 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. $\text{rd}(\widehat{C}) = 0$. Then, $n = 0$ and $T_{\widehat{C}}$ consists only of one node v_0 (the root), it has no edges and $\widehat{\ell}_T(v_0) = \{\widehat{C}_1, \dots, \widehat{C}_m\}$. The mapping $\phi(v_0) = d$ is a τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$. Note that, for each \widehat{C}_i of the form $A \in \mathbf{N}_C$ we know $A \in \mathcal{L}_{\mathcal{I}}(d)$, and consequently ϕ satisfies Condition 1 in Definition 13. In case \widehat{C}_i is of the form $E_{\sim q}$, the fact that $d \in [\widehat{C}_i]^{\mathcal{I}}$ implies that ϕ satisfies Condition 2 in Definition 13.

Induction Step. Assume that the claim holds for all the concepts with *role-depth* smaller than k . We show that it also holds for $\text{rd}(\widehat{C}) = k$. First, consider the concept $\widehat{D}_0 = \widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_m$. 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 $\text{rd}(\widehat{D}_0) = 0$, 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 *top-level* concept $\exists r_j. \widehat{D}_j$ 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 $\text{rd}(\widehat{D}_j) < k$, the application of induction hypothesis on 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 difficult to see that for each node $v \in V_T$, there exists exactly one τ -homomorphism ϕ_j such that $v \in \text{dom}(\phi_j)$. Based on this, we build the 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 . It remains to show that ϕ is τ -homomorphism.

1. **ϕ is homomorphism from T_C to $G_{\mathcal{I}}$:** Let v be any node from V_T . We know there exists ϕ_j such that $\phi(v) = \phi_j(v)$. Since ϕ_j is also a homomorphism, $\ell(v) = \ell_j(v)$ and $\ell_j(v) \subseteq \ell_{\mathcal{I}}(\phi_j(v))$, it follows that $\ell(v) \subseteq \ell_{\mathcal{I}}(\phi(v))$. Now, let vrw be any edge from E_T . There are two cases:
 - vrw is of the form $v_0r_jv_j$. As explained before we have $\phi(v_0) = d$, $\phi(v_j) = d_j$ and $dr_jd_j \in E_{\mathcal{I}}$. Hence, $\phi(v_0)r_j\phi(v_j) \in E_{\mathcal{I}}$.
 - $v, w \in \text{dom}(\phi_j)$ for some $j \in \{1 \dots n\}$. 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 2 in Definition 13 can be verified in a similar way.

Thus, ϕ is 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 $\widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_m$ and $\widehat{\ell}_T(v_0) = \{\widehat{C}_1, \dots, \widehat{C}_m\}$. Consider any $\widehat{C}_i \in \widehat{\ell}_T(v_0)$. We distinguish two cases:

- \widehat{C}_i is of the form $A \in \mathbf{N}_{\mathcal{C}}$. Since ϕ is τ -homomorphism, it is also a classical homomorphism in the sense of Definition 3 and then, ignoring the labels of the form E_{\sim_q} we have $\ell_T(v_0) \subseteq \ell_{\mathcal{I}}(d)$. Hence, $d \in A^{\mathcal{I}}$.
- \widehat{C}_i is of the form E_{\sim_q} . By Condition 2 in Definition 13 we also have $d \in [E_{\sim_q}]^{\mathcal{I}}$.

Thus, we have shown $d \in [\widehat{C}_i]^{\mathcal{I}}$ for each conjunct \widehat{C}_i of \widehat{C} . Consequently, $d \in \widehat{C}^{\mathcal{I}}$.

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, \dots, v_n in V_T such that $v_0r_jv_j \in E_T$. This also means that \widehat{C} is of the form $\widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_m \sqcap \exists r_1.\widehat{D}_1 \sqcap \dots \sqcap \exists r_n.\widehat{D}_n$ 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:

- $m > 0$. 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 $j \in \{1 \dots n\}$. Since ϕ is also a homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ and $v_0r_jv_j \in E_T$, then there exists $e_j \in \Delta^{\mathcal{I}}$ such that $dr_je_j \in E_{\mathcal{I}}$ 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 induction hypothesis can be applied and its application yields that $e_j \in [\widehat{D}_j]^{\mathcal{I}}$. Hence, $d \in [\exists r_j.\widehat{D}_j]^{\mathcal{I}}$.

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

Missing proofs of Section 4

Lemma 24. Let \mathcal{I} and \mathcal{J} be two interpretations such that there exists a homomorphism φ from $G_{\mathcal{I}}$ to $G_{\mathcal{J}}$. Then, for any individual $d \in \Delta^{\mathcal{I}}$ and any \mathcal{EL} concept description C it holds: $\text{deg}^{\mathcal{I}}(d, C) \leq \text{deg}^{\mathcal{J}}(\varphi(d), C)$.

Proof. Let C^r be the reduced form of C and h be any *ptgh* from T_{C^r} to $G_{\mathcal{I}}$ with $h(v_0) = d$. Since φ is a homomorphism from $G_{\mathcal{I}}$ to $G_{\mathcal{J}}$, the mapping $\varphi \circ h$ is a *ptgh* from T_{C^r} to $G_{\mathcal{J}}$ with $(\varphi \circ h)(v_0) = \varphi(d)$.

Then, we have that for each $v \in \text{dom}(h)$ the homomorphism φ makes $\ell_{\mathcal{I}}(h(v)) \subseteq \ell_{\mathcal{J}}((\varphi \circ h)(v))$. In addition, for each r -successor $w \in \text{dom}(h)$ of v in T_{C^r} , we have that if $h(w)$ is an r -successor of $h(v)$ in $G_{\mathcal{I}}$, then $(\varphi \circ h)(w)$ is also an r -successor of $(\varphi \circ h)(v)$ in $G_{\mathcal{J}}$. Hence, it follows from Definition 19 that $h_w(v_0) \leq (\varphi \circ h)_w(v_0)$ for each *ptgh* h from T_{C^r} to $G_{\mathcal{I}}$ with $h(v_0) = d$.

Thus, we can conclude that $\text{deg}^{\mathcal{I}}(d, C) \leq \text{deg}^{\mathcal{J}}(\varphi(d), C)$. \square

Definition 60. Let C be an \mathcal{EL} concept description and T_C the corresponding \mathcal{EL} description tree. For any node $v \in V_{T_C}$ we denote by $T_C[v]$ the subtree of T_C rooted at v . In addition, the \mathcal{EL} concept description $C[v]$ is the one having the description tree $T_C[v]$.

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 .

In the proof of Lemma 26, we will use concepts and description trees of the form $T_C[v]$ and $C[v]$. We remark that for each reduced concept C^r the concept $C^r[v]$ is also in reduced form, for all nodes v in 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).

Lemma 26. Let C be an \mathcal{EL} concept description, \mathcal{I} a finite interpretation and $d \in \Delta^{\mathcal{I}}$. Then, Algorithm 4 terminates on input (C, \mathcal{I}, d) and outputs $\text{deg}^{\mathcal{I}}(d, C)$, i.e., $S(v_0, d) = \text{deg}^{\mathcal{I}}(d, C^r)$.

Proof. To see that the algorithm terminates, it is enough to observe that T_{C^r} and $G_{\mathcal{I}}$ are finite and the algorithm consists of nested iterations over the nodes and edges in T_{C^r} and $G_{\mathcal{I}}$. To show that $S(v_0, d) = \text{deg}^{\mathcal{I}}(d, C^r)$, we prove a more general claim:

Claim: After a run of the algorithm, $S(v, e) = \text{deg}^{\mathcal{I}}(e, C^r[v])$ for all $v \in V_{T_{C^r}}$ and $e \in \Delta^{\mathcal{I}}$.

Note first, 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 $v \in V_{T_{C^r}}$.

Induction Base. $\eta(v) = 0$. Then v is a leaf in T_{C^r} . This means that v has no successors and for each $e \in \Delta^{\mathcal{I}}$ there exists a unique *ptgh* h from $T_{C^r}[v]$ to $G_{\mathcal{I}}$ with $h(v) = e$. One can see in Algorithm 4, that the special case where $|\ell_{T_{C^r}}(v)| + k^*(v) = 0$ is treated properly. Otherwise, we have $c = |\ell_{T_{C^r}}(v) \cap \ell_{\mathcal{I}}(e)|$ and $S(v, e) = \frac{c}{|\ell_{T_{C^r}}(v)|}$. Note that this is exactly the value of $h_w(v)$ in Definition 19. Since h is unique, this means that $\text{deg}^{\mathcal{I}}(e, C^r[v]) = S(v, e)$.

Induction Step. $\eta(v) > 0$. Let v_1, \dots, v_k be the children of v in T_{C^r} such that if v_1 is an r -successor of v in T_{C^r} , then e has at least one r -successor in $G_{\mathcal{I}}$. The application of the max operator in line 10, selects for each r -successor v_i of v an r -successor e_i of e in $\Delta^{\mathcal{I}}$ that has the maximum value for $S(v_i, e_i)$, and then is used in the computation of c . Let (v_i, e_i) be the pair representing such a selection for each v_i . Two observations are in order:

- Since v_i is a child of v , it occurs first in the post-order selected in line 1. Therefore, the value of $S(v_i, e_i)$ is computed before the computation of c for (v, e) .
- Since $\eta(v_i) < \eta(v)$, the application of induction hypothesis yields $S(v_i, e_i) = \text{deg}^{\mathcal{I}}(e_i, C^r[v_i])$.

For each $1 \leq i \leq k$, let h_i be a *ptgh* from $T_{C^r}[v_i]$ to $G_{\mathcal{I}}$ such that $h_i(v_i) = e_i$ and $h_{i_w}(v_i) = \text{deg}^{\mathcal{I}}(e_i, C^r[v_i])$. It is easy to see that the mapping $h = h_1 \cup \dots \cup h_k \cup \{(v, e)\}$ is a *ptgh* from $T_{C^r}[v]$ to $G_{\mathcal{I}}$ with $h(v) = e$ and that $h_w(v) = S(v, e)$. Hence, by Definition 20 we have $S(v, e) \in \mathcal{V}^{\mathcal{I}}(e, C^r[v])$. Suppose, however, that $S(v, e) < \max \mathcal{V}^{\mathcal{I}}(e, C^r[v])$. We show that this is not the case by reaching a contradiction.

Since $S(v, e) < \max \mathcal{V}^{\mathcal{I}}(e, C^r[v])$, there exists a *ptgh* h' from $T_{C^r}[v]$ to $G_{\mathcal{I}}$ with $h'(v) = e$ such that $h'_w(v) > h_w(v)$. Looking at h_w in Definition 19, the fact that $h(v) = h'(v)$ implies that the difference must be on the values of $h_w(v_i)$ and $h'_w(v_i)$. More precisely, there must exist at least one successor v_i of v such that $h'_w(v_i) > h_w(v_i)$. Based on this, we distinguish two cases:

- $h'(v_i) \neq h(v_i)$, i.e., the *ptgh* h' maps v_i to a different element in $\Delta^{\mathcal{I}}$. But, if that were the case, then the application of the max operator in line 10 would have chosen $h'(v)$ as the pairing for v_i , instead of e_i .
- $h'(v_i) = h(v_i) = e_i$. This case would contradict the induction hypothesis, since $h'_w(v_i) > h_w(v_i)$ would imply $S(v_i, e_i) < \text{deg}^{\mathcal{I}}(e_i, C^r[v_i])$.

Hence, we obtain by contradiction that $S(v, e) = \max \mathcal{V}^{\mathcal{I}}(e, C^r[v])$ and consequently, $S(v, e) = \text{deg}^{\mathcal{I}}(e, C^r[v])$. Since $S(v_0, d)$ is a particular case, we thus have shown that $S(v_0, d) = \text{deg}^{\mathcal{I}}(d, C^r)$. \square

Missing proofs of Section 5

Proposition 40. Let \mathcal{A} be an ABox. Then, \mathcal{A} is consistent iff there exists a consistent *pre-processing* \mathcal{A}' of \mathcal{A} .

Proof. (\Rightarrow) Let \mathcal{I} be an interpretation such that $\mathcal{I} \models \mathcal{A}$. One can see that for any assertion $\neg\widehat{C}(a)$ which a rule is applicable to, if $\mathcal{I} \models \neg\widehat{C}(a)$ there is a way to apply the rule such that \mathcal{I} also satisfies the new introduced assertion. The case for $\rightarrow_{\neg\exists}$ is clear. For the rule $\rightarrow_{\neg\sqcap}$, if $\mathcal{I} \models \neg\widehat{C}$, then there exists a conjunct \widehat{C}_i such that $\mathcal{I} \models \neg\widehat{C}_i(a)$. This can be the *non-deterministic* choice made by $\rightarrow_{\neg\sqcap}$. Last, for assertions of the form $\neg E_{\sim q}$ and $\neg A$ the applicable rules are $\rightarrow_{\neg\sim}$ and $\rightarrow_{\neg A}$, respectively. Since $\neg E_{\sim q} \equiv E_{\gamma(\sim)q}$ and $\neg A \equiv A_{<1}$, we have that \mathcal{I} satisfies $E_{\gamma(\sim)q}$ and $A_{<1}$.

Thus, since \mathcal{I} satisfies every assertion in \mathcal{A} we can conclude that there exists *pre-processing* \mathcal{A}' such that $\mathcal{I} \models \mathcal{A}'$.

(\Leftarrow) This direction is trivial since $\mathcal{A} \subseteq \mathcal{A}'$. □

Lemma 43. Let \mathcal{A} be an ABox and \mathcal{I} an interpretation such that $\mathcal{I} \models \mathcal{A}$. In addition, let \mathcal{A}' be a *pre-processing* of \mathcal{A} such that $\mathcal{I} \models \mathcal{A}'$. For each $a \in \text{Ind}(\mathcal{A})$, let \mathcal{I}_a be a *tree-shaped* interpretation such that:

- $\mathcal{I}_a \models \mathcal{A}'(a)$,
- there exists a homomorphism φ_a from $G_{\mathcal{I}_a}$ to $G_{\mathcal{I}}$ with $\varphi_a(a^{\mathcal{I}_a}) = a^{\mathcal{I}}$.

Last, let \mathcal{J} be the following interpretation:

- $\Delta^{\mathcal{J}} := \bigcup_{a \in \text{Ind}(\mathcal{A})} \Delta^{\mathcal{I}_a}$,
- $A^{\mathcal{J}} := \bigcup_{a \in \text{Ind}(\mathcal{A})} A^{\mathcal{I}_a}$ for all $A \in \mathbf{N}_C$,
- $r^{\mathcal{J}} := \{a^{\mathcal{I}_a} r b^{\mathcal{I}_b} \mid r(a, b) \in \mathcal{A}\} \cup \bigcup_{a \in \text{Ind}(\mathcal{A})} r^{\mathcal{I}_a}$ for all $r \in \mathbf{N}_R$, and
- $a^{\mathcal{J}} := a^{\mathcal{I}_a}$, for all $a \in \text{Ind}(\mathcal{A})$.

where the sets $\Delta^{\mathcal{I}_a}$ are pair-wise disjoint. Then, $\mathcal{J} \models \mathcal{A}$.

Proof. We start by considering the following mapping from $V_{\mathcal{J}}$ to $V_{\mathcal{I}}$:

$$\varphi^* := \bigcup_{a \in \text{Ind}(\mathcal{A})} \varphi_a$$

Note that since the sets $\Delta^{\mathcal{I}_a}$ are disjoint the mapping is unambiguous. In addition, we have $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ for all $r(a, b) \in \mathcal{A}$, and since $\varphi_a(a^{\mathcal{I}_a}) = a^{\mathcal{I}}$ and $\varphi_b(b^{\mathcal{I}_b}) = b^{\mathcal{I}}$,

this implies that $(\varphi^*(a^{\mathcal{I}_a}), \varphi^*(b^{\mathcal{I}_b})) \in r^{\mathcal{I}}$ for all $(a^{\mathcal{I}_a}, b^{\mathcal{I}_b}) \in r^{\mathcal{J}}$. Then, it is clear that φ^* is a homomorphism from $G_{\mathcal{J}}$ to $G_{\mathcal{I}}$ with $\varphi^*(a^{\mathcal{J}}) = a^{\mathcal{I}}$ for all $a \in \text{Ind}(\mathcal{A})$.

We now show that $\mathcal{J} \models \mathcal{A}'$ and since $\mathcal{A} \subseteq \mathcal{A}'$, this will imply $\mathcal{J} \models \mathcal{A}$. First note that if $r(a, b) \in \mathcal{A}'$, then it is also in \mathcal{A} . By construction $(a^{\mathcal{I}_a}, b^{\mathcal{I}_b}) \in r^{\mathcal{J}}$ and since $a^{\mathcal{J}} = a^{\mathcal{I}_a}$ for all $a \in \text{Ind}(\mathcal{A})$, it follows that $(a^{\mathcal{J}}, b^{\mathcal{J}}) \in r^{\mathcal{J}}$. Therefore, it remains to show that each concept assertion in \mathcal{A}' is satisfied by \mathcal{J} .

We first show that $\mathcal{J} \models \mathcal{A}'^+$. Let a be any individual name in \mathcal{A} and $\widehat{C}(a) \in \mathcal{A}'^+$. Since $\mathcal{I}_a \models \widehat{C}(a)$, we have $a^{\mathcal{I}_a} \in \widehat{C}^{\mathcal{I}_a}$ and the application of Theorem 14 yields a τ -homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}_a}$ with $\phi(v_0) = a^{\mathcal{I}_a}$. Since we require the interpretation \mathcal{I}_a to be *tree-shaped*, it is not hard to see that $\phi(v) = a^{\mathcal{I}_a}$ only if $v = v_0$. We then show that ϕ is also τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$. Note that since $\mathcal{I}_a \subseteq \mathcal{J}$, Condition 1 from Definition 13 is obviously satisfied by ϕ and \mathcal{J} . Hence, it remains to show that Condition 2 is also satisfied.

Let $v \in T_{\widehat{C}}$ and $E_{\sim q} \in \widehat{\ell}_{T_{\widehat{C}}}(v)$, we distinguish two cases:

- $v = v_0$. By the relationship that exists between $\tau\mathcal{EL}(m)$ concept descriptions and $\tau\mathcal{EL}(m)$ descriptions trees (see Section 3), we have that $E_{\sim q}$ is top-level atom of \widehat{C} . Therefore, $a^{\mathcal{I}_a} \in [E_{\sim q}]^{\mathcal{I}_a}$ and $a^{\mathcal{I}} \in [E_{\sim q}]^{\mathcal{I}}$.

We now distinguish between $\sim \in \{>, \geq\}$ or $\sim \in \{<, \leq\}$. Then Lemma 24 can be applied accordingly to obtain $a^{\mathcal{J}} \in [E_{\sim q}]^{\mathcal{J}}$, since on the one hand, we have $\mathcal{I}_a \subseteq \mathcal{J}$, and on the other hand φ^* is homomorphism from $G_{\mathcal{J}}$ to $G_{\mathcal{I}}$ with $\varphi^*(a^{\mathcal{J}}) = a^{\mathcal{I}}$.

- $v \neq v_0$. As said before, we have $\phi(v) = e$ with $e \neq a^{\mathcal{I}_a}$ and $e \in \Delta^{\mathcal{I}_a}$. Since $G_{\mathcal{I}_a}$ is a tree, the reachable elements from e in $\Delta^{\mathcal{J}}$ through role relations are exactly the same as in $\Delta^{\mathcal{I}_a}$. Then it is easy to see that $\text{deg}^{\mathcal{I}_a}(e, E) = \text{deg}^{\mathcal{J}}(e, E)$, and since $e \in [E_{\sim q}]^{\mathcal{I}_a}$ we also have $e \in [E_{\sim q}]^{\mathcal{J}}$.

Thus, ϕ is τ -homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{J}}$ with $\phi(v_0) = a^{\mathcal{J}}$. The application of Theorem 14 yields $a^{\mathcal{J}} \in \widehat{C}^{\mathcal{J}}$. Since we have chosen a and $\widehat{C}(a)$ arbitrarily, we can thus conclude that $\mathcal{J} \models \mathcal{A}'^+$.

We now turn into \mathcal{A}'^- , i.e., we prove $\mathcal{J} \models \neg\widehat{C}(a)$ for each assertion $\neg\widehat{C}(a) \in \mathcal{A}'^-$. We do not consider assertions where \widehat{C} is of the form $\widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_n$. The application of rule $\rightarrow_{\neg\sqcap}$ ensures that we always have $\neg\widehat{C}_i(a) \in \mathcal{A}'^-$ for some \widehat{C}_i which is not a conjunction and moreover, $a^{\mathcal{J}} \notin [\widehat{C}_i]^{\mathcal{J}}$ implies $a^{\mathcal{J}} \notin \widehat{C}^{\mathcal{J}}$. We use induction on the *role-depth* of \widehat{C} :

1. $\text{rd}(\widehat{C}) = 0$. Then, $\neg\widehat{C}$ is of the form $\neg E_{\sim q}$ or $\neg A$. The application of the rules $\rightarrow_{\neg\sim}$ and $\rightarrow_{\neg A}$ yields, $E_{\gamma(\sim)q}(a) \in \mathcal{A}'^+$ and $A_{<1}(a) \in \mathcal{A}'^+$. Since $\neg E_{\sim q} \equiv E_{\gamma(\sim)q}$ and $\neg A \equiv A_{<1}$ (see Section 4.1 and Proposition 8) and since $\mathcal{J} \models \mathcal{A}'^+$, this yields $a^{\mathcal{J}} \notin \widehat{C}^{\mathcal{J}}$.

2. $\text{rd}(\widehat{C}) > 0$. Then, $\neg\widehat{C}$ is of the form $\neg\exists r.\widehat{D}$ and $(\neg\exists r.\widehat{D})(a) \in \mathcal{A}'$. Assume that $(a^{\mathcal{J}}, d) \in r^{\mathcal{J}}$ for some $d \in \Delta^{\mathcal{J}}$. We have two cases:

- $d = b^{\mathcal{J}}$ for some $b \in \text{Ind}(\mathcal{A})$. By construction of \mathcal{J} we have $r(a, b) \in \mathcal{A}$. Then, the application of rule $\rightarrow_{\neg\exists}$ implies that $\neg\widehat{D}(b) \in \mathcal{A}'$. Obviously, $\text{rd}(\widehat{D}) < \text{rd}(\exists r.\widehat{D})$ and therefore the application of induction hypothesis yields $b^{\mathcal{J}} \notin \widehat{D}^{\mathcal{J}}$.
- $d \neq b^{\mathcal{J}}$ for all $b \in \text{Ind}(\mathcal{A})$. Then, by construction of \mathcal{J} we have $d \in \Delta^{\mathcal{I}_a}$. Moreover, $(\neg\exists r.\widehat{D})(a) \in \mathcal{A}'(a)$ and since $\mathcal{I}_a \models \mathcal{A}'(a)$, we have $d \notin \widehat{D}^{\mathcal{I}_a}$. Now, suppose that $d \in \widehat{D}^{\mathcal{J}}$. By Theorem 14 there exists a τ -homomorphism ϕ from $T_{\widehat{D}}$ to $G_{\mathcal{J}}$ with $\phi(v_0) = d$. But, if that is the case, by the disjointness assumptions made to build \mathcal{J} and the fact that $G_{\mathcal{I}_a}$ is a tree, we would have that it is also a τ -homomorphism from $T_{\widehat{D}}$ to $G_{\mathcal{I}_a}$, contradicting the fact that $d \notin \widehat{D}^{\mathcal{I}_a}$. Thus, $d \notin \widehat{D}^{\mathcal{J}}$.

We just have shown that for each r -successor d of $a^{\mathcal{J}}$ it is the case that $d \notin \widehat{D}^{\mathcal{J}}$. Hence, $a^{\mathcal{J}} \notin [\exists r.\widehat{D}]^{\mathcal{J}}$.

Thus, $\mathcal{J} \models \mathcal{A}'^-$ and consequently $\mathcal{J} \models \mathcal{A}'$. □

Lemma 42. Let \mathcal{A} be a *single-element* ABox and \mathcal{I} an interpretation such that $\mathcal{I} \models \mathcal{A}$. In addition, let \mathcal{J} be the bounded model of \mathcal{A}^+ constructed in Lemma 37. Then, there exists a *tree-shaped* interpretation \mathcal{K} such that:

1. $\mathcal{K} \models \mathcal{A}$,
2. there exists a homomorphism φ from $G_{\mathcal{K}}$ to $G_{\mathcal{I}}$ such that $\varphi(a^{\mathcal{K}}) = a^{\mathcal{I}}$, and
3. $|\Delta^{\mathcal{K}}| \leq |\Delta^{\mathcal{J}}| \times n$, where:

$$n := \begin{cases} 1, & \text{if } \mathcal{A}^- = \emptyset \\ \prod_{\neg\widehat{C}(a) \in \mathcal{A}^-} s(\widehat{C}), & \text{otherwise.} \end{cases}$$

Proof. Let \mathcal{I} be an interpretation such that $\mathcal{I} \models \mathcal{A}$ and \mathcal{J} the interpretation constructed in Lemma 37 with respect to \mathcal{A}^+ . We start by recalling some elements from the proof of Lemma 37 which will be used to prove our claim.

- ϕ is a τ -homomorphism from $\widehat{G}(\mathcal{A})$ to $G_{\mathcal{I}}$ with $\phi(a) = a^{\mathcal{I}}$ for all $a \in \text{Ind}(\mathcal{A})$.
- φ is a homomorphism from $V_{\mathcal{J}}$ to $V_{\mathcal{I}}$ with $\varphi(v) = \phi(v)$ for all $v \in V_{\mathcal{A}}$.

Let $\#(\mathcal{A})$ denote the number of concept assertions occurring in \mathcal{A} . We prove our claim by induction on the number $\#(\mathcal{A}^-)$.

Induction Base. $\#(\mathcal{A}^-) = 0$. Then, we have that $\mathcal{A}^- = \emptyset$. Therefore, $\mathcal{A} = \mathcal{A}^+$ is a $\tau\mathcal{EL}(\text{deg})$ ABox. We choose \mathcal{K} to be the interpretation \mathcal{J} . Then we have $\mathcal{K} \models \mathcal{A}$ and obviously $|\Delta^{\mathcal{K}}| \leq |\Delta^{\mathcal{J}}|$. Since $a^{\mathcal{J}} = a$, this means that $\varphi(a^{\mathcal{K}}) = a^{\mathcal{J}}$ and Condition 2 is satisfied. Moreover, since \mathcal{A} contains a single element and no role assertions, we have that $\widehat{G}(\mathcal{A})$ is a tree and by construction $G_{\mathcal{K}}$ is also a tree.

Thus, we have shown our claims for the chosen interpretation \mathcal{K} .

Induction Step. We assume that the claim holds for all *single-element* ABoxes \mathcal{B} with $0 \leq \#(\mathcal{B}^-) < k$ and show that it holds for an ABox \mathcal{A} with $\#(\mathcal{A}^-) = k$.

Consider the bounded model \mathcal{J} of \mathcal{A}^+ (as in the base case). We know that $\mathcal{J} \models \mathcal{A}^+$. However, \mathcal{J} does not necessarily satisfies \mathcal{A}^- since the assertions from \mathcal{A}^- were not taken into account to obtain it. The idea for the rest of the proof is to build an ABox $\mathcal{A}_{\mathcal{J}}$ which reflects the structure of \mathcal{J} . Then, we will consider a *pre-processing* \mathcal{A}' of $\mathcal{A}_{\mathcal{J}} \cup \mathcal{A}^-$ based on \mathcal{I} and show how to use it to extend \mathcal{J} into an interpretation \mathcal{K} that satisfies our claims.

Let $G_{\mathcal{J}}$ be the description graph associated to \mathcal{J} , from the base case we know that it is a tree. The ABox $\mathcal{A}_{\mathcal{J}}$ is built as follows:

$$\mathcal{A}_{\mathcal{J}} := \bigcup_{\substack{b \in V_{\mathcal{J}} \\ A \in \ell_{\mathcal{J}}(b)}} \{A(b)\} \cup \bigcup_{brc \in E_{\mathcal{J}}} \{r(b, c)\}$$

where $\text{Ind}(\mathcal{A}_{\mathcal{J}}) = \Delta^{\mathcal{J}} = V_{\mathcal{J}}$.

We name the element $a^{\mathcal{J}}$ in \mathcal{J} as a in the new ABox $\mathcal{A}_{\mathcal{J}}$. In addition, for each individual $b \in \text{Ind}(\mathcal{A}_{\mathcal{J}})$ with $b \neq a$, we make $b^{\mathcal{J}} = b$. Then, since all the concept assertions in $\mathcal{A}_{\mathcal{J}}$ are of the form $A(a)$ with $A \in \mathbf{N}_{\mathcal{C}}$, it is easy to see that $\mathcal{J} \models \mathcal{A}_{\mathcal{J}}$. We now use the homomorphism φ to extend the interpretation of \mathcal{I} to the individual names in $\mathcal{A}_{\mathcal{J}}$ as: $b^{\mathcal{I}} = \varphi(b^{\mathcal{J}})$. Since $\varphi(a^{\mathcal{J}}) = a^{\mathcal{I}}$, this means that the element $a^{\mathcal{I}}$ does not change. Hence, φ is a homomorphism from $G_{\mathcal{J}}$ to $G_{\mathcal{I}}$ with $\varphi(b^{\mathcal{J}}) = b^{\mathcal{I}}$ for all $b \in \text{Ind}(\mathcal{A}_{\mathcal{J}})$. Then, for any $A(b) \in \mathcal{A}_{\mathcal{J}}$ we have $b^{\mathcal{J}} \in A^{\mathcal{J}}$ and using φ we also have $b^{\mathcal{I}} \in A^{\mathcal{I}}$. A similar argument yields $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ for all $r(a, b) \in \mathcal{A}_{\mathcal{J}}$. Therefore $\mathcal{I} \models \mathcal{A}_{\mathcal{J}}$ and consequently, $\mathcal{I} \models \mathcal{A}_{\mathcal{J}} \cup \mathcal{A}^-$.

By Remark 41 there exists a *pre-processing* \mathcal{A}' of $\mathcal{A}_{\mathcal{J}} \cup \mathcal{A}^-$ such that $\mathcal{I} \models \mathcal{A}'$ and, in addition, $\text{Ind}(\mathcal{A}_{\mathcal{J}}) = \text{Ind}(\mathcal{A}')$. For each individual $b \in \text{Ind}(\mathcal{A}_{\mathcal{J}})$ let \mathcal{A}'_b be the ABox:

$$\mathcal{A}'_b := \bigcup_{E_{\gamma(\sim)_q}(b) \in \mathcal{A}'} \{E_{\gamma(\sim)_q}(b)\} \cup \bigcup_{\neg\exists r.\widehat{D}(b) \in \mathcal{A}'} \{\neg\exists r.\widehat{D}(b)\}$$

Here, $E_{\gamma(\sim)_q}(b)$ is an assertion that results from the application of rule \rightarrow_{\sim} or rule $\rightarrow_{\neg A}$. For the rule $\rightarrow_{\neg A}$, we consider $A_{<1}$ as $E_{\gamma(\sim)_q}$, since it is obtained from

$\neg A$ and $A \equiv A_{\geq 1}$. Then, we define the ABox \mathcal{A}_b as:

$$\mathcal{A}_b := \mathcal{A}'_b \cup \bigcup_{A(b) \in \mathcal{A}'} \{A(b)\}$$

The idea now is to show that $\#(\mathcal{A}_b^-) < \#(\mathcal{A}^-)$ and then to apply induction hypothesis on \mathcal{A}_b . As a first step one can show that $\#(\mathcal{A}'_b) \leq \#(\mathcal{A}^-)$. The main reasons why this is possible are that \mathcal{J} is *tree-shaped* and $\Delta^{\mathcal{J}} = \text{Ind}(\mathcal{A}_{\mathcal{J}})$. For example, if we consider the individual b in $\mathcal{A}_{\mathcal{J}}$ corresponding to the root of $G_{\mathcal{J}}$, it is not hard to see that for each concept assertion $\alpha(b)$ in \mathcal{A}^- , either it is in \mathcal{A}'_b or there is at most one concept assertion in \mathcal{A}'_b obtained from $\alpha(b)$ through the pre-processing rule applications. Moreover, no assertion in \mathcal{A}'_b can be obtained in a different way. Consequently, $\#(\mathcal{A}'_b) \leq \#(\mathcal{A}^-)$. Taking this as the base case, the same can be shown for the rest of the individuals by induction on the depth² of each node in $V_{\mathcal{J}}$.

Consider the set (possibly empty) $B \subseteq \text{Ind}(\mathcal{A}_{\mathcal{J}})$ such that $b \in B$ **if, and only if**, \mathcal{A}'_b contains at least one assertion of the form $E_{\gamma(\sim)_q}(b)$. Then, for each $b \in B$ we have $\#(\mathcal{A}_b^-) < \#(\mathcal{A}^-)$ and therefore, the application of induction hypothesis to \mathcal{A}_b yields a *tree-shaped* interpretation \mathcal{I}_b and a homomorphism φ_b from $G_{\mathcal{I}_b}$ to $G_{\mathcal{I}}$ such that: $\varphi_b(b^{\mathcal{I}_b}) = b^{\mathcal{I}}$ and $\mathcal{I}_b \models \mathcal{A}_b$.

For the individuals $b \in \text{Ind}(\mathcal{A}_{\mathcal{J}})$ such that $b \notin B$, we consider the *single-pointed* interpretation $\mathcal{I}_b = (\{b\}, \cdot^{\mathcal{I}_b})$ which is the restriction of \mathcal{J} to $\{b\}$. Note then that \mathcal{A}_b can only contain assertions of the form $\neg \exists r. \widehat{D}(b)$ or assertions from $\mathcal{A}_{\mathcal{J}}$. Since $\mathcal{J} \models \mathcal{A}_{\mathcal{J}}$, it is clear that $\mathcal{I}_b \models \mathcal{A}_b$ and φ_b with $\varphi_b(b^{\mathcal{I}_b}) = b^{\mathcal{I}}$ is a homomorphism from $G_{\mathcal{I}_b}$ to $G_{\mathcal{I}}$.

Now, for assertions of the form $\neg E_{\sim_q}(b)$ and $\neg A(b)$ in \mathcal{A}' , the application of the rules \rightarrow_{\sim} and $\rightarrow_{\neg A}$ ensures that $E_{\gamma(\sim)_q}(b)$ and $A_{<1}(b)$ are in \mathcal{A}_b . Since $\mathcal{I}_b \models \mathcal{A}_b$, $\neg E_{\sim_q} \equiv E_{\gamma(\sim)_q}$ and $\neg A \equiv A_{<1}$, we have $\mathcal{I}_b \models \neg E_{\sim_q}(b)$ and $\mathcal{I}_b \models \neg A(b)$. In addition, for assertions $\neg \widehat{C}(b) \in \mathcal{A}'$ where \widehat{C} is of the form $\widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_n$, by the application of rule $\rightarrow_{\neg \sqcap}$ we know that there exists some \widehat{C}_i such that $\neg \widehat{C}_i(b) \in \mathcal{A}'$. Since $\neg \widehat{C}_i$ is of one of the previous considered forms, we also have $\mathcal{I}_b \models \neg \widehat{C}(b)$.

Hence, it follows that $\mathcal{I}_b \models \mathcal{A}'(b)$ for all $b \in \text{Ind}(\mathcal{A}_{\mathcal{J}})$. Therefore, considering the sets $\Delta^{\mathcal{I}_b}$ pair-wise disjoint, for all $b \in \text{Ind}(\mathcal{A}_{\mathcal{J}})$, we can apply Lemma 43 to $\mathcal{A}_{\mathcal{J}} \cup \mathcal{A}^-$ to obtain an interpretation \mathcal{K} such that $\mathcal{K} \models \mathcal{A}_{\mathcal{J}} \cup \mathcal{A}^-$. However, we still need to show that $\mathcal{K} \models \mathcal{A}^+$. We use again the same idea that shows, in Lemma 32 and Lemma 37, that ϕ_{id} is still τ -homomorphism after extending \mathcal{I}_0 into \mathcal{J} .

More precisley, in our case we know that ϕ_{id} is τ -homomorphism from $\widehat{G}(\mathcal{A}^+)$ to $G_{\mathcal{J}}$ by construction of \mathcal{J} . Since the construction of \mathcal{K} in Lemma 43 is based

²The depth of a node in a tree is the length of the path from the root of the tree to the node. The root of the tree has depth 0.

on a pre-processing of $\mathcal{A}_{\mathcal{J}} \cup \mathcal{A}^-$, it is not hard to see that renaming $a^{\mathcal{I}_a}$ to a for each $a \in \text{Ind}(\mathcal{A})$ makes $\mathcal{J} \subseteq \mathcal{K}$. Then we want to show that ϕ_{id} is also τ -homomorphism from $\widehat{G}(\mathcal{A}^+)$ to $G_{\mathcal{K}}$, which amounts to show that Condition 2 of Definition 13 is satisfied. Let v be any node from $\widehat{G}(\mathcal{A}^+)$ with $E_{\sim q} \in \widehat{\ell}_{\mathcal{A}^+}(v)$. As in Lemma 32, we make the same case distinction on whether $\sim \in \{>, \geq\}$ or $\sim \in \{<, \leq\}$. Since we know that $\phi_{id}(v) \in [E_{\sim q}]^{\mathcal{J}}$, the case for $\{>, \geq\}$ is obvious because $\mathcal{J} \subseteq \mathcal{K}$.

For $\sim \in \{<, \leq\}$, we have $\phi(v) \in [E_{\sim q}]^{\mathcal{I}}$ and by Lemma 43 φ^* is a homomorphism from $G_{\mathcal{K}}$ to $G_{\mathcal{I}}$. Hence, if we had $\varphi^*(v) = \phi(v)$, we could apply Lemma 24 to show that $v \in [E_{\sim q}]^{\mathcal{K}}$. Now, note that by construction of \mathcal{J} in Lemma 37 and by construction of $\mathcal{A}_{\mathcal{J}}$, v is actually an individual name in $\mathcal{A}_{\mathcal{J}}$. Therefore, $\varphi^*(v) = \varphi_v(v) = \varphi_v(v^{\mathcal{I}_v}) = v^{\mathcal{I}}$. We defined above $v^{\mathcal{I}} = \varphi(v)$ and since $\varphi(v) = \phi(v)$ for all $v \in V_{\mathcal{A}^+}$, we can conclude that $\varphi^*(v) = \phi(v)$ and consequently, $v \in [E_{\sim q}]^{\mathcal{K}}$. Hence, ϕ_{id} is τ -homomorphism from $\widehat{G}(\mathcal{A}^+)$ to $G_{\mathcal{K}}$ with $\phi_{id}(a) = a^{\mathcal{K}}$. Thus, $\mathcal{K} \models \mathcal{A}^+$.

To see that \mathcal{K} is *tree-shaped*, note first that \mathcal{J} and each interpretation \mathcal{I}_b are also *tree-shaped*. Then by the construction of \mathcal{K} in Lemma 43 one can easily see that $G_{\mathcal{K}}$ is a tree.

Regarding the size of \mathcal{K} , if $b \notin B$ we have $|\Delta^{\mathcal{I}_b}| = 1$, otherwise \mathcal{I}_b is obtained by the application of induction hypothesis to \mathcal{A}_b . Let \mathcal{J}_b be the bounded model for \mathcal{A}_b^+ constructed in Lemma 37. Then,

$$|\Delta^{\mathcal{I}_b}| \leq |\Delta^{\mathcal{J}_b}| \times \prod_{\neg \widehat{D}(b) \in \mathcal{A}_b^-} \mathfrak{s}(\widehat{D}) \quad (1)$$

A closer look at \mathcal{A}_b^+ shows that it only contains assertions of the form $E_{\gamma(\sim)q}(b)$ or $A(b)$, with $A(b) \in \mathcal{A}_{\mathcal{J}}$ and $A \in \text{N}_{\mathcal{C}}$. Since, in addition, it only contains one individual and no role assertions, the construction of \mathcal{J}_b in Lemma 37 yields:

$$|\Delta^{\mathcal{J}_b}| \leq \sum_{E_{\gamma(\sim)q}(b) \in \mathcal{A}_b^+} \mathfrak{s}(E_{\gamma(\sim)q})$$

Furthermore, $|E_{\gamma(\sim)q}| > 1$ allows to transform this inequality into the following:

$$|\Delta^{\mathcal{J}_b}| \leq \prod_{E_{\gamma(\sim)q}(b) \in \mathcal{A}_b^+} \mathfrak{s}(E_{\gamma(\sim)q}) \quad (2)$$

It is not hard to see that for each assertion of the form $\neg \widehat{D}(b) \in \mathcal{A}_b^-$, the concept \widehat{D} is a sub-description of a concept \widehat{C} such that $\neg \widehat{C}(a) \in \mathcal{A}^-$. In addition, each threshold concept $E_{\gamma(\sim)q} \in \mathcal{A}_b^+$ is obtained during the pre-processing by the application of the rule \rightarrow_{\sim} to the concept $\neg E_{\sim q}$, and moreover, $E_{\sim q}$ is also a sub-description of a concept \widehat{C} with $\neg \widehat{C}(a) \in \mathcal{A}^-$. Note that since the pre-processing is applied to $\mathcal{A}_{\mathcal{J}}$ and $G_{\mathcal{J}}$ is a tree, there is at most one of these concepts in \mathcal{A}_b^+ for

each concept assertion $\neg\widehat{C}(a) \in \mathcal{A}^-$. Hence, since $\#(\mathcal{A}'_b) \leq \#(\mathcal{A}^-)$, we combine (1) and (2) to obtain:

$$|\Delta^{\mathcal{I}_b}| \leq \prod_{\neg\widehat{C}(a) \in \mathcal{A}^-} s(\widehat{C})$$

Finally, $|\Delta^{\mathcal{J}}| = |\text{Ind}(\mathcal{A}_{\mathcal{J}})|$ and the construction of $\Delta^{\mathcal{K}}$ yield:

$$|\Delta^{\mathcal{K}}| = \sum_{b \in \text{Ind}(\mathcal{A}_{\mathcal{J}})} |\Delta^{\mathcal{I}_b}| \leq |\Delta^{\mathcal{J}}| \times n$$

□