Semantic Web 0 (0) 1
IOS Press

Towards Explainable Knowledge Graph Embeddings by Respecting Logical Commitments

Mena Leemhuis [a,*], Özgür Lütfü Özçep [a] and Diedrich Wolter [b]
[a] *Institute of Information Systems, University of Lübeck, Lübeck, Germany*
E-mails: leemhuis@ifis.uni-luebeck.de, oezcep@ifis.uni-luebeck.de
[b] *Information Systems and Applied Computer Sciences, University of Bamberg, Bamberg, Germany*
E-mail: diedrich.wolter@uni-bamberg.de

Abstract. Knowledge graph embeddings (KGEs) can be seen as opportunity to integrate machine learning (ML) with knowledge representation and reasoning. In KGEs, concepts and relations are represented by geometric structures that are induced by ML. Explicit representation of concepts and relations empowers reasoning which can augment ML. This additional symbolic layer linked to ML models is widely advocated to foster explainability. However, symbolic reasoning and ML need to be aligned beyond the level of concept symbols in order to obtain explanations of what was actually learned. We characterize explainability as the alignment of reasoning in two agents, which calls for a rigorous understanding of reasoning grounded in KGEs. The desired alignment can be achieved by investigating the *logical commitments* made in KGE approaches by identifying models of logics that are aligned with ML models. Not until logical commitments of KGEs are aligned with common modes of reasoning, explanations for learnt models can be generated that are both effective and semantically congruent with what has been learnt. We critically review existing approaches to KGEs and then analyze a cone-based model capable of grasping full negation, a property common to symbolic reasoning but not yet captured in current KGE approaches. To this end, we propose orthologics as basis to characterize cone-based models.

Keywords: logical commitment, knowledge graph embedding, orthologic, cone

1. Introduction

There are roughly two lines of research in explainable AI (XAI), the first grounded in the knowledge representation and reasoning (KRR) community, the other grounded in the machine learning (ML) community. Common to both is the aim of explaining (under a very wide reading) *decisions* an intelligent agent takes in a human comprehensible way [1]. One example for explanations in the former community is *pinpointing* which justifies, i.e., explains, a conclusion by pinpointing to a (minimal) set of premises from which the conclusion follows proof-theoretically [2, 3]. Examples for explanations in the ML community are post-hoc justifications for predictions of neural networks based on feature importance as in the systems SHAPE [4] or LIME [5].

In the KRR community, the focus of explanations is couched in logical frameworks and mainly refers to their notions of entailment or derivation. In the ML community, explanations are mostly of non-logical nature. However, there seems to be an agreement [6] that in the area of knowledge graph embedding (KGE) [7], logic plays or at

*Corresponding author. E-mail: leemhuis@ifis.uni-luebeck.de.

least must play an important role in explanations. The line of argument is that knowledge graphs (KGs) as sets of triples $(s\ R\ o)$ (s stands in R-relation to o) already present qualitative relational structures that are tailored towards a treatment in a logic. So one might, e.g., add background knowledge to KGs by formulating logical axioms in an ontology that constraints the "meanings" of the relations R and the objects s, o to the intended ones. So this line of argument suggests that we can "import" the good properties of logics w.r.t. explainability to the realm of ML.

Knowledge graph embeddings present an attractive setting for learning since they do not rely on a pre-defined grounding, for example, needed for logic-based ML approaches such as inductive logic programming. Instead, embeddings reveal similarities of objects, concepts, and relations in geometrical terms.

However, one cannot simply "import" desirable logic structures into an existing KGE, but one has to specifically develop a KGE to be compatible with a selected set of logic features. The selection of logic features is no cherry-picking as there is usually a trade-off between rich logic structures and efficient learning. The choice of a particular approach may further induce logic properties that are not easy to discover. In this paper, we introduce the term *logical commitment* to refer to the choice of logic properties underlying a KGE, both explicitly by constructing an KGE method for a given logic or implicitly by the logic properties a chosen KGE implies. We review selected KGEs with respect to the logical commitments they make and show how we can advance logical commitment towards fundamental logic features with a cone-based embedding that can be characterized by an orthologic. Our work is in line with recent publications in KGE which explicitly state logic properties for KGEs [8–11]. Any logical commitment in context of KGE is related to some *geometric commitments* since KGEs embed objects in a continuous vector space and thus all logic operations are grounded in some geometric operations. Balancing between KGEs that are efficient from a ML perspective and well-aligned with logic structures used when explaining what has been learnt thus comprises investigations of geometric and logic structures. The contribution of this paper is to advocate the formal treatment and understanding of logic reasoning with knowledge graphs. On a technical level, we analyze selected KGE approaches with respect to their logical commitment. We discuss the utility of cone-based geometric models for grasping a richer collection of logic features that seems necessary to advance reasoning in knowledge graphs. In our main line of arguments we claim that progress towards explainable knowledge graphs hinges on a sufficiently rich and, most importantly, well-understood logic characterization. While design and assessment of explainability ultimately requires an understanding of human reasoning, human cognition is beyond the scope of this paper (an attempt at describing human reasoning formally and thereby complementing our work can for example be found in [12]). We restrict ourselves to a technical definition of explainability based on query-answering that involves two agents. Beyond this formal treatment we and parts of the KGE community draw motivation from Gärdenfors' theory of *conceptual spaces* [13] that postulates that human reasoning can be approached as a form of geometric reasoning. Also in this sense, knowledge graph embeddings can be seen as opportunity to integrate machine learning with knowledge representation and reasoning to build trustworthy and explainable systems.

The remainder of this paper is structured as follows. In Section 2 we introduce the term logical commitment and outline a roadmap to achieve an understanding of logical commitments made. Section 3 then identifies logical commitments made by selected KGE approaches. Section 4 discusses the importance of cone-based geometric models and reviews the level of logical commitments from the perspective of orthologics. The paper then concludes by a brief summary.

2. Logical Commitments

This paper aims to coin the term *logical commitment* as an advancement of *ontological commitments* used in context of discussing knowledge representations [14, 15]. Ontological commitments are the views taken on the world when designing a knowledge representation. They primarily address the concept level of knowledge representation, i.e., the choice of basic concepts and how other concepts are composed from them. We may adopt different ontological commitments when constructing representations, for example we may model cats as subclasses of *pets* distinct from *wild animals*, where others might model them as subclasses of *mammals* and not incorporate a distinction between *pets* and *wild animals* at all. Our ontological commitment crucially determines what kind of facts a knowledge representation allows us to capture and to reason about. Yet, reasoning capabilities hinge only in parts on the choice of concepts. Beyond the concept level, reasoning capabilities are determined by the language to compose statements and the calculus to draw conclusions—put differently, they are determined by *logic*. Researchers in KRR have developed a rich repertoire of logics that support various forms of reasoning, most of them providing some unique features. In order to arrive at meaningful conclusions for a given set of facts, an appropriate logic apparatus has to be chosen—similar to how an appropriate choice of concepts has to be made. Logics may differ in fundamental aspects such as the choice of validity semantics, in the set of operators and quantifiers, but they may also differ in nuances like specific axioms that hold or do not hold. While it may be argued whether authors using the term *ontological commitment* already meant it to encompass the choice of a particular logic, we argue that it is helpful to coin a dedicated term in order to analyze, what we believe, is at the core of explainable knowledge representations: an understanding of consequences that can be drawn from a set of facts. From a practical point of view, both terms a interrelated: A simple set of concepts may call for a rich logic to capture interesting facts, whereas a clever choice of concepts may avoid the need of deep logic reasoning for drawing useful conclusions [16].

Logical commitments explicate the choice of logic on the level of semantics of operators and other elements of the language. We may distinguish certain levels of logical commitments which we have portrayed in Figure 1: On the basic level, a designer has to commit to some validity semantics. The basic task is to define the notion of an extension of a concept. This includes stating under which circumstances an individual is considered to be an element of a concept. In order to capture semantics beyond the classical ones we also consider associating an anti-extension of a concept as identifying the logical commitment on this level. (For an application of the notion of an anti-extension in the concrete case of the truth concept see [17]). Using this dual approach also acknowledges the recent trend in (proof-theoretical) semantics known as bilateralism [18]. An entry "binary-total" then means that the extension of a concept is a classical set and its anti-extension is given by set complement. An entry "binary-partial" means that both, extension and anti-extension, are given as classical sets, but they do not necessarily cover the whole domain.

On the second level, commitments to logic operators ($\wedge$, $\vee$, $\neg$, $\ldots$) are made, deciding which are supported and how their semantics is defined. As we will see, contemporary KGEs only support a subset of the operators and there are good reasons to subscribe to non-Boolean algebras for defining semantics of them. In particular, the kind of non-Boolean algebras we are interested in are those termed *(algebraic) ortholattices*. These are defined by the following axioms:

- $a \vee a = a$, $a \wedge a = a$. (idempotence)
- $a \vee b = b \vee a$, $a \wedge b = b \wedge a$. (commutativity)
- $(a \vee b) \vee c = a \vee (b \vee c)$, $(a \wedge b) \wedge c = a \wedge (b \wedge c)$ (associativity)
- $a \vee (a \wedge b) = a$, $a \wedge (a \vee b) = a$ (absorption)
- $a \wedge \bot = \bot$, $a \vee \bot = a$, $a \vee \top = \top$, $a \wedge \top = a$,
- $\neg\neg a = a$ (double negation elimination)
- $\bot = a \wedge \neg a$ (intuitionistic absurdity)
- $\neg(a \vee b) = \neg a \wedge \neg b$, $\neg(a \wedge b) = \neg a \vee \neg b$ (De Morgan)

As can be seen, ortholattices are a weakening of Boolean algebras.

On the third level, commitments to quantifiers and quantifier-like constructs are made. With quantifier-like constructs we refer to relations or operators that introduce new views such as modal operators or roles. Algebraically, quantifiers can be handled by extending the structures of the second level with additional functions. For example, extending Boolean algebras with normal additive functions $f_\exists$ leads to BAOs (Boolean algebras with operators) [19, 20]. The functions $f_\exists$ represent existentials $\exists$, normality ensuring that $f_\exists(\bot) = \bot$ and additivity ensuring that $f_\exists$ distributes over disjunction $\vee$. Similarly, one can extend ortholattices with functions representing (generalized) quantifiers [21].

On a fourth level, a specific logic calculus may be chosen. In the context of this work we assume one strives for a sound and complete logic calculus that mirrors the commitments of earlier stages. (Regarding soundness and completeness consider also the discussion after Definition 1.) The type of calculus may also depend on the kind of application or, more concretely, inference service that one wants to support. We will mainly deal with the inference service of query answering and define the notion of explainability along this notion.

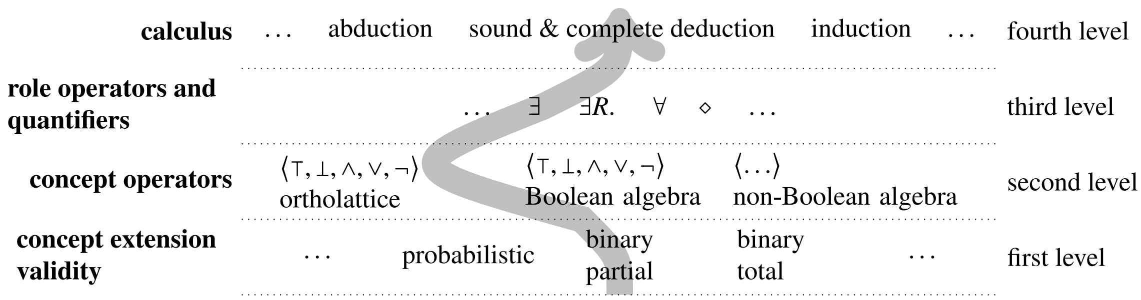

Fig. 1. Different levels of logical commitments, implicit or explicit choices in design of a knowledge representation. The arrow indicates a possible combination of ingredients.

While it is possible to add further levels to address higher-order constructs as well, we will only discuss the levels described here in context of KGEs. Although distinct levels are enumerated, this is not meant to suggest one level would be more important than another. Rather, these levels present partially independent dimensions in which we choose ingredients for composing a logic.

It may not be clear which composition of logic ingredients is best-suited to build an explainable system, but from a methodological perspective on designing explainable systems it is highly desirable to identify the logical commitments inherent to some approach. From a perspective of knowledge representation and reasoning we define explainability as follows:

Definition 1. *We say an agent A_1 is* explainable to agent A_2 with respect to the inference service of query answering for a query class $\mathbf{Q}$ *if there are*

- *a logic L containing sub-logics L_1, L_2 where L_1 is supported by the agent A_1 and L_2 by agent A_2;*
- *a knowledge base KB over L_1;*
- *a computable transformation $\tau_{2\to1}$ of L_2-queries $Q_i \in \mathbf{Q}$ to queries $\tau_{2\to1}(Q_i) \in L_1$;*
- *and a computable transformation $\tau_{1\to2}$ from answers in L_1 to query answers in L_2*

such that

$$\text{answers}(A_2, Q_i) = \tau_{1\to2}(\text{answers}(\text{KB}, \tau_{2\to1}(Q_i)).$$

As an example for explainability in the above sense think of agent A_2 querying agent A_1 and being informed that "all As are Bs" as well as "all Bs are Cs". Then, for an explainable system, the agent would need to be able to expect that also "As are Cs" iff the fact is deducible in its logic L_2. The mapping $\tau_{1\to2}$ then allows the query answers provided by agent A_1 to be made fully comprehensible to agent A_2, for example by explicating differences between L_1 and L_2. In order to make progress towards truly explainable AI we obviously need to make progress in grasping human thinking in logic terms, i.e., characterize L_2 (assuming A_2 to be the human) and in the understanding of logics underlying KGEs, i.e., understanding L_1. With this paper we aim to bring the second aspect into attention by discussing KGEs from the perspective of logical commitments, and show how advancements can be made.

The above definition presumes a notion of logic that comes with a syntax and semantics with the usual notions of structures/interpretations and of models as wells as derived notions such as entailment which can be used to formally define query answering. Zooming into agent A_1, it may already provide a class X of structures (e.g., obtained by a knowledge graph embedding) which gives rise to some logic L_1' that needs to provide the basis for answering queries posed in L_1. To this end, it is essential to show that the class X is sound and complete for L_1. We note that L_1 has be considered as an umbrella logic that encompasses a query language and a language for the knowledge base.

2.1. Logical Commitments in Machine Learning

Logical commitments exist in most machine learning approaches, albeit often implicit and maybe even unconsciously. For example, in an incremental (multi-classification or multi-labeling) learning scenario, an object b could have been classified as an instance of various concepts $C_1 \ldots C_n$. But then, one natural reading of the "and" operator suggests that object b is also an instance of $C_1 \sqcap \cdots \sqcap C_n$, where $\sqcap$ stands for and concept-concept constructor. Sure, this natural reading may be debatable as sometimes one would like to express additional knowledge in the syntax: $(C_1 \sqcap C_2)(b)$ could be read as a stronger assertion than $\{C_1(b), C_2(b)\}$: Though both are truth conditionally equivalent, the first indicates justification of $C_1(b)$ and $C_2(b)$ on the same grounds whereas the latter indicates possibly different justifications for $C_1(b)$ and $C_2(b)$. Nonetheless, the told story then indicates an implicit reference to a relation of specific meaning, namely that of justification, that could be laid down also in a logical manner. So we would even go further from a simple logic with an "and" operator to a modal-logic-style logic based on an accessibility relation.

As another example consider the case of negation. Some form of negation is implicitly an essential part of concept learning simply as one draws a line between what one expects to be in the extension of a concept and what one expects to be not (sic) in it. The question then is whether this label level negation can be extended to a full negation on arbitrary concepts, say negation of $(C_1 \sqcap C_2)$. As we will discuss, such extension following the usual axioms of (Boolean) negation is neither easy nor a commitment desirable in all circumstances.

2.2. A Roadmap to Understanding Logical Commitments

The best of all worlds may not always exist as an option. Some classes of (geometric) objects may just not be axiomatizable—mainly due to the fact that one cannot ensure completeness—or a suitable logic is not found yet because no characterization has been achieved so far—again due to the challenge of showing completeness. A case in point for the latter situation are the geometrical structures that play a fundamental role in quantum logics, namely closed subspaces of a Hilbert space [22]. But even if one does not have a complete characterization, we argue, one should identify logically properties of geometrical objects as far as known and exploit them for understanding the geometrical objects induced by an embedding approach.

In which sense might this be helpful? We believe that such properties might give an additional justification and hence explanation for the geometrical commitments. If a KGE commits itself to a set of geometrical objects, then it commits to structures expressible with these objects which may introduce some bias with respect to what can be learnt and which valid logical conclusions can be drawn. For bridging the gap between ML/KGE and humans we have to bridge the gap between the respective forms of reasoning employed. This requires us to make progress in building foundations on both sides of the gap. With studying logical commitments we strengthen the foundation on the AI side.

As a concrete instance of the kind of argument explicated above we will investigate embeddings based on closed convex cones. These have been discussed in our own works [11, 23] but recently got attention from researchers from the ML community [24]. As mentioned above, in contrast to previous works which started by a specific KR language and then identified suitable geometric models, we take a complementary approach by investigating logical structures exhibited by a class of geometric models which are relevant to KGE. We identify orthologic as a logic framework in which we can express relevant properties of cone-based models. It turns out that the structure of cones is not distributive. However, this is not a bug but a desirable feature, i.e., a logical commitment we may want to make: non-distributive concept hierarchies can model uncertainty or partial information w.r.t. the extension of concepts [25]—a typical challenge faced in many ML tasks.

3. Logics for Knowledge Graph Embeddings

In this section we want to outline some methodological considerations on how a systemic survey of KGE approaches w.r.t. logical commitments could look like and provide a proof-of-concept survey of some recent KGE approaches with explicitly stated or easily identifiable logical commitments. In order to do so, we start with the notion of *full expressivity*, because, identifying logical commitments leads to assertions on the expressivity of KGEs. We will then argue that the investigation cannot stop at this point and motivate this by describing the general embedding scenario we have in mind. Afterwards we discuss current categorizations existing in contemporary KGEs.

Only recently, investigations have started that try to reveal and explicate the logical commitments made by knowledge graph embeddings. Earlier approaches to knowledge graph embedding—including the well-known TransE [26]—were motivated by efficient learning algorithms with a specific task in mind (link prediction), hence resolving the ubiquitous expressivity vs. feasibility dilemma strictly in favor of feasibility. Hence, expressivity aspects, let alone expressivity w.r.t. logics, were outside the focus of those approaches.

The intuition that early embedding approaches (such as TransE) are weakly expressive can be formally captured by the notion of *full expressivity* [27]. The basic idea is simple: A KGE approach is considered to be fully expressive iff for any pair of disjoint sets of positive triples P and negative triples N, the embedding adheres to this dichotomy. If one assumes as in [27] that the embedding approach rests on a scoring function σ_R for binary relations R, this conditions becomes more concretely: for any disjoint pair (P, N) of triples there is a sufficient large dimension of the embedding space and thresholds λ_R such that

- for all $(s\ R\ o) \in P$: $\sigma_R(e(s), e(o)) \leq \lambda_R$ and
- for all $(s\ R\ o) \in N$: $\sigma_R(e(s), e(o)) > \lambda_R$.

Observing that a given KGE approach is not fully expressive is, in our eyes, the starting point (not the end point) of investigations on the expressivity of this particular approach. So far, there exists no practical and useful KGE approach guaranteeing to satisfy full expressivity for arbitrary pairs (P, N) and it might be argued whether this would even be sensible to expect. Not until we gain a precise understanding of the expressivity achieved by a specific KGE, we can employ it in a methodological sound manner. This calls for a rigorous analysis of KGE in order to understand the different expressivity classes (for the non-fully expressive KGE) approaches. How could one formally define those notions?

As the notion of full expressivity bears some combinatorial character, fundamental information-theoretic approaches such as that of the Vapnik-Chervonenkis (VC) dimension [28] seem obvious starting points. VC measures the (combinatorial) complexity of a set of binary classification models as the maximal set of elements on which each binary partition (dichotomy) is correctly classified. Without giving an explicit definition here, similarly, one could define the expressivity of embedding approaches by considering the maximal cardinality $|X|$ of sets of triples X such that each partition $KG = N \uplus P$ is correctly classified by the embedding approach, i.e., the maximal-cardinal set X for which the embedding approach would be fully expressive. While this may be a be an interesting line of work, it does not foster understanding explainability of KGEs. First, notions like VC-dimensionality would lead to a linear hierarchy of expressivity levels not allowing for orthogonal aspects in expressivity. Second, for a transparent, explainable system one would expect a more qualitative approach along logical commitments which could be easily followed by humans. And in fact, identifying logical commitments has already been implicitly addressed by KGE research such as simpleE [27] which discuss kinds of background knowledge expressible in their approach, yet without describing expressiveness on a logical level. In several cases precise logic characterizations of expressivity are not hard to achieve. For example, one can derive that the KGE approach DistMult [29] can model only symmetric relations by observing their use of a bilinear scoring function $\mathbf{y}_{e_1}^T \mathbf{M}_r \mathbf{y}_{e_2}^T$ [29, Equation (2)] where $\mathbf{y}_{e_i} \in \mathbb{R}^n$, $i \in \{1, 2\}$ and $\mathbf{M}_r \in \mathbb{R}^{n \times n}$ is a diagonal matrix. Thus, DistMult commits to symmetric relations.

Moreover, considering the simpleE KGE [27] reveals that transition-based embeddings fulfill the property that if a relation is reflexive then it must also be symmetric and transitive. This is the starting point for kinds of expressivity considerations that must be pursued under the search for identifying logical commitments. We say "starting point" because considering relations only—though this might seem obvious in the consideration of knowledge graph triples—would be too short-sighted. This will become clear in the following formalization of the general embedding scenarios for which we at all see a chance for identifying logical commitments so that definition of explainability can be applied.

At the end of the day, an embedding leads to some structure $\mathfrak{E}$ which contains information about individuals, binary relations, and concepts (unary relations). This is clear w.r.t. the former two information bearers (individuals and relations) due to the fact that KGs are sets of triples $(s\ R\ o)$ with individuals s, o as first and last arguments and a relation R as second argument. Here and in the following we will use the logical notation $R(s, o)$ and talk of role assertions. Yet, we also have to account for concepts in KGEs for two reasons: First, at least for RDF graphs, there is a special binary relation rdf:type which can be used to state that some individual is an instance of a concept (class in RDF speak), as in $(s \text{ rdf:type } C)$ stating that s is an instance of a concept C. We use the logical notation $C(a)$ instead and talk of concept assertions. Note that these logical notions are in-line with the Tarskian style semantics of concepts which are interpreted by sets. In RDF-speak, the concepts are "reified" objects in the domain (and not sets of elements of the domain). Nonetheless, to empower reasoning, concepts have to be associated with sets representing their extensions. This leads us to the second reason for consideration of concepts in the realm of KGEs: With every relation two natural concepts are introduced, namely the domain and the ranges of the role. These are basic concepts that one would like to talk about or constrain in background knowledge (see below), for example when stating that a role-filler is of a certain type. Note that domain and range of relations now are concepts described as entities with an extension. Summarizing the above points, in order to characterize relevant entities in an embedding scenario we need a signature as it is usually used in description logics, namely a set which is the union of individual constants N_c, binary relation symbols, also called roles, N_R, and last but not least concept symbols N_C. Then a KG can be described as a set of sentences of the form $R(a, b)$ and $C(a)$. (In DL speak such a set is an abox). Not until we have identified such signature for a given KGE we can assess and construct explainability in the sense of Definition 1.

The structure induced by an embedding can be used for different purposes—usually termed "downstream applications" in the embedding community. We rather would like to think of special inference services that can be accomplished on those structures. Put differently, we propose to investigate which types of reasoning a structure fosters and how it can be characterized as a logical calculus. One kind of inference frequently discussed in the context of KGEs is analogical reasoning as popular in word puzzles. For example, consider the following puzzle: If man is to woman as X is to queen, what is X? The kind of logic required here is hidden behind the 'is to' and 'as' which are relations or mappings which have to the best of our knowledge never been characterized in terms of logical operators. By contrast, solving these word puzzles is a form of inductive reasoning that has already been studied comprehensively in the logics community [30] and is useful for learning and knowledge discovery.

Another, even more basic yet useful inference is termed (link) prediction or knowledge graph completion: the embedding structure may lead to new assertions/triples being true in the original KGE. This can be considered as a form of simple query answering: given the resulting structure $\mathfrak{E}$ of an embedding and a triple t, is it the case that t is true in $\mathfrak{E}$? One crucial point is how to explicate the semantics of a query being made true in the structure $\mathfrak{E}$. This question rests on what kind of information bearers are represented in $\mathfrak{E}$ and how they are represented. A survey such as that of [7] describes the types of $\mathfrak{E}$ by describing how individuals and relations are represented and how the scoring functions looks like.

The survey of [7] already categorizes approaches that rely on scoring functions, but it does not generalize to approaches that are not based on a scoring function such as [8, 11]. Let us give here an argument why we think that a more logic-based approach would do more justice to capturing the meanings of role and concept symbols. We give the argument for a typical widely discussed score-threshold-based approach such as that of rotatE [31]. Consider an object represented by a vector x in the embedding space. A relation R is represented in rotatE by a rotation M_R. Vector $y := M_R x$ gives only some "prototypical" object to which x stands in R-relation. Other objects y' to which x might stand in R-relation are given by $\|M_R x - y'\| \leq \lambda$ for some threshold λ. In particular, this means that the objects to which x is R-related are always close to $M_R x$ ("threshold balls"). In consequence, x can not be related to some objects y' and y'' that are quite different in that they belong to complementary concepts, say $y' \in C$ and $y'' \in \neg C$.

We argue that one must change the perspective of investigations of scoring-threshold based KGE approaches towards a logic perspective that provides a unifying framework for all kinds of KGE approaches. Explainability hinges on the kinds of relations inducible, thus on the embedding structures $\mathfrak{E}$ that provides the foundations for the semantics of any logic induced by a specific KGE approach.

3.1. Logics Grounded in Geometric Structues

In order to characterize an embedding structure $\mathfrak{E}$ from a logic point of view, we must associate it with some logical structure $\mathcal{I}$. How such a logical structure $\mathcal{I}$ might look in general depends on the choice of the semantics of concept validity (first level in Figure 1). Common to all of them are the following properties:

- The domain is a continuous space (the embedding space), typically R^n.
- Constants are interpreted by elements in this space.
- Concept symbols are associated with some sets of elements of the domain (possibly exhibiting some additional structure: probability, many-valued semantics, etc.).
- The sets concept symbols are associated with belong to a specific class of geometrical objects.
- (Binary) relations are associated with some set of pairs (possibly exhibiting some additional structure: probability, many-valued semantics, etc.).

An instance of such structure exhibited is called a *geometric model* according to [8]. Motivation for putting an emphasis on geometry, i.e., the spatial properties of the model, like in the works by [8], can be found in the theory of *conceptual spaces* [13] which characterizes reasoning in general to be of a spatial nature. Let for an embedding approach EA the set of geometric models induced by all embedding structures $\mathfrak{E}$ be denoted as $\text{gMod}(EA)$. Now, beyond the simple inference of link prediction or the more advanced analogical reasoning one might ask more complex queries that do not ask just whether a triple holds in the structure. So there may queries consisting of conjunctions of triples (CQs) or unions of conjuncts or even arbitrary first-order logical queries. Of course, this can be considered as an add-on service—completely independent of the embedding. But if an embedding structure is supposed to be used as basis to answer such queries then already in the beginning it has to provide operators for the various logical operators in the query language such that for an query (complex) Boolean query q over some query language over a DL signature there is a well-defined notion of a structure making the query true: $\mathcal{I} \models q$.

Several embedding approaches already take into account the existence of background knowledge [8, 27]. We will assume here that the background knowledge is given in form of terminological axioms, termed $\mathcal{T}$. The idea is that the ontology contains axioms that do not have be to learnt at all during construction of the embedding but must be accounted for as (hard) constraints during learning. In DL speak we can consider the terminological axioms $\mathcal{T}$ as a tbox and the knowledge graph as an abox $\mathcal{A}$. The embedding must result in a geometric model such that $\mathcal{I} \in \text{gMod}(EA)$ is true w.r.t. the pair $(\mathcal{T}, \mathcal{A})$. So we need here again a notion of modeling $\models$. But this modeling relation may not necessarily be the same as the one used in the query language: The set of logical operators used in the tbox language may be different from the logical operators used in the query language.

To make the last point clear, assume that we consider fragments of first-order logic as potential logics for the tbox (training phase) and the query answering service (application phase). Testing the consistency of a tbox in full first-order logic is not decidable (only semi-decidable). Hence, one could not use full first-order logic here. On the other hand, full first-order logic (or at least a safe fragment of it) could be used for the simpler problem of query answering on the structure (if we assume that the structure is a classical logical structure and then query answering amounts to model checking.) Of course the logic of the query language and the background language must be such that they can interact. Hence, we assume that for both languages there is the pre-defined signature and a logic L_1 as used in Definition 1 which contains sub-logics for queries and background knowledge, respectively.

Now consider the structure $\mathfrak{E}$ resulting from application of an embedding approach. In order to make a system explainable according to Definition 1, we would have to find (next to some transformations) the logics L, L_1, L_2 and a knowledge base KB in L_1. This knowledge base can be defined as the theory of (set of sentences true in) the structure resulting from the embedding. Concretely, assume $\mathfrak{E}$ to be the embedding structure and $\mathcal{I}_{\mathfrak{E}}$ its associated logical structure. Then the set of all L_1-sentences true in $\mathcal{I}$ represents the knowledge base KB we were aiming at:

$$KB = \{\alpha \in L_1 \mid \mathcal{I}_{\mathfrak{E}} \models \alpha\}$$

Of course everything hinges upon finding appropriate logics L, L_1, L_2 and such that L_1 is sound and complete for the class of structures of an embedding. Because this may not always be possible to achieve completely, we proposed a level-based definition of logical commitments.

3.2. Logic Signatures of KGEs

In the following, we focus on those KGE approaches that mention the kind of geometries used for embedding and the logic that characterizes them—or at least allow to associate some induced geometry or logic. For the two linear

Table 1
Comparison of approaches for embedding w.r.t. the geometric objects logical commitments.

Geometries	Logic L	Concept extension	Operators	Quantifiers	Reference
convex sets	Quasi-Chained Datalog$^{\pm}$	binary total	sub-Boolean algebra: no disjunct implicit atomic negation	quasi-chain guarded	[8]
hyperspheres	$\mathcal{EL}$	binary total	sub-Boolean algebra: no disjunct atomic negation	guarded existentials	[9]
axis-aligned cones	$\mathcal{ALC}$	binary partial	Boolean algebra: full negation full disjunction	guarded	[11]
cones	Minimal Orthologic + POM +?	binary partial	non-Boolean algebra: orthonegation disjunction partial orthomodular	no roles	[32]
closed subspaces in Hilbert space	Minimal Quantum Logic + ?	binary total	non-Boolean algebra: orthomodular orthonegation	no roles	[10]
hyperbolic cones	taxonomic	binary total	implicit	implicit universal	[33]
Cartesian products of 2D-Cones	∃FOL queries	binary total	non-Boolean algebra: negation as failure	no universal	[24]
threshold balls	Some restricted relational language	binary total	unknown induced algebra	N/A	[26, 29, 31]

approaches TransE [26] and RotatE [31] as well as the bilinear approach DistMult [29] no explicit discussion of (Boolean) operators is given. It is not obvious whether one can associate with the given geometries simple geometric operations that could be interpreted as (Boolean) operators, let alone quantifiers. For example, in case of TransE all relations are total which may induce some properties in a possible algebra of TransE relations that, to the best of our knowledge, has not been investigated yet.

The only operator that is implicit is a form of (atomic) negation-as failure for the intended prediction scenarios of those embeddings: If some (pair of) objects is not classified as being in that concept (relation) then it is not of that type, it is in its anti-extension. A characterization of logical commitment for those approaches, as far as possible, is summarized in Table 1. We consider the overarching logic L required for the respective embedding approach in the table. The accompanying explanations in the main text gives further pointers regarding possible distinctions of query logic and background logic.

Gutiérrez-Basulto and Schockaert [8] identify a fragment of existential Datalog (fulfilling the quasi-chainedness property) as an appropriate logic for arbitrary convex regions in Euclidean spaces as concerns the background logic. Kulmanov et al. [9] discover a correspondence for hyperspheres and the lightweight description logic $\mathcal{EL}$. Here, too, the background logic is considered. Özçep et al. [11] identify axis-aligned cones as a geometrical class that presents a model for the semi-descriptive logic $\mathcal{ALC}$. The query language considered is mainly that used for multi-label learning but conjunctive queries are possible. [8, 9] do not support full negation but in essence disjointness—hence we talk of atomic negation). In both approaches there is not explicit atomic negation operator on concepts but could be defined by the integrity constraints (using $\perp$). In contrast to [8, 9], the approach in [11] defines on the basis of a polarity operator a full negation operator on concepts represented by axis-aligned cones. On the other hand, in [11] binary relations are allowed to be arbitrary pairs of vectors, whereas [8] models also relations (of any arity) by convex regions.

In all three approaches the expressible concept hierarchy fulfills distributivity of conjunction over disjunction. For [8] and [9] this is trivially true as there is no explicit disjunction defined (only an implicit one in the head of rules, i.e., general inclusions). The approach of [10] considers minimal quantum logic which does not fulfil distributivity but (only) a weakening: orthomodularity. But, as argued for in [25], the ability to express non-distributive concept hierarchies means a benefit, as it enables modeling uncertainty with respect to the extensions of concepts: in many ML learning tasks, concepts do not have a crisp boundary, but rather have areas of uncertainty and as such provide partial information only.

Regarding partiality and non-distributivity, the approach of [32] is similar to that of [10]. The class of cones leads to a logic that is not even orthomodular but fulfills a weakening called partial orthomodularity. Whether this logic is axiomatizable at all is not known so far.

The approach of [33] uses hyperbolic cones in order to model relation hierarchy graphs and to grasp properties that follow by traversing the edges. The approach is special in that terminological knowledge which we considered as part of the background knowledge is directly encoded in the knowledge graph by relying on reification. The exact logic captured by this approach is not obvious, as the authors allow next to subclass relations also part-of relations. Moreover, it is not clear how quantifiers are handled. There is implicit universal quantification due to the subsumption hierarchy. But quantifiers, negation (and other Boolean operators) are not handled explicitly in this approach. Due to the above observations we tentatively described the logic as taxonomic.

The approach take by Zhang et al. [24] also uses the idea laid down in [11] to handle negation of concepts by using cones. In contrast to [11], Zhang et al. do not consider negation as polarity, but as set-complement in 2D. Other concepts' extensions then are constructed as Cartesian products. The authors claim to embed FOL queries. We think that this concerns only the query language (and not the background language). Interestingly, they exclude the universal quantifiers $\forall$ from their considerations. Given the fact that $\forall$ is dual to $\exists$ via negation we consider this as a sign that negation is not treated as part of the background language and must be treated rather as a form of negation as failure. Regarding the other operators we note that [24] cannot guarantee de Morgan rules to hold. Due to this fact and despite the paper suggesting that classical FOL queries are embedded, we would consider their approach as being based on some non-Boolean algebra.

In the next section we give a more detailed discussion of the logical commitments of our own proposal for KGE based on cones [11, 32]. This will give us an opportunity to motivate orthologics as framework to compare different approaches.

4. Orthologics for Cone-Based Embeddings

As our discussion of logical commitments in KGEs revealed, not all approaches yield structures that induce a well-known logic. This observation is not to be interpreted as a shortcoming since these may exhibit features particularly suited to certain applications or to strengthen explainability. For example, designing a logic that captures some properties of human reasoning [12, for example] may ease bridging between KGE and humans. In order to facilitate comparisons between distinct logic commitments identified, we strive for a general logic framework in which specifics of several KGEs can be expressed. Clearly, one has to balance generality with respect to effectiveness of performing the analysis. We propose to choose orthologics [34] as one such general framework. Orthologics are a calls of classic yet little restricted Boolean logics that can encompass several properties underlying KGE approaches. In particular, orthologics avoid strong logic commitments on the fundamental negation operation (unlike Boolean logics do) that are not met by geometric models in KGEs.

4.1. Cones for Representing Concepts

Geometric models based on cones have recently received some attention [11, 24] and may offer an interesting balance between feasibility of computing an embedding and achieving high expressivity. A *convex cone A* is a set such that from $x, y \in A$ it follows that $\lambda x + \mu y \in A$ for any $\lambda, \mu \in \mathbb{R}_{\geq 0}$. In a crisp Boolean semantics we say that some constant x is part of some concept represented by cone a iff $x \in a$. One of the nice properties of closed convex cones we exploit is that they allow a polarity operation that resembles negation in orthologic, called orthocomplent. The *polar cone a°* for a is defined for Euclidean spaces with a dot product $\langle \cdot, \cdot \rangle$ as $a^\circ = \{x \in \mathbb{R}^n \mid \forall y \in a : \langle x, y \rangle \leq 0\}$. Intuitively, the polar of a cone consists of all vectors that differ in orientation to any element of the cone by at

Axioms

$$A \vdash A \quad A \,\&\, B \vdash A \quad A \,\&\, B \vdash B \quad A \dashv\vdash {\sim}{\sim}A \quad A \,\&\, {\sim}A \vdash B \quad A \vee B \dashv\vdash {\sim}({\sim}A \,\&\, {\sim}B)$$

Rules

$$\frac{A \vdash B,\ B \vdash C}{A \vdash C} \qquad \frac{A \vdash B,\ A \vdash C}{A \vdash B \,\&\, C} \qquad \frac{A \vdash B}{{\sim}B \vdash {\sim}A}$$

Fig. 2. Minimal Orthologic *Omin*

least 90°. Clearly, the polar of a convex cone is also a convex cone. It appears natural to define conjunction as set intersection and disjunction as set union, but this creates problems since convex cones are obviously not closed under set union. One way out of this problem is to define union via De Morgan exploiting the availability of negation, i.e., $a \vee b := \neg(a \wedge b)$. However, this raises the question of which logic commitment is induced by such definition.

4.2. Orthologic as Framework for Logical Commitments

Orthologics are a family of logics based on the algebra of ortholattices, which, roughly speaking, can be understood as Boolean algebras without the distributivity rule. Orthonegation is a weakening of Boolean negation, which does not fulfill the principle of the excluded middle. As argued by Conradie et al. [25], this leads to the possibility of modeling incomplete (unknowable) information, which can be captured by non-distributive and even non-orthomodular logics. In this way, semantics acknowledging uncertainty in form of partial Boolean validity can be captured (cp. Table 1). This turns out to be relevant in many application areas, e.g., in a terminology of species, where one species has more than two direct subspecies. It is easy to prove that the resulting lattice is non-distributive. So, we could then use closed subspaces of Hilbert spaces as geometrical structures or equivalently a minimal orthologic to deal with non-distributive (but orthomodular) concept hierarchies, as done by Garg et al. [10] (see also Table 1). But actually, it may also be useful to weaken distributivity even beyond orthomodularity. Consider, e.g., a medical scenario where a positive outcome of a clinical test is correct with high probability, whereas a negative one is correct by chance only. This means that a person under suspicion can't be classified as healthy even if tested. As many other embedding approaches do not allow for non-distributivity (even if they allow for negation) and have problems with handling incomplete knowledge, this leads to an increased representability.

We are interested in extensions of minimal orthologic Omin [34], portrayed in Figure 2, with rules that describe properties of arbitrary ortholattices of cones. For this we consider structures $(X, \llcorner)$ called orthoframes, i.e., pairs with a set X and an orthogonality (i.e irreflexive and symmetric) relation $\llcorner$. The $\llcorner$-closure of a set $A \subseteq X$ is defined as $A^{\llcorner} = \{X \in X \mid X \llcorner y \forall y \in A\}$. A set is called $\llcorner$-closed iff $A^{\llcorner\llcorner} = A$. A special subclass of orthoframes are *cone-based orthoframes*, where where $X = \mathbb{R}^n$ and $\llcorner$ is the binary relation defined by $\langle x, y\rangle \leq 0$ for $x, y \in X$ and not both of x, y are equal to $\vec{0}$. We call the resulting logic $\mathcal{L}^{min}_{pOM} = Omin + (pOM)$ the *minimal logic of partial modularity*.

For the discussion of the rule we consider orthomodels $\mathcal{I}$ where $(A)^{\mathcal{I}} = a, (B)^{\mathcal{I}} = b$ and $(C)^{\mathcal{I}} = c$. The intuition behind this rule is the following (see also Figure 3, bottom): The conclusion of the rule is exactly the conclusion of the orthomodularity. Without all the premisses this rule does not hold for cones: Considering only the premise (PM1) gives us exactly orthomodularity (Omr) which can be shown not to hold for cones.

But the conclusion of (pOM) can be guaranteed to hold under specific circumstances where b and a° are contained in a halfspace c and where b is at the correct border of a namely when b and $a \sqcap c$ are *perspective* [35] w.r.t. c°.

(PM1)	B	$\vdash$	A
(PM2)	${\sim}A \vee B$	$\vdash$	C
(PM3)	$B \vee {\sim}C$	$\dashv\vdash$	$(A \,\&\, C) \vee {\sim}C$
(C)	$A \,\&\, ({\sim}A \vee B)$	$\vdash$	B

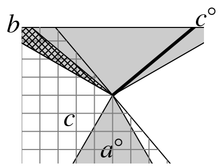

Fig. 3. Rule (pOM) and its illustration in $\mathbb{R}^2$

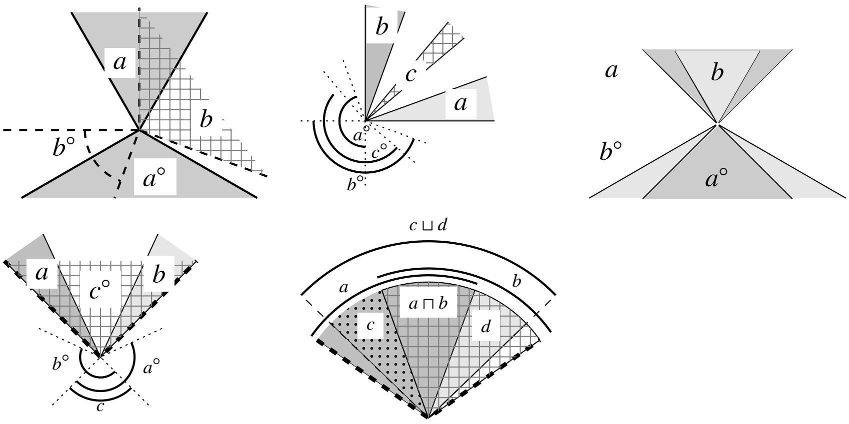

Fig. 4. Counter-examples for rules falsified by cones. Top left: excluded middle. Top middle: distributivity. Top right: orthomodularity. Bottom left: Join-Semi-Distributivity. Bottom right: Whitman Condition

Theorem 1 ([32]). *Any subortholattice of closed convex cones in a (finite-dimensional) Hilbert-space fulfils (pOM).*

Rule (pOM) marks a starting point for understanding the logical commitments of embedding approaches based on cones. Whether closed convex cones are axiomatizable at all with a sound and complete orthologic is still an open question.

Justification for investigating cones and the logics they introduce can be found in the following observation, showing that cones are structures that naturally arise when an orthogonality relation is constrained to rely on an arbitrary symmetric positive semi-definite bilinear form (e.g., scalar product) since $\llcorner$-closure enforces cones.

Proposition 1. *For a given vector space V (over $\mathbb{R}$ or any other field with an order $\leq$) and a symmetric positive semidefinite bilinearform $\langle\cdot,\cdot\rangle$ let $X = V \setminus \{\vec{0}\}$ and define for all $u, v \in X$: $u \llcorner v$ iff $\langle u, v\rangle \leq 0$. Then: $(X, \llcorner)$ is an orthoframe and the $\llcorner$-closed sets are closed convex cones (without $\vec{0}$).*

Proof. Due to symmetry/positive semidefiniteness of $\langle\cdot,\cdot\rangle$ the relation $\llcorner$ is symmetric/irreflexive. Now consider a $\llcorner$-closed set Y. We have to show that for all $\lambda \geq 0$ and $y \in Y$ also $x := \lambda y \in Y$ (or $\lambda y = \vec{0}$) and with $u, v \in Y$ also $u + v \in Y$ (or $u + v = \vec{0}$). So let $\lambda \geq 0$ and $y \in Y$. Further assume $\lambda y \neq \vec{0}$, thus $\lambda \neq 0$. Due to $\llcorner$-closure there is a y' with $y' \llcorner Y$ and $y' \not\llcorner x$. But this means in particular $\langle y', y\rangle \leq 0$ and $\langle y', x\rangle = \langle y', \lambda y\rangle = \lambda\langle y', y\rangle > 0$. As $\lambda \neq 0$ this is a contradiction. Now let $u, v \in Y$ and $\lambda \geq 0$ and assume that $u + v \neq \vec{0}$. So u is not $-v$. Towards contradiction assume $x := u + v \notin Y$. Due to $\llcorner$-closure there is a y' with $y' \llcorner Y$ and $y' \not\llcorner x$. The first in particular means we have $\langle y', u\rangle \leq 0$ and $\langle y', v\rangle \leq 0$; the latter means that $\langle y', u + v\rangle = \langle y', u\rangle + \langle y', v\rangle > 0$, a contradiction. □

4.3. Aligning Polarity-Based Negation with Orthologics

Reviewing the rich body of orthologics we discovered several candidate rules that grasp certain aspects of negation, but neither of them is satisfied by a geometric model of cones using polarity as negation. To the best of our knowledge, no rule has yet been proposed that captures the structure of cones. In other words, there is no precise understanding of the logical commitments of cones-based KGEs. While the in-depth discussion of rules not satisfied is not relevant for this paper, Table 2 presents a selection of falsified prominent rules. The table is complemented by Figure 4 which we hope to help conveying the intuition of KGE using cones.

Table 2 summarizes some prominent rules for which falsifying cone-based orthomodels exist. The rules above the double line are rules in a propositional calculus that are either discussed directly in the logic literature or that are obvious translations of rules discussed in the context of lattice theory. The rules below the double line are defined without a reference to a calculus but directly with lattice-theoretic notions because these are not expressible in a propositional calculus without propositional quantifiers—as in the case of the calculus of Goldblatt [34] which we consider in this paper.

Table 2

Well-known rules excluded from a logic of cones.

Name	**Propositional Calculus Rule**	**Comment**
Distributivity (D)	$A \mathbin{\&} (B \vee C) \dashv\vdash (A \mathbin{\&} B) \vee (A \mathbin{\&} C)$	Adding (D) to ortholattices gives Boolean logic
Meet-Semi-Distributivity (MSD)	$\dfrac{A \mathbin{\&} C \dashv\vdash B \mathbin{\&} C}{A \mathbin{\&} C \dashv\vdash (A \vee B) \mathbin{\&} C}$	Weakening of (D)
Join-Semi-Distributivity (JSD)	$\dfrac{A \vee C \dashv\vdash B \vee C}{A \vee C \dashv\vdash (A \mathbin{\&} B) \vee C}$	Weakening of (D)
Modularity (M)	$\dfrac{C \vdash A}{A \mathbin{\&} (B \vee C) \vdash (A \mathbin{\&} B) \vee C}$	Weakening of (D)
Orthomodularity (Omr)	$\dfrac{A \vdash B, \sim A \vdash C}{A \vee (B \mathbin{\&} C) \dashv\vdash (A \vee B) \mathbin{\&} (A \vee C)}$	Weakening of (M); gives quantum logic
Johannson's minimal negation (LLJ)	$\dfrac{A \mathbin{\&} B \vdash C}{A \mathbin{\&} \sim C \vdash \sim B}$	Only rule in Dunn's kite [21] falsified by cones.
Weak J.'s negation (wLLJ)	$\dfrac{A \mathbin{\&} B \vdash D \mathbin{\&} \sim D}{A \vdash \sim B}$	Weakening of (LLJ)

Name	**Lattice rule**	**Comment**
M-symmetry [36, p. 66]	If $M(a,b)$ then $M(b,a)$	Weakening of (M)
Mac Lane's Cond. (Mac_1) [36, p. 111]	If $b \wedge c < a < c < b \vee a$ then there is a d with $b \wedge c < d \leq b$ and $a = (a \vee d) \wedge c$	Weakening of (M)
Semimodularity (SM) [36, p. 2]	If $a \wedge b <: a$ then $b <: a \vee b$ (<: denotes the covering relation)	Weakening of (MS) and (Mac_1) equivalent in the finite
Birkhoffs's Covering (Bi) [36, p. 3]	If $a \wedge b <: a, b$ then $a, b <: a \vee b$	Weakening of (SM) and (Mac_1)
Whitman Cond. (W) [37, p.479]	If $a \wedge b \leq c \vee d$ then $a \leq c \vee d$ or $b \leq c \vee d$ or $a \wedge b \leq c$ or $a \wedge b \leq d$	Used in discussion of free lattices

Counterexamples for (D), (MSD), (wLLJ), (LLJ), (JSD) and (W) are shown in Figure 4. The instance on the top right of Figure 4 is a counterexample to orthomodularity. Thus it is also a counterexample for (M) and (D). Moreover, the same instance serves as a counterexample for the condition of Birkhoff (Bi) and hence also of (MS), (Mac1), (SM): Let a and b of the rule (Bi) be instantiated by b and $\neg a$ of the instance on the top right of Figure 4. Then the precondition is fulfilled as $\perp = b \wedge \neg a <: b$ and $\perp = b \wedge \neg a <: \neg a$. On the other hand we do not have $b <: b \vee a^{\perp} = \top$ nor $\neg a <: b \vee \neg a = \top$.

As we can find cone-based orthomodels not fulfilling the distributivity law (D) of $\wedge$ over $\vee$, the logic generating cones is indeed a non-classical propositional logic. Moreover, some intuitive rewritings known from classical logics are not possible. For example, Johansson's constructive contraposition (LLJ) and its weakening (wLLJ) do not hold. Closed convex cones falsifying (wLLJ) is a demonstration that there are two different notions of complement. The first says that b is a complement of a iff $a \wedge b \leq \perp$, the second iff $a \leq \neg b$. In fact, it is known that any ortholattice which has a unique complement must in fact be distributive and hence must be a Boolean algebra (though this is not the case for arbitrary lattices [38]).

As orthomodularity is not fulfilled, the kind of logic induced by closed cones cannot be that of quantum logics [39]. (In fact, the main underlying geometric structure for quantum logics is not that of a cone, at least in the pioneering work of von Neumann and Birkhoff [40], but that of a closed subspace of a Hilbert space). Hence as a weakening of orthomodularity (pOM) is introduced (see Figure 3) in [32].

The Whitman condition is discussed in the context of free product lattices. A counterexample in $\mathbb{R}^2$ is portrayed in the bottom right of Figure 4.

As can be seen, the task of precisely identifying logical commitments is not an easy one and even simple structures like cones induce logical commitments that differ from logics studied so far. However, the negative results summarized in Table 2 are indeed not negative with respect to improving explainability but important stepstones towards a (fully) explainable agent that has learned knowledge by means of a cone embedding. Knowing that a specific rule is not met allows explanations to be augmented accordingly and enables the receiver to construct a consistent model of what the agent has learned.

5. Conclusion

Studying the logic structure of geometric models teaches about expressivity of knowledge graph embeddings and can contribute to interpretability of learned models. We argue that a precise understanding of logic structures is required to assess explainability of a representation. We introduce the term *logical commitment* to capture the implicit and explicit consequences of choosing a certain geometric model in a KGE approach. We show that existing approaches have taken a widespread selection of logical commitments, often implicitly as a consequence of identifying geometric models that can be learned from data in an efficient and robust manner. The key point this paper makes is that we must complement those works by a rigorous treatment of logical commitments and we have laid out a first roadmap of aspects that need to be investigated, including validity semantics, signature and semantics of operators, and of quantifiers. In the line with recent works on logic-based KGEs we focus on cone-based geometric models and discuss how orthologics can be used to explicate logical commitments. We have to acknowledge, though, that our call to action regarding the study of logical commitments is no easy research agenda and possibly touches fundamental problems of mathematics.

With a logic-based approach to understanding explainability as model alignment problem between two agents (cp. Definition 1) we aim to pave the way for engineering explainability. Above all, understanding logical commitments also paves the way for meta-cognition as it empowers an agent to reflect on what it has leaned using a calculus fitting to logical commitments made.

References

[1] F. Lecue, On the role of knowledge graphs in explainable AI, *Semantic Web Journal* (2019).
[2] R. Peñaloza, Axiom Pinpointing, *arXiv e-prints* (2020), arXiv:2003.08298–.
[3] D.L. McGuinness and A.T. Borgida, Explaining Subsumption in Description Logics, in: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'95, 1995, pp. 816–821. ISBN 1558603638.
[4] S.M. Lundberg and S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 2017, pp. 4768–4777. ISBN 9781510860964.
[5] M.T. Ribeiro, S. Singh and C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, ACM, New York, NY, USA, 2016, pp. 1135–1144. ISBN 978-1-4503-4232-2. doi:10.1145/2939672.2939778.
[6] F. Bianchi, G. Rossiello, L. Costabello, M. Palmonari and P. Minervini, Knowledge Graph Embeddings and Explainable AI, in: *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, I. Tiddi, F. Lécué and P. Hitzler, eds, Studies on the Semantic Web, Vol. 47, IOS Press, 2020, pp. 49–72. doi:10.3233/SSW200011.
[7] Q. Wang, Z. Mao, B. Wang and L. Guo, Knowledge Graph Embedding: A Survey of Approaches and Applications, *IEEE Transactions on Knowledge and Data Engineering* **29**(12) (2017), 2724–2743.
[8] V. Gutiérrez-Basulto and S. Schockaert, From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules, in: *Proc. of KR 2018*, M. Thielscher, F. Toni and F. Wolter, eds, AAAI Press, 2018, pp. 379–388.
[9] M. Kulmanov, W. Liu-Wei, Y. Yan and R. Hoehndorf, EL Embeddings: Geometric Construction of Models for the Description Logic EL++, in: *Proceedings of IJCAI-19*, 2019.
[10] D. Garg, S. Ikbal, S.K. Srivastava, H. Vishwakarma, H. Karanam and L.V. Subramaniam, Quantum Embedding of Knowledge for Reasoning, in: *Advances in Neural Information Processing Systems*, Vol. 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. Alché-Buc, E. Fox and R. Garnett, eds, Curran Associates, Inc., 2019.

[11] Ö.L. Özçep, M. Leemhuis and D. Wolter, Cone Semantics for Logics with Negation, in: *Proc. of IJCAI 2020*, C. Bessiere, ed., ijcai.org, 2020, pp. 1820–1826.

[12] M. Ragni, C. Eichhorn and G. Kern-Isberner, Simulating human inferences in the light of new information: A formal analysis, in: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-2016)*, AAAI Press, 2016, pp. 2604–2610.

[13] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*, The MIT Press, Cambridge, Massachusetts, 2000.

[14] R. Davis, H. Shrobe, and P. Szolovits, What is a Knowledge Representation?, *AI Magazine* **14**(1) (1993), 17–33.

[15] T.R. Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *International Journal of Human-Computer Studies* **43**(5–6) (1995), 907–928, special issue on the role of formal ontology in the information technology.

[16] K.D. Forbus, *Qualitative Representations: How People Reason and Learn about the Continuous World*, MIT Press, 2019.

[17] S. Soames, *Understanding Truth*, Oxford and New York: Oxford University Press USA, 1998.

[18] N. Francez, Bilateralism in Proof-Theoretic Semantics, *Journal of Philosophical Logic* **43** (2014), 239–259.

[19] B. Jonsson and A. Tarski, Boolean Algebras with Operators. Part I, *American Journal of Mathematics* **73**(4) (1951), 891–939.

[20] B. Jonnson and A. Tarski, Boolean Algebras with Operators, *American Journal of Mathematics* **74**(1) (1952), 127–162.

[21] C. Hartonas, *Reasoning with Incomplete Information in Generalized Galois Logics Without Distribution: The Case of Negation and Modal Operators*, in: *J. Michael Dunn on Information Based Logics*, Springer International Publishing, Cham, 2016, pp. 279–312. ISBN 978-3-319-29300-4.

[22] M.L. Dalla Chiara and R. Giuntini, Quantum Logic, *arXiv e-prints* (2001), quant-ph/0101028.

[23] M. Leemhuis, O.L. Özçep and D. Wolter, Multi-Label Learning with a Cone-Based Geometric Model, in: *Proceedings of the 25th International Conference on Conceptual Strutures (ICCS 2020)*, M. Alam, T. Braun and B. Yun, eds, 2020, pp. 177–185.

[24] Z. Zhang, J. Wang, C. Jiajun, J. Shuiwang and W. Feng, ConE: Cone Embeddings for Multi-Hop Reasoning over Knowledge Graphs, in: *Advances in Neural Information Processing Systems*, 2021.

[25] W. Conradie, A. Palmigiano, C. Robinson and N. Wijnberg, Non-distributive logics: from semantics to meaning, *arXiv e-prints* (2020), arXiv:2002.04257–.

[26] A. Bordes, N. Usunier, A. García-Durán, J. Weston and O. Yakhnenko, Translating Embeddings for Modeling Multi-relational Data, in: *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, C.J.C. Burges, L. Bottou, Z. Ghahramani and K.Q. Weinberger, eds, 2013, pp. 2787–2795. http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.

[27] S. Mehran Kazemi and D. Poole, SimplE Embedding for Link Prediction in Knowledge Graphs, *arXiv e-prints* (2018), arXiv:1802.04868–.

[28] V. Vapnik and A. Chervonenkis, On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities, *Theor of Probability and its Applications* **16**(2) (1971), 264–280.

[29] B. Yang, W. Yih, X. He, J. Gao and L. Deng, Embedding Entities and Relations for Learning and Inference in Knowledge Bases, in: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, eds, 2015. http://arxiv.org/abs/1412.6575.

[30] S. Stratulat, A Unified View of Induction Reasoning for First-Order Logic, in: *Turing-100: The Alan Turing Centenary Conference*, 2012.

[31] Z. Sun, Z.-H. Deng, J.-Y. Nie and J. Tang, RotatE Knowledge Graph Embedding by Relational Rotation in Complex Space, in: *Proceedings of the International Conference on Learning Representations (ICLR 2019)*, 2019.

[32] M. Leemhuis, Ö.L. Özçep and D. Wolter, Orthologics for Cones, *arXiv e-prints* (2020), arXiv:2008.03172–.

[33] Y. Bai, R. Ying, H. Ren and J. Leskovec, Modeling Heterogeneous Hierarchies with Relation-specific Hyperbolic Cones, in: *Proc. 35th Annual Conference on Neural Information Processing Systems (NeurIPS 2021)*, 2021, pp. arXiv:2110.14923–.

[34] R.I. Goldblatt, Semantic analysis of orthologic, *Journal of Philosophical Logic* **3**(1) (1974), 19–35. ISBN 1573-0433.

[35] F. Maeda and S. Maeda, *Theory of symmetric lattices*, Springer, 1970.

[36] M. Stern, *Semimodular Lattice*, Cambridge University Press, 1999.

[37] G. Grätzer, *Lattice Theory: Foundation*, Springer Basel, 2011.

[38] R.P. Dilworth, Lattices With Unique Complements, *Transactions of the American Mathematical Society* **57**(1) (1945), 123–154.

[39] K. Engesser, D. Gabbay and D. Lehmann, *A New Approach to Quantum Logics*, College Publications, 2007.

[40] G. Birkhoff and J. von Neumann, The Logic of Quantum Mechanics, *Annals of Mathematics* **37**(4) (1936), 823–843.