Machine Learning and Knowledge Graphs: Existing Gaps and Future Research Challenges

Claudia d'Amato
University of Bari, Italy

Louis Mahon
Oxford University, United Kingdom

Pierre Monnin
Université Côte d'Azur, Inria, CNRS, I3S, France

Giorgos Stamou
National Technical University of Athens, Greece

Abstract

The graph model is nowadays largely adopted to model a wide range of knowledge and data, spanning from social networks to knowledge graphs (KGs), representing a successful paradigm of how symbolic and transparent AI can scale on the World Wide Web. However, due to their unprecedented volume, they are generally tackled by Machine Learning (ML) and mostly numeric based methods such as graph embedding models (KGE) and deep neural networks (DNNs). The latter methods have been proved lately very efficient, leading the current AI spring. In this vision paper, we introduce some of the main existing methods for combining KGs and ML, divided into two categories: those using ML to improve KGs, and those using KGs to improve results on ML tasks. From this introduction, we highlight research gaps and perspectives that we deem promising and currently under-explored for the involved research communities, spanning from KG support for LLM prompting, integration of KG semantics in ML models to symbol-based methods, interpretability of ML models, and the need for improved benchmark datasets. In our opinion, such perspectives are stepping stones in an ultimate view of KGs as central assets for neuro-symbolic and explainable AI.

2012 ACM Subject Classification Information systems → World Wide Web; Computing methodologies → Artificial intelligence

Keywords and phrases Graph-based Learning, Knowledge Graph Embeddings, Large Language Models, Explainable AI, Knowledge Graph Completion & Curation

Digital Object Identifier 10.4230/TGDK.1.1.42

Category Vision

Funding *Claudia d'Amato*: Partially supported by the project FAIR - Future AI Research (PE00000013), spoke 6 - Symbiotic AI (`https://future-ai-research.it/`), under the NRRP MUR program funded by the NextGenerationEU and by the project HypeKG - Hybrid Prediction and Explanation with Knowledge Graphs (H53D23003700006), under PRIN 2022 program funded by MUR.
Pierre Monnin: Supported by the AT2TA project (`https://at2ta.loria.fr`), funded by the French National Research Agency ("Agence Nationale de la Recherche" – ANR) under grant ANR-22-CE23-0023.

Received Date of submission **Accepted** Date of acceptance **Published** Date of publishing

Editor TGDK section area editor

1 Introduction

Graph data refers to data that lends itself naturally to being represented as a graph-based data model. Examples of graph data are social networks, computer networks, entailment graphs [94], concept graphs [26]. Several standards have been proposed to represent graph data, including the W3C devised standards OWL, RDF, and RDFS. These enable easy sharing and combining of graph data from different sources, and so further facilitate the adoption of the graph formalism.

Amongst the several types of graph data in widespread use, one prominent example is the Knowledge Graph (KG). A KG aims to convey knowledge of the real world and represent it conforming to a graph-based data model, where nodes represent concepts of interest, such as human or lion, and edges represent possibly different relations between these entities, such as isTypeOf or isPredatorOf [68]. A closely related concept that we do not discuss any further is Property Graph, where both nodes and edges can have multiple properties which are represented as key-value pairs (the interested reader may refer to [68] for further details). Graphs data may be stored in native graph databases or relational databases [68].

When referring to the representation of information, the term 'knowledge', as opposed to 'data', is usually what is predicated of humans. It suggests the information is stored in a more structured and actionable manner, e.g. that it enables reasoning. This distinction from 'data' was first made in relation to the concept of a knowledge base (KB) [105], in the context of expert systems [64], in order to distinguish them from databases using, *e.g.*, lookup tables or hash tables. A KB is a representation of information as a set of facts or sentences [163].

A KG can be formalized as a triple of sets $\langle E, R, T\rangle$, where E is a set of entities, R a set of relations, and T is of the form $\{(s, p, o) \mid s, o \in E, p \in R\}$ [29], by which it is immediately equivalent to a KB, considered as a set of facts. Moreover, a graph $G = (\mathcal{N}, \mathcal{E})$, can be written equivalently as a set of facts, by equating $\mathcal{N}$ with the set of all entities appearing as arguments to facts, and equating each fact $\langle s, p, o\rangle$ to a directed edge from s (*subject*) to o (*object*) labelled p (*predicate*). On a higher level, one difference between a KG and KB as a set of facts, is that the former has a greater emphasis on the connection to the graph-based data model, and is more directly associated with the agreed formatting standards for graph data. Our discussion here does not require precise disambiguation of the term and in the remainder of this paper, we use the two terms interchangeably.

A closely related concept to a KG is an ontology. Intuitively, an ontology defines a set of object types, and how these types relate to each other. For example, if the domain is living things, then an ontology would specify that human and lion are two distinct types of a mammal, mammal and reptile are two distinct types of vertebrates etc. Formally, an ontology has been defined as comprising two components, the TBox, which introduces the vocabulary of an application domain, and the ABox, which contains assertions about named individuals in terms of this vocabulary [11]. Often the set of concepts in a KG forms an ontology, and their ontological relations can be incorporated into the structure of the KG. In the remainder of this paper, we will treat the term 'ontology' as interchangeable with 'knowledge base', as defined above.

Many important applications, such as e-Commerce [211], financial trading [29], semantic search [208], fact-checking [167], recommendation [198], (medical) decision support systems [205], question answering [73] and even machine translation [224, 136] benefit from access to real-world knowledge in a form that is both machine-readable and human-interpretable (i.e. entities, properties, relations and types). There has thus been a general convergence on KGs as the means to represent and store such knowledge. This interest from academia and especially from industry, has led to several large-scale efforts at constructing KGs. Some are freely available and accessible, such as DBpedia [9][1], Freebase [18][2], YAGO [174][3], Wikidata [190][4]. Others are private, developed for commercial use by companies such as Google, Amazon, IKEA, Uber, Microsoft, Facebook and LinkedIn. The interested reader could refer to [69] for a comprehensive overview of the history

[1] https://www.dbpedia.org/
[2] https://developers.google.com/freebase
[3] https://yago-knowledge.org/
[4] https://www.wikidata.org/

and current use of KGs.

The amount of data that may be of interest to KG applications is very large, e.g., English-language Wikipedia contains close to 7 million articles at the time of writing[5]. Developing KGs of this size is a difficult, expensive process, requiring the integration of multiple sources of information, along with input from human experts and crowdsourcing. Despite significant efforts for making KGs as comprehensive and reliable as possible, they tend to suffer from incompleteness and noise, due to the complex building process [69, 196]. This has prompted a search for automatic construction and enrichment [83, 193], often through the use of machine learning (ML). Indeed, the ML world has advanced considerably in the past decade, particularly with the rise of deep learning. From the victory of AlexNet in the ILSVRC in 2012 [97], to the release of ChatGPT in 2022, deep learning has come to dominate ML research and powers many industry applications.

One method of combining the world of knowledge and KGs with ML, and especially deep learning, is to form a vector representation of each node and edge in the KG, by optimizing some loss function based on the graph structure. The resulting set of vector representations is known as a knowledge graph embedding (KGE) and it enables several important use cases. In one direction, KGEs allow the use of predictive machine learning techniques to improve the KG, for example, by KG completion, where sparse KGs, such as those automatically constructed from text [90], are augmented with missing triples. Also, by using the deep neural network (DNN) feature vector extracted from a video, KGEs have been used to represent the content of a video as a graph [121]. Other uses of KGEs include triple (fact) classification, for assessing if a fact within the KG is correct or not, KG question answering and node clustering. Node clustering indeed can reveal similarities and differences between groups of nodes in the KG [59] and this can, for example, help uncover certain types of users in a social network, or article subjects, in a citation network. KG question answering uses the information in a KG to answer natural language questions [73]. In the other direction, KGEs allow KGs to be used to improve ML performance: for example, knowledge-aware visual question-answering [108], or reasoning of large language models (LLMs) [215].

In this paper, we introduce some of the main existing methods for combining KGs and ML, divided into two categories: those using ML to improve KGs, and those using KGs to improve results on ML tasks. From this introduction, we draw research gaps and perspectives that we consider urgent as well as promising. These gaps and perspectives are summarised in Table 1 (and analyzed and developed in section 3) and are concerned with the topics: LLM prompting, KG semantics and KGE models, symbol-based methods, ML model interpretability, and benchmark datasets. For each topic, we provide a description of some unsolved problems (gaps) that we consider to be of particular importance for future research work, and provide our views, claims, and proposals to overcome them. In particular, we support the use of KGs to formalise LLM prompting (*e.g.*, providing concept, defining sequencing). We claim that KGE could benefit from the injection of KG semantics and usage of various reasoning capabilities, *e.g.*, in terms of performance or negative generation. Informative negatives could also be generated by exploiting symbol-based method learning disjointness axioms (that are often missing). With respect to interpretability using KG, we argue that little progress has been made, and that in-model KG-based approaches that demonstrably produce reliable explanations are needed to validate ML results. Assessing these improvements in KGE performance or interpretability also calls for extensive empirical evaluations. Such evaluations require benchmark datasets that feature various schema constructs or levels of semantics that are currently lacking, unnoticed, or uncommon in the state of the art. That is why, we call for a systematic characterization and collection of available datasets as well

[5] https://en.wikipedia.org/wiki/Wikipedia:Statistics

as the creation of synthetic KG generators to produce tailored datasets to support experiments.

The remainder of this paper is organised as follows. Section 2 provides an overview of existing work linking KG and ML, under the framework of KGs for ML (Section 2.1) and ML for KGs (Section 2.2). Section 3 describes some gaps in the literature that we deem important, and outlines our vision of future research directions for filling these gaps. We particularly focus on: the use of KGs for prompting LLMs (Section 3.1), the integration of KG semantics and associated reasoning capabilities in KGE models for improved performance and negatives handling (Section 3.2), the potential of symbol-based ML for KGs (Section 3.3), the attempts to use KGs for explainable AI (Section 3.4), and the need for further benchmark datasets and metrics to assess improvements brought by aforementioned directions (Section 3.5). Finally, Section 4 concludes and summarises this work.

2 Machine Learning and Knowledge Graphs

In this section we focus on the interlink between ML and KGs. As sketched in [19], two main perspectives can be drawn: a) KGs as input to ML, whose main goal is to improve the performance in many learning tasks, e.g. question answering, image classification, instance disambiguation, text summarization, etc.; b) ML as input to KG, whose main goal is to improve the KG itself, e.g. in terms of coverage, quality, and adding new facts. In the following, we analyze the most impactful approaches in the literature, along these two perspectives.

2.1 Knowledge Graphs as Input to Machine Learning

KGs, as representations of background and contextual knowledge in a structured form, have gained significant interest from both academia and industry in the area of machine learning, enabling models to tackle complicated tasks that need prior knowledge [44]. ML models are knowledge-aware and thus can benefit from the incorporation of information that effectively captures the semantic meanings [84]. From traditional ML to modern DNNs, KGs can offer advantages, enhancing the functionality of ML systems by addressing various challenges and solving problems. In the following, we will briefly describe key applications of KGs in ML. Specifically, in Section 2.1.1, we elaborate on the key methodologies for incorporating KGs in ML, with a particular emphasis on the shortcomings they seek to mitigate. In Section 2.1.2, our focus shifts to recent advancements in describing large language models (LLMs) enhancement using KGs, a domain we believe will be increasingly significant in the future, given the widespread adoption of LLMs.

2.1.1 Addressing Machine Learning Challenges with Knowledge Graphs

KGs represent semantic descriptions of entity types and properties with a well-defined meaning. Hence, they can be employed when attempting to automatically extract features (that are difficult to measure or quantify directly) from data points [93, 134]. A feature extractor is a transformation function that maps data from a higher-dimensional space to a lower-dimensional vector space, encompassing a wide range of dimensionality reduction techniques. Early approaches map the output of feature extractors to hierarchies [101, 41] or use hierarchies as input to feature extraction [164], or use large-scale real world labels and their inter-relations [142, 39]. Many recent approaches rely on image annotation that is linked to KGs, such as WordNet [128], like the image databases that have been established based on these concepts (see for example [40, 95]). On the other hand, knowledge graph embedding methods can be also seen as methods to build semantic feature extractors. This involves the mapping of entities and relations into low-dimensional vectors, effectively capturing their semantic meanings in a form that is more compatible to deep learning

Table 1 Overview of the research topics considered, the identified gaps, and our claims and proposals to address them.

Topics	Gaps	Claims & Proposals
LLM prompting	▪ LLM hallucinations ▪ No formalized process to interact with LLMs	▪ Use KG at inference time to formalize the dialogue process between humans and LLM ▪ Ground prompts in knowledge (*e.g.*, adding context, analyzing response, defining prompt sequence)
KG semantics & KGE models	▪ Semantics not (fully) considered ▪ Deductive capabilities not (fully) considered	▪ Investigate the full exploitation of KG semantics (*e.g.*, to improve model performance, to generate informative negatives) ▪ Possibly with different reasoning types (deductive, analogical) ▪ Empirical full assessment of the role of semantics
Symbol-based methods	▪ Largely disregarded ▪ Scalability issues	▪ Leverage mining of disjointness axioms to generate informative negatives needed in ML models training ▪ Alleviate scalability issues
Interpretability of ML models	▪ Pre-/post-model approaches do not fulfil necessary requirements ▪ In-model KG-based explainable approaches not proved to improve interpretability	▪ Infuse KG in ML training ▪ Demonstrate that this improves ML interpretability
Benchmark datasets	▪ Lack of needed characteristics (e.g., schemas) ▪ Some datasets under-used or unnoticed	▪ Develop a unified repository of datasets ▪ Automatically crawl in the wild and qualify datasets w.r.t. needed characteristics ▪ Create synthetic KG generators that generate both tailored schemas and KGs

models [195, 139]. This field of research offers significant opportunities for exploration and advancement [147, 113] and will be analysed in more detail in the next sections.

DNNs require a substantial amount of data for training. Sometimes, the data can either be unavailable or come with a high cost of collection. This issue, commonly referred to as the *sample shortage*, comes with different settings. Among them, the *zero-shot learning* (ZSL) [143] [49, 197] and the *few-shot learning* (FSL) [199] has recently gained significant research attention and call for the use of structured knowledge [71]. ZSL is formally defined as the task of predicting labels for new classes that have never been encountered during training, while FSL involves the task of predicting labels for new classes for which only a small number of labeled samples are provided. In both cases, the proposed solutions try to somehow transfer knowledge from seen classes to unseen classes (see [231] for recent advances on transfer learning, specifically describing knowledge transfer). Here, KGs play an important role, since they can represents background knowledge such as class hierarchies, instances of classes (samples), features, properties, relations as well as meta information like model parameters, providing the necessary auxiliary information. The interested reader can check [197] and [199] for a systematic review on ZSL and FSL, [71] and [27] for ZSL and FSL based on external knowledge (covering some works that use KGs as the background knowledge), [134] covers the use of knowledge graphs specifically for visual transfer learning and [28] that is a recent thorough survey paper that specifically classifies and analyzes methods utilizing KGs for ZSL and FSL.

The capabilities of DNNs have enabled the development of numerous models and techniques to address challenging problems, particularly those involving multimodal data. In this context, multimodal machine learning [14, 61, 133] has emerged as one of the rapidly advancing fields within artificial intelligence, addressing various challenging problems, including visual question answering, visual reasoning, image captioning, image-text retrieval, visual storytelling, visual dialoging and others [3, 66, 220, 207, 171, 45, 45, 96]. Not surprisingly, the proposed DNNs models (mainly based on transformers) often struggle with generalization to various concepts and scenarios that demand commonsense knowledge, or understanding of abstract entities, facts, and real-world events, due to the lack of formal representation of background, contextual and commonsense knowledge [152, 74, 91]. Hence, integrating external knowledge at different stages of multimodal learning, especially in pre-training or fine-tuning, augments the capabilities of models, enabling them to better address a broader range of real-world scenarios. Several proposed DNNs models are based on external knowledge that is represented using semantic descriptions stored in KGs. In particular, there have been proposed datasets that leverage external knowledge [123, 179, 151, 203] linked to web resources and KGs [107] to learn the alignment between visual and textual information [30] in order to solve multimodal learning tasks. The interested reader can find information in several survey papers classifying and analyzing methods in the area of multimodal learning (see for example [14, 61, 133], specifically presenting works that make use of KGs [120]).

The adoption of symbolic knowledge representation and reasoning as a means to address the opacity of machine learning classifiers is a research domain that has recently garnered significant attention from researchers [58]. The need to provide explanations grounded in domain knowledge with formal semantics has driven the utilization of KGs in explainable AI [32, 112, 42, 25, 183]. As this field holds considerable interest and offers numerous prospects for future research, we discuss it in more detail in section 3.4.

2.1.2 Knowledge Graphs for Large Language Models

The current ML literature is dominated by Deep Learning solutions that have been proved very effective in multiple domains and for multiple tasks. Particularly, nowadays LLMs and related systems are catalyzing the attention of the scientific and industrial community for their impressive ability to provide highly accurate results in a very limited amount of time, as for the case of ChatGPT[6] and similar solutions. LLMs behind these systems (like the GPT models [22] that currently guide ChatGPT) are usually deep learning models that have been trained on huge amounts of text data and are capable of understanding and generating human-like text. Typically, they get a text in their input and provide a text as a response. Lately, they can be also directly connected to other generative models like Midjourney[7] and DALLE-3[8] that get text as input and give image or videos in the output, advancing the user experience and extending the scope of application domains.

There are many ways of using KGs to improve or understand the operation of LLMs. There are works that aim to enhance the text generation (see for example the survey [219]) or more generally to enhance visiolinguistic learning with knowledge (see for example the survey [120]). In [144] several methods are discussed that try to unify LLMs and KGs, combining their advantages. Among others, methods that use KGs to improve the operation of LLMs are analysed. An interesting approach is to incorporate knowledge graph information into LLMs in order to enhance their performance, by advancing the factual knowledge understanding. This is a way to improve the LLM performance on knowledge-intensive tasks, and to generate more informed and contextually grounded text. In particular, there are works that try to enhance word representations with knowledge graph embeddings providing context, improving the model's performance [148], or to learn contextualized representations that capture both linguistic and factual knowledge [119], or to use KGs in pre-training to enhance the model's understanding of factual knowledge [176, 110]. Other works in the area try to decompose knowledge into separate modules to improve its natural language understanding capabilities [222], or to integrate KG and language understanding in a joint pre-training framework [218].

Moreover, there are other approaches for graph-to-text generation integrating knowledge from a knowledge graph into the text generation process, trying to produce more informative and coherent outputs [217]. In this framework, combining language representations with knowledge graph embeddings can be used to enhance the representation of contextualised knowledge [175, 173, 65]. Sentiment knowledge can be also incorporated with the use of KGs, thus enhancing the performance of language models with respect to sentiment analysis accuracy [181, 180].

Finally, KGs can be used to prob and possibly understand different aspects of the operation of LLMs. In particular, KGs can be used to elicit knowledge from language models using automatically generated prompts, enabling targeted information retrieval from the model's knowledge base [166], or for querying language models effectively, through a query generation technique that leverages explicit context [2], or to contrastively probing LLMs to investigate the domain knowledge of pretrained language models by comparing their performance to specially designed contrast models [126]. Prompting can be also used for understanding the limitations LLMs, revealing scenarios where language models may produce unreliable or incorrect responses [122], or to enable the exploration and understanding of the underlying knowledge captured by LLMs [178], or to understand how LLMs capture factual knowledge and identify the key factors that contribute to their acquisition of factual information [109].

Of particular significance in this context is the utilization of KGs to validate LLMs, mitigating the issue of hallucination, that causes the generation of factually incorrect content [85]. Hallucination of LLMs poses a substantial challenge to their reliability [15]. Although some LLMs are

[6] https://openai.com/blog/chatgpt
[7] https://www.midjourney.com/home/
[8] https://openai.com/dall-e-3

equipped with the ability to explain their predictions, their explanatory capabilities also suffer from hallucination and this has been particularly connected to the criticism that LLMs have limited ability to encode factual knowledge [232, 191, 56]. Hence, it becomes crucial to examine and authenticate the knowledge embedded within LLMs to prevent hallucination. Recently, there is some work in the area of utilizing KGs for hallucination detection. Specifically, KGs are used as an external source to validate LLMs reliability [86], or to develop fact-checking models, identifying and mitigating hallucinations [48]. This is a very interesting area for future work.

In Section 3.1, we outline our viewpoint on the most important research areas that require attention in order to address the challenges discussed here.

2.2 Machine Learning as Input to Knowledge Graphs

From the perspective of ML as input to KGs, the main objective is to improve the quality of existing KGs overall. Particularly, given the well-known issues concerning noise and incompleteness of KGs, most solutions have focussed on KG refinement which actually encompasses several tasks. Among the others, triple classification (aiming at assessing the correctness of a statement in a KG and generally regarded as a binary classification problem) and mostly link/type prediction (aiming at predicting missing links/types between entities and generally regarded as a learning to rank problem) gained most of the attention, aiming at improving/limiting KG incompleteness.

Different approaches have been developed over the years, with the goal of improving effectiveness (mostly targeting the link prediction problems) while scaling to very large KGs. Mostly, numeric-based methods have been investigated. Among the very first proposals, *probabilistic latent variable models* from the Statistical Relational Learning (SRL) [54] field (having as main goal the creation of statistical models for relational/graph-based data) have been formalized. Successive and very efficient solutions have been represented by *Knowldge Graph Emebedding* (KGE) models. Other approaches focusing on propositionalization techniques, recently also exploiting Graph Neural Networs (GNN) [204]) have been also pursued. Complementary to these numeric-based solutions, research directions targeting symbol-based models have been also proposed, particularly focusing on rule-based methods for predicting triples in KGs.

In the following we summarize the most representative methods for each of the aforementioned categories. We dedicate particular attention to KGE methods that represent the main subject of study for our successive proposals, illustrated in Section 3.2.

2.2.1 Probabilistic Latent Variable Models

Probabilistic Latent Variable Models explain relations between entities by associating each resource to a set of intrinsic latent attributes (i.e. attributes not directly observable in the data) and conditions the probability distribution of the relations between two resources on their latent attributes. All relations are considered conditionally independent given the latent attributes. This allows the information to propagate through the network of interconnected latent variables.

One of the first solutions belonging to this category is the *Infinite Hidden Semantic Model* (IHSM) [155]. It formalises a probabilistic latent variable that associates a latent class variable with each node and makes use of constraints expressed in First Order Logic during the learning process. IHSM showed promising results but resulted limited in scaling on large data collections, because of the complexity of the probabilistic inference and learning, which is intractable in general [92].

2.2.2 Knowledge Graph Embedding Models

KGE models have received considerable attention because of their impressive ability to scale on very large KGs. KGE are numeric-based approaches that convert the data graph into an optimal low-dimensional space in which *graph structural information* and *graph properties* are preserved as much as possible [23, 83]. The embedding procedure consists of learning embeddings such that the score of a valid (positive) triple is lower than the score of an invalid triple, i.e. the invalid triples function as negative examples. Graph embedding methods may differ in their main building blocks: the *representation space* (e.g. point-wise, complex, discrete, Gaussian, manifold), the *encoding model* (e.g. linear, factorization, neural models) and the *scoring function* (that can be based on distance, energy, semantic matching or other criteria) [83]. Over the years, several models have been developed. Some are presented below. It should also be noted that several libraries or frameworks such as Deep Graph Library[9] [194], PyKEEN[10] [6], or PyTorch-BigGraph[11] [106] have been developed and provide unified implementations of wide ranges of models.

One of the first solutions that has been proposed is RESCAL [141], which performs graph embedding by computing a three-way factorization of an adjacency tensor that represents the multi-graph structure of the data collection. It resulted in a powerful model that was also able to capture complex relational patterns over multiple hops in a graph, however it was not able to scale on very large graph-based data collections (e.g. the whole YAGO or DBpedia). The main limitation was represented by the parameter learning phase, which may take rather long for converging to optimal solutions.

The very first highly scalable embedding model is TransE [20]. It introduces a simple but effective and efficient model: each entity is represented by an embedding vector and each predicate is represented by a (vector) *translation operation*. The score of a triple is given by the similarity of the translated subject embedding to the object embedding. The optimal embedding and translation vectors for predicates are learned jointly. The method relies on a *stochastic optimization process*, that iteratively updates the distributed representations by increasing the score of the positive triples i.e. the observed triples, while lowering the score of unobserved triples standing as negative examples. The embedding of all entities and predicates in the KG is learned by minimizing a *margin-based ranking loss*.

Despite its scalability and effectiveness, TransE remained limited in properly representing various types of properties such as *reflexivity*, and 1-to-N, N-to-1 and N-to-N relations that can be easily found in KGs (e.g. typeOf as an example of N-to-N relationship). To tackle this limitation while keeping the ability to scale to very large KGs, a large family of models has been developed that build on TransE, such as TransH [200] and TransR [114].

Specifically, TransR adopts a score function that projects entities into a different vectorial space for each relation through a suitable projection matrix. TransR associates to typeOf, and to all other properties, a specific vector space in which entity vectors are projected. This leads to training specific projection matrices for typeOf (and any other relation) so that the projected entities can be located more suitably to be linked by the vector translation associated to the (typeOf) relation. This differs from TransE, which models typeOf as simple vector translation. The considered individuals and classes may be quite different in terms of the properties and attributes they are involved in, thus determining strong semantic differences (according to [213]) taking place at small reciprocal distances in the underlying vector space, hence revealing the weakness of employing the mere translation.

[9] https://www.dgl.ai/
[10] https://github.com/pykeen/pykeen
[11] https://github.com/facebookresearch/PyTorch-BigGraph

With the goal of capturing additional properties in the data, such as inverse relationship, symmetry, anti-symmetry and composition, more complex embedding models have been formalized, either targeting more complex vector representation spaces, such as the complex representation, as for the case of ComplEx [187] and (Path-)RotatE [227], Gaussian representation, as for the case of KG2E [67] and TransG [206], and manifold representation, as for the case of MuRP [13] and DihEdral [209], or targeting more complex encoding models such as neural models, as for the case of ConvKB [138] and CompGCN [188]. Nevertheless, these additional models became rather computationally expensive, which limits their usefulness.

Nevertheless, several additional semantic aspects that are generally available within KGs, such as hierarchies of concepts and roles, type constraints and transitivity of relationships are currently almost disregarded by existing KGE models. The need for *semantic embedding methods* has been argued [33, 146, 82]. In [60] a KG embedding method considering logical rules has been proposed, where triples in the KG and rules are represented in a unified framework. Specifically, triples are represented as atomic formulae while rules are represented as more complex formulae modelled by t-norm fuzzy logics. A common loss function over both representations is defined, which is minimized to learn the embeddings. This proposal resulted in a novel solution but the specific form of prior knowledge that has to be available constitutes its main drawback. A similar drawback also applies to [130], where a solution based on adversarial training is formalized, exploiting Datalog clauses to encode assumptions which are used to regularize neural link predictors.

Complementary solutions, directly targeting rich representation languages as RDFS and OWL and not requiring additional formalism for representing prior knowledge have been proposed. Particularly, [129] has proven the effectiveness of combining embedding methods and strategies relying on reasoning services for the injection of prior Background Knowledge (BK) to enhance the performance of a specific predictive model. Following this line, TransOWL, aiming at injecting schema level information, particularly during the learning process, and its upgraded version TransROWL, have been formalized [36, 35]. The main focus is on the application of this idea to enhance well-known basic scalable models, namely TransE [20] and TransR [114], even if, in principle, the proposed approach could be applied to more complex embedding methods, with an additional formalization. In TransOWL the original TransE setting is maintained while resorting to reasoning with schema axioms to derive further triples to be considered for training and that are generated consistently with the semantics of the properties. Particularly, for each considered axiom, TransOWL defines, on the score function, specific constraints that guide the way embedding vectors are learned. A set of different axioms, specifically equivalentClass, equivalentProperty, inverseOf and subClassOf, are employed for the definition of constraints on the score function so that the resulting vectors, related to such axioms, reflect their specific properties. As a consequence, new triples are added to the training set on the grounds of the specified axioms. TransROWL further develops TransOWL by adopting TransR as the base model in order to handle non 1-to-1 properties in a more proper way. TransOWL and TransROWL have been proven to improve their effectiveness on link prediction and triple classification tasks when compared to the baseline models (TransE and TransR) that focus on structural graph properties. Some additional efforts in the formalization of KGE and Deep Learning solutions taking into account limited semantics can be found in the literature [57, 12, 72, 62, 100]. Nevertheless, none of the existing KGE model is able to exploit the full expressiveness that a KG may have in principle.

Independently of the specific model, another important issue needs to be highlighted: most of the existing KGs only contain positive (training) examples, since usually false facts are generally not encoded. However, training a learning model in all-positive examples could be tricky, because the model might easily overgeneralize. As such, for obtaining negative examples, that are needed when training KGE models, two different approaches are generally adopted: either *corrupting* true/observed triples randomly, with the goal of generating plausible negative examples or adopting a *local-closed world assumption* (LCWA) in which the data collection is assumed as *locally* complete [140]. In both cases, wrong negative triples may be generated and thus used when training and learning the embedding models.

In Section 3.2, we present our perspective on the research directions that need to be tackled to cope with the problems illustrated particularly in this section.

2.2.3 Neural Methods for Vector Space Embeddings

Another research direction focused on the exploitation of vector space embeddings for obtaining a propositional feature vector representation of a KG. One of the first solutions targeting this research direction is RDF2Vec [156], which adapts the well-known Word2Vec technique, devised for natural language processing, to graph representations. A two-step approach is adopted. First the data graph is converted into a set of sequences of entities (two different approaches can be used for this purpose: graph walks and Weisfeiler-Lehman Subtree RDF graph kernels). In the second step, the obtained sequences are used to train a neural language model to estimate the likelihood of a sequence of entities appearing in a graph. The result is that each entity in the graph is represented as a vector of latent numerical features. In order to show that the obtained vector representation is independent of the downstream task and the specific algorithm, an experimental evaluation involving a number of classification and regression tasks has been performed.

An upgrade of RDF2Vec has been presented in [31], where global patterns are considered (differently from the intial RDF2Vec proposition grounded on local patterns). These solutions cannot cope with literals.

Another way to better capture global information is to use a more powerful model, such as a graph neural network (GNN). These are a class of methods for allowing artificial neural networks to operate on graph data. Given that graphs are a very general data structure, GNNs can take a wide variety of forms. It has also been shown that many popular deep learning architectures, such as convolutional neural networks, recurrent neural networks, and transformers, can be seen as a GNN for a suitably defined graph [21]. In a GNN, as for RDF2Vec and KGE models, nodes are represented as vectors. These vectors are fed through a sequence of message-passing layers, where nodes update their values based on their neighbours' values, and local pooling layers, where groups of neighbouring nodes are combined into a single vector representation. The final layer aggregates the entire input into a single vector representation for the entire graph. Because of this iterative process, GNNs are better able to capture multi-hop relations and global graph structure, compared to RDF2Vec [156]. They are also able to reduce an entire graph to a single embedding vector, as well as computing embedding vectors for each node. See [226] or [229] for an overview of GNN design and applications.

Several works have applied GNNs to construct or enhance KGs. [230] integrates Bellman-Ford into the GNNs training procedure, and then uses the resulting model for link prediction on KGs. [145] show that GNNs can be trained, in a supervised setting, to accurately estimate node importance in a KG. GNNs have also been used for entity alignment, which seeks to discover when the same entity appears in two different knowledge graphs. [201] embeds entities in both KGs and then uses the distance between the embeddings to identify when nodes in different KGs correspond to the same entity. More recent works have built on this method, for example by capturing time-sensitive information [210] or multi-modal inputs [172]. Another common uses of GNNs for KG is to improve the use of KGs in recommender systems [52], and inference [137]. For an overview of the use for GNNs for KGs, see [216].

2.2.4 Rule Learning Solutions

With the goal of finding new facts (namely new triples) that are missing in a KG, AMIE [51, 50][12] has been proposed. AMIE represents one of the most well-known and efficient solutions grounded on a symbol-based approach. Inspired by *association rule mining* [4] and the Inductive Logic Programming (ILP) literature, AMIE is able to learn logic rules from KGs, that are ultimately used for predicting new unseen triples. Interestingly AMIE is tailored to support the Open World Assumption (OWA) characterizing KGs, differently from all numeric-based solutions that are grounded on the Closed World Assumption (CWA). Nevertheless, AMIE mines rules inspecting the triples that are directly observable in the KG and it does not exploit the additional semantics that is available in the KG as well as any form of deductive reasoning.

A related rule mining system, based on a level-wise generate and test strategy has been further proposed [37], with the goal of learning SWRL rules [70] while exploiting schema level information and deductive reasoning capabilities during learning. As for AMIE, the goal was to exploit the discovered rules for predicting new facts. This system actually outperformed AMIE in terms of new predicted triples, and this was due to the exploitation of the schema level information and reasoning capabilities. Nevertheless, they have been also the main cause of the reduced ability of the system to scale on large KGs, when compared to AMIE.

More recently AnyBURL [124] has been proposed. It is a scalable bottom-up rule learning system for KG completion that works by sampling random paths, that are generalized into Horn rules. Reinforcement learning is exploited to guide path sampling and make efficient use of computational resources. AnyBURL also showed improved scalability and competitive performance in comparison to numeric-based approaches. Even more so, it has been also shown that AnyBURL can be used to explain predictions made by a latent model when restricting the types of learned rules. Nevertheless, as for AMIE, no exploitation of the KG semantics and reasoning capabilities can be found.

3 Gaps in Machine Learning and Knowledge Graphs and Next Challenges

In this section we analyse existing gaps of the class of methods illustrated in Section 2 that we identify as important. Hence, for each of them, we provide our perspective on the research directions that need to be pursued in order to fill these gaps. Specifically, the following Section 3.1 primarily focuses on the need of having a clear methodology for interleaving LLMs with KGs and drafts a preliminary proposal. Section 3.2 primarily focuses and provides preliminary proposals for the need of taking into account reasoning capabilities and schema level information of KGs, to be used for setting up a more informative way for generating negative training examples as well as for injecting schema level information in KGE. Beyond the gaps, Section 3.3 presents our view supporting that symbolic ML methods may still have a role in KG, particularly for KG refinement and more specifically for mining disjointness axioms, that are quite often missing in KGs and related ontologies. Section 3.4 presents our position on the need for an approach that demonstrably produces reliable explanations to validate ML results when applied to KGs. Hence, Section 3.5 shows the need for diverse, high-quality benchmark datasets when combining ML and KGs as well as new metrics for capturing new behaviours.

[12] AMIE system is currently at its third version. For more details see `https://github.com/dig-team/amie`

3.1 Knowledge Graphs for Prompting Large Language Models

From what has been described in section 2.1, we understand that the use of KGs, as an additional tool, during the (pre-)training phase or during the inference phase of LLMs are important fields of research, attracting the interest of many researchers, and could potentially improve the operation of the LLM and the results of LLMs, respectively. Although the operation of modern LLMs and respective systems (like chatGPT) is impressive and traditional machine learning gaps (like reasoning capabilities) have started to close, **there is still room for improvement**, and the use of KGs as an additional tool during the training and fine-tuning phases can play an important role, here. Specifically, KGs can provide background knowledge (encyclopaedic, commonsense, domain-specific, multimodal etc), represent human-oriented processes, and explain opaque machine operation. On the other hand, the practical use of LLMs increases dramatically and **there is a great need for advancing the use of LLMs inference, making the process of dialoguing LLMs more formal and systematic**. Therefore, the use of KGs during the phase of the design of the input to be given to LLMs and during the phase of the analysis of the LLM response seems to have a great potential.

Following the above, interesting open problems and challenges is the use of KGs in LLM *prompt engineering* or simply LLM *prompting* [144, 117]. Prompting is the process of providing a sequence of instructions or queries to a LLM in order to get the desired output or to check the LLM's operation and characteristics. It is actually a dialogue between a user (human or agent) and a LLM, that reflects the user's intent and finally results in the desired task or information that the user wants to get from the model. Although the field is new, there are some attempts to formalise the process (see for example the *Automatic Prompt Engineer (APE)* approach [228]). The formalization of the dialogue process should be grounded on some type of background knowledge, so there is a need for representing and using this knowledge. Here, we describe the great potential of using KGs in LLM prompting, based on the nature of prompts, their types and effectiveness, the tasks and the methodology to provide adequate prompts during the prompting process, focusing on the potential use of KGs.

There are many ways to modify the prompt that is given to LLM, using KGs. First, the instruction or question can be more explicit and specific, capturing the user requirements, since it is well-understood that the more specific the prompt the better chance of guiding the LLM to the desired response. For example, the instruction "*Summarise* text A" can be specified as "Summarise the text A *in 200 words*", using the knowledge that an abstract should be between 200 and 300 words. Or the question "Is there any recent paper in the area of prompting machine learning systems?" can be specified as "Is there any recent paper in the area of prompting *LLMs*?". On the other hand, sometimes it may be helpful, depending on the instruction or the question, to generalise it, for example, the question "Is there any recent paper in the area of prompting machine learning systems?" can be generalised as "Is there any recent *work* in the area of prompting machine learning systems?". Also, may be useful to *contextualise* or *style* the prompt, by providing examples ("Suggest romantic musicals, *like "La La Land*"), or conditions ("Suggest papers for prompting LLM, *published in top conferences*"), or style ("Paraphrase text A, *using more formal language*). It is not difficult to see that KGs can be very helpful in constructing knowledge-enhanced prompts like the above (and not restricted to them), guiding prompt changes, as they capture formal domain knowledge descriptions. Interesting ideas can be found in [228] that the instruction generation is framed as natural language program synthesis, in [168] that simple and effective prompts are constructed to improve GPT-3's reliability, in [192] that multi-step reasoning tasks are tackled by constructing planning and solving prompts, in [225] that LLMs are asked to provide explanations for their choices (in this case for a specific task that is model selection) and in [117] that prompting with generated knowledge rectifies model prediction.

Response analysis Another interesting issue that could be considered is to use KGs to characterise the prompt, for example to measure its *effectiveness* or *reliability*, by analysing and evaluating the response. The effectiveness of prompts depends on the response of the LLM, i.e. the answer to a specific prompt in comparison with the desired output, given the task. Depending on the prompt and response languages, it is important to formalise effectiveness or reliability evaluation measures that guide a process of iterative refinement of the results, by using formal knowledge represented in KGs. Interesting ideas can be found in [144, 119, 168].

Prompt sequencing Designing and controlling prompting, i.e. producing a sequence of prompts to elicit a desired output, can be a challenging task that requires a systematic strategy, evaluation and experimentation. Although LLMs are powerful, their operation is complex and unpredictable and thus a dialogue for producing a sequence of prompts may be helpful to understand LLM characteristics, like complex reasoning capabilities. There is lately some work in the area, for example: Chain-of-Thought (CoT) prompts [202] decompose complex reasoning capabilities into a set of simpler reasoning steps; In [117], the usefulness of using knowledge in common sense reasoning is discovered, extracting knowledge from an LLM and then using this knowledge as additional input to refine the prompt result. The APE methodology proposed in [228] uses ideas from program synthesis in order to optimise the prompt selection process, based on efficient score estimations. Future steps would benefit from the use of KGs as formal knowledge representations, because there is a clear requirement formalising the prompting extraction methodology.

3.2 Handling Semantics, Reasoning and Negative Information in Knowledge Graph Embedding Methods

One of the key features of KGs is that they can be enriched with schema-level information. For the purpose ontologies are generally adopted, which coupled with deductive reasoners, allow to make explicit knowledge which is implicitly coded in a KG[13]. For example, given a KG containing the triple <c typeOf Woman> (or equivalently Woman(c), by adopting a Description Logic formalism) and referring to the following simple ontology formalizing a hierarchy of concepts $\mathsf{Man} \sqsubseteq \mathsf{Human}$ and $\mathsf{Woman} \sqsubseteq \mathsf{Human}$, the fact $\mathsf{Human}(c)$ can be derived by the use of a deductive reasoner. Similarly, new knowledge can be derived when additional axioms are available, such as equivalence axioms, disjointness axioms, as well as restrictions on domain and ranges[14]. However, due to the limited ability of reasoners to scale on very large KGs, deductive reasoning is currently almost disregarded.

Indeed, when talking about ML methods coupled with KGs, as for the case of KGE methods, generally only facts that can be directly observed are taken into account e.g. when projecting the data graph into a lower vectorial representation space. This is clearly a limitation, since knowledge that is somehow already available within the KGs (as for the fact Human(c) in the example above) and that may play a role when considering KGE is ignored. For instance, by considering the fact Human(c), a more appropriate vectorial representation for the entity c could be provided thus limiting errors also when solving downstream tasks. By only considering observable facts, schema level information, that is a seminal element of knowledge, and all additional knowledge that can be derived are actually fully dismissed.

[13] Several reasoners exist and may be used for the purpose. Some examples are RDFox (`https://www.oxfordsemantic.tech/rdfox`), HermiT (`http://www.hermit-reasoner.com/`), FaCT++ (`http://owl.cs.manchester.ac.uk/tools/fact/`). See `http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/` for the full list of reasoners.

[14] See `https://www.w3.org/TR/owl2-overview/` for details on the representation language.

Abboud et al. [1] analyzed the shortcomings of the existing embedding models. These shortcomings can be summarized in: theoretical inexpressiveness, lack of support for inference patterns and higher-arity relations, need for logical rule incorporation.

Here, we specifically claim that KGE methods need to be equipped with the full usage of KGs semantics which comprises the exploitation of all axioms that can be found in the ontologies that are used for supplying (rich) schema level information to KGs, as well as the exploitation of deductive reasoning services that allow to obtain additional knowledge both at schema and assertion level. Indeed, whilst the need for *semantic embedding methods* has been advocated [33, 146, 82], only a few proposals can be found in the literature that actually address this problem (see section 2.2.2 for details) and mainly focusing on equivalentClass, equivalentProperty, inverseOf and subClassOf axioms. To the best of our knowledge, none of the existing methods is able to exploit all kinds of axioms that in principle can be found in expressive ontologies. Even more so, a complementary research direction would be needed, calling for a solid and extensive experimental evaluation aiming at providing a clear position on the need (or not) to fully exploit the KG semantics as well as reasoning capabilities. Specifically, we claim that a comprehensive experimental evaluation, involving most of the KGE methods currently available is needed. Two main scenarios should be considered: the first one (currently adopted) where only observable facts are considered; the second one where the full knowledge available within KG is made explicit by considering schema-level information (e.g. transitivity, equivalence axioms, same as axioms etc,) and reasoning capabilities. Hence performances on the very same downstream tasks, adopting the two settings, should be compared, in order to experimentally prove the value added, if any, of exploiting the KGs entirely. Importantly the second scenario could be possibly divided into two intermediate steps, one where knowledge is partially completed by considering the schema level information but no exploitation of deductive reasoners and a second step where the actual full knowledge is gained by adopting available deductive reasoners. This is on one hand, for assessing the impact of the usage of the full knowledge and on the other hand, for assessing if some complexity, due to reasoning, can be saved whilst still trying to make knowledge explicit as much as possible.

Another issue with KGE models is given by the need of negative examples (for training KGE models) that anyhow are generally missing in KGs, where generally only positive information is coded. As illustrated in section 2.2.2, this problem is usually addressed either by *corrupting* true/observed triples randomly, that is by replacing either the subject or the object of the observed triple with an entity picked randomly from the KG, or by adopting a *local-closed world assumption* (LCWA), in which the data collection is assumed as *locally* complete [140]. In both cases, wrong negative triples may be generated and thus used when training and learning the embedding models. In order to mitigate this issue, preliminary proposals tried to take under control the number of negatives that are randomly generated [43]. Clearly this solution does not solve the problem of generating false negatives, it simply try to somehow control the effect of the false negatives. One of the first proposal trying to generate and materialize actual negative triples has been formalized in [8]. Nevertheless, the proposed solution is grounded on the exploitation of additional and external sources of information besides KGs. Specifically, the proposed solution is grounded on two complementary approaches: a statistical ranking for statements obtained based on related entities, and a pattern-based text extraction, applied to search engine query logs.

On the contrary, here we claim that KGs semantics should be fully and solely exploited for making explicit correct negative statements. For instance, given a restriction on domain and/or range of predicate appearing in a true observed triple, the restriction can be exploited for generating negative triples where e.g. the object entity of the negative triple can be deductively proved to be out of the declared range restriction. Similarly, given an observed true triple with a predicate having a functional restriction, negative statements may be generated by constructing triples having objects that are different from the object in the true statement. More generally, the approach for generating correct negative statements that is envisioned, is deeply grounded on the semantics of the schema axioms. The approach should basically construct triples that are in the complement of the set of triples representing the semantics of a given schema axiom.

An initial proposal in this direction can be found in [36, 35, 118], where only domain, range, disjointWith and functionalProperty constraints are considered. Whilst we consider this proposal a valuable way to go as in agreement with the envisioned solution, it needs to be extended for comprising all axioms and constraints that can be possibly found in a KGs, e.g. transitivity, same-as, equivalence axioms, for citing a few. Even more so, we consider it worthwhile to conduct an extensive experimental study comparing the different settings for generating negative examples in order to prove experimentally the actual role of semantics, if any.

Up to now, when referring to reasoning we basically meant deductive reasoning applied to ontologies/KGs [11]. Nevertheless, besides deductive reasoning, other forms of reasoning could be investigated. These different reasoning forms could be useful in KG-related tasks, and conversely, knowledge contained in KGs could be leveraged in their reasoning process. Here we specifically focus on analogical reasoning that is a remarkable capability of the human mind [132] relying on analogical proportions. They are statements of the form "A is to B as C is to D" that can be formalized as quadruples $A : B :: C : D$ [127]. An example of such a quadruple is "leg : human :: paw : dog". Analogical reasoning relies on similarity and dissimilarity to extrapolate knowledge between objects of potentially different domains. To illustrate, the given example quadruple leverages the similarity between body parts and whole, and the relation linking them to constitute a valid analogy. Analogical reasoning is mainly concerned with two tasks: *analogy detection* that aims to determine whether a quadruple $A : B :: C : D$ is a valid analogy, and *analogy solving* that aims to predict a missing element X, given three elements A, B, and C such that $A : B :: C : D$ constitutes a valid analogy. When elements are represented as vectors, analogies can be thought of as parallelograms, *i.e.*, $e_B - e_A = e_D - e_C$. Such a view can thus be adopted with embeddings, which attracted works on ML-based analogy for various Natural Language Processing tasks, *e.g.*, word morphology [7] or machine translation [102]. In the realm of KGs, to the best of our knowledge, only few works consider analogical reasoning. However, KG embeddings are suited for analogical formalization. For instance, by using translations to model relations, TransE inherently validates the parallelogram rule. This motivated Portisch *et al.* [149] to investigate whether some KG embedding models are well-suited for the task of analogy detection with standard analogical datasets. But analogical reasoning could also be directly applied to KGs. In the link prediction task, it is natural to extrapolate edges from one (part of a) KG to another (part), which motivated the ANALOGY model [116]. Interestingly, ANALOGY is based on the parallelogram rule and the authors showed that it subsumes some other models such as DistMult, ComplEx, and HolE. Analogical reasoning can also be considered as an enhancer of existing KGE models by using triples, relations or entities in analogies to enrich the training process [214]. In fact, the integration of analogical reasoning into KG-related tasks and KGE models is not limited to one formalization or one task. Jarnac *et al.* re-used a convolutional model for analogy detection and applied it on pre-trained graph embeddings to select subgraphs of interest from Wikidata to bootstrap a domain-specific KG [81]. Analogies also inherently appear in several other tasks, *e.g.*, Semantic Table Interpretation, matching, or recommendation [135]. It remains to explore both theoretically and empirically the best formalizations, models, improvement in performance, and interactions with other forms of reasoning, especially deductive reasoning that is inherently permitted by ontologies.

3.3 Symbol-based Methods for Knowledge Graphs

Given KGs volumes, **the need for scalable ML solutions has obfuscated the attention to symbol-based ML solutions**. Nevertheless, the important gain, in terms of scalability, that numeric-based methods (such as KGEs) are obtaining is penalizing: a) the possibility to have interpretable models as a result of a learning process (see Section 3.4 for more details); b) the ability to exploit deductive (and complementary forms of) reasoning (see Section 3.2 for more details); c) the expressiveness of the representations to be considered and related assumptions (such as the Open World Assumption (OWA)).

Indeed, suitable symbol-based methods, often inspired by the *Inductive Logic Programming* (ILP) [153] field (aiming at inducing a hypothesised logic program from background knowledge and a collection of examples), have been proposed [34, 87, 104, 51, 182]. Most of them are able to cope with expressive representation languages such as Description Logics (DLs) [11], theoretical foundation for OWL, and the *Open World Assumption* (OWA) typically adopted, differently from the *Closed Wold Assumption* (CWA) that is usually assumed in the traditional ML settings. Also, problems such as ontology refinement and enrichment at terminology/schema level have been proposed [46, 47, 103, 189, 159].

Particularly, with the purpose of enriching ontologies at the terminological level, methods for learning concept descriptions for a concept name have been formalized. The problem has been regarded as a supervised concept learning problem aiming at approximating an intensional DLs definition, given a set of individuals of an ontological KB acting as positive/negative training examples. Various solutions, e.g. DL-Foil[15] [46] and celoe [103] (part of the DL-Learner suite[16]), have been formalized. They are mostly grounded on a *separate-and-conquer* (sequential covering) strategy: a new concept description is built by specializing, via suitable *refinement operators*, a partial solution to correctly cover (i.e. decide a consistent classification for) as many training instances as possible. Whilst DL-Foil works under OWA, celoe works under CWA. Both of them may suffer of ending up in sub-optimal solutions. In order to overcome such issue, DL-Focl[17] [161, 160], Parcel [185] and SpACEL [186] have been proposed. DL-Focl is an optimized version of DL-Foil, implementing a base greedy covering strategy. Parcel combines top-down and bottom-up refinements in the search space. Specifically, the learning problem is split into various sub-problems, according to a divide-and-conquer strategy, that are solved by running celoe as a subroutine. Once the partial solutions are obtained, they are combined in a bottom-up fashion. SpACEL extends Parcel by performing a symmetrical specialization of a concept description. All these solutions proved to be able to learn approximated concept descriptions for a target concept name to be used for possibly introducing new (inclusion or equality) axioms in the KB. Nevertheless, quite often, relatively small ontological KBs have been considered for the experiments, revealing that, currently, they have **limited ability to scale** on very large KGs.

A few scalable exceptions are represented by rule learning systems for KG completion such as AMIE and most of all AnyBURL (see section 2.2.4 for more details). Nevertheless, most of the existing symbol-based methods cannot scale to very large KGs [160].

Here we want to highlight particularly the **role that symbolic ML solutions may play in assessing disjointness axioms** within ontologies. Indeed, disjointness axioms are essential for making explicit the negative knowledge about a domain, yet they are often overlooked during the modelling process [196]. Furthermore, disjointness axioms would be absolutely beneficial for setting up an informed generation of negative examples in KGE models (see section 3.2 for details),

[15] System publicly available at: `https://bitbucket.org/grizzo001/dl-foil/src/master/`
[16] Suite publicly available at: `https://dl-learner.org/`
[17] System publicly available at: `https://bitbucket.org/grizzo001/dlfocl/src/master/`

thus limiting false negatives that random corruption may inject.

To tackle this problem, automated methods for discovering disjointness axioms from the data distribution have been devised. A solution grounded on *association rule mining* [4] has been proposed in [189]. It is based on studying the correlation between classes comparatively, namely by considering *association rules*, *negative association rules* and *correlation coefficient*. Background knowledge and reasoning capabilities are used to a limited extent. A different solution has been proposed in [159, 158], where, moving from the assumption that two or more concepts may be mutually disjoint when the sets of their (known) instances do not *overlap*, the problem has been regarded as a clustering problem, aiming at finding partitions of similar individuals of the knowledge base, according to a *cohesion* criterion quantifying the degree of homogeneity of the individuals in an element of the partition. Specifically, the problem has been cast as a *conceptual clustering* problem, where the goal is both to find the best possible partitioning of the individuals and also to induce intensional definitions of the corresponding classes expressed in the standard representation languages. Emerging disjointness axioms are captured by the employment of *terminological cluster trees* (TCTs) and by minimizing the risk of mutual overlap between concepts. Once the TCT is grown, groups of (disjoint) clusters located at sibling nodes identify concepts involved in candidate disjointness axioms to be derived[18]. Unlike [189], that is based on the statistical correlation between instances, the empirical evaluation of [159, 158] showed the system ability to discover disjointness axioms also involving complex concept descriptions, thanks to the exploitation of the underlying ontology as background knowledge.

Here, we claim that, when tackling the problem of learning disjointness axioms, a two-level analysis needs to be conducted. One level relates to the expressiveness of the axioms that can be learned. The other level is related to the usage of the learned axioms from a user/knowledge engineering perspective. The goal of this two-level analysis should be finding a trade-off between expressiveness and utility from a user modelling perspective. Whilst the former analysis, concerning the expressiveness of the discovered axioms, has been conducted (as reported just above) the latter, requiring an actual user study is currently missing, whilst we consider it necessary for coming up with the aforementioned trade-off between expressiveness and utility of the discovered disjoitness axioms. Furthermore, additional efforts should be devoted to the scalability of the developed methods that, even if not very limited, still they do not appear to be able to scale on the existing KGs.

3.4 Knowledge Graphs for Interpretable Machine Learning

When considering the relation of KGs to deep learning, via KGEs for example, a popular research objective is to use KGs for interpretability. The internal dynamics of DNNs are typically opaque, and there is hope that KGs can be used to help provide (satisfying) explanations of their behaviour. The general goal of producing explanations for behaviour of machine learning models is sometimes referred to as *explainable AI* (XAI).

As argued in [55], the concepts of explainability and interpretability are intertwined in the context of XAI, because what we really seek is an interpretable explanation. One could, for example, detail exactly the activations of each hidden layer in a neural network to explain why it produced the output from the corresponding input, but this is not a human-interpretable explanation, so is unhelpful for XAI. Despite a strong incentive for interpretable machine learning [115], especially in the area of healthcare [131, 5], and despite significant research attention, how to make complex machine learning model interpretable and explainable remains an open problem [88, 111].

[18] System publicly available at: `https://github.com/Giuseppe-Rizzo/TCTnew`

In this section, we give an overview of existing work, and needed future work, on using KGs for interpretable machine learning. We follow our above framework and divide the discussion into two parts: ML for KG and KG for ML.

The former uses ML techniques to augment or construct a KG. With respect to interpretability, the idea is that a KG is a human-readable representation of information. Once it is constructed, it can be used to produce an answer that is highly interpretable, because we can identify the facts and inference rules from which the answer was derived. The problem is that the construction itself, which is often a complex process, remains uninterpretable. The same also applies to work that uses LLMs for KG construction, such as [63, 99], which use BERT-based models to build a clinical KGs for clinical and financial applications, respectively. Once constructed, the KG can perhaps be used in an interpretable way, but the LLM that constructs it is not interpretable. Methods which use then use the KG as input to another stage, may see interpretability gains at those other stages. For example, [16] iteratively use a KG to augment the training data, and then use predictions from augmented training data to extend the KG. However, the initial creation of the KG remains uninterpretable.

In the other direction, there are several works which aim to use KGs to enhance the performance of ML models. There, the possible approaches to using KG for interpretable ML models can, following [154] be divided into three types, pre-model, post-model and in-model.

Pre-model, refers to using the KG as input to a DNN often referred to as "conditioning on the KG", [100]. The idea is that the KG contains higher-quality structured information than images or free-form text, which can then be used by the DNN to solve the given task. This could potentially help interpretability if the network uses an attention mechanism that can be inspected to see which parts of the KG are attended to, as shown by [212] (although, interestingly, the authors were not motivated by explainability in the design of their model). A similar method was later also used by [221]. Similarly, [223] proposed a question-answering model that attends to paths in a KG from a question to the answer, and claims the attention map over these paths constitutes an explanation of the model output. However, these provide at best, only partial interpretability, because it is unclear how/why the model's attention mechanism focuses on the information from the KG that it does.

Post-model, refers to obtaining the output of a ML model, and then invoking a KG to try to produce an explanation for where that output came from. For example, [53] proposes a visual classifier that matches the predicted classes to KG entities, and then uses the KG structure to give an explanation. Similarly, [169] claims to propose an explainable textual entailment model that, after predicting whether one text entails another, finds evidence for this entailment in a KG. The problem with generating post-hoc explanations is that they depend only on the model output and not on the processes internal to the model which produced that output, even though it is precisely the latter that explanations are supposed to shed light on. Two different ML models that produced the same output by very different means would, by methods such as [169] and [53], automatically receive the same 'explanation'. For example, consider two visual classifiers which both assign the same label to an input image. Suppose one of these classifiers has been trained on and memorized the test set, while the other has actually learnt relevant visual features and used these to infer the label. We would surely want the explanation for the outputs of these two classifiers to be different, but if we use only the assigned label to produce an explanation, then they will automatically be the same. Thus, post-model XAI methods that invoke a KG after prediction are precluded from the outset from producing satisfactory explanations, because the explanation is independent of internal model behaviour (given the output), which is exactly the thing we want to explain.

In-model, the third manner of enhancing ML models with a KG, involves the KG during the training of the model itself. In the case of DNNs, this faces the difficulty of connecting discrete data from the KG, to a continuous loss function. Beyond some exploratory work, [100, 165], few methods have attempted this approach. Additionally, even if one successfully improved predictive performance, it is not immediately obvious that it would improve interpretability. It is possible that such an in-model method, were it to be designed, would involve a complex interactive passing of information between a KG and a DNN, which is highly uninterpretable. One such method did explicitly target explanations [162], however this was a bespoke system that requires the KG to consist of part-whole relationships only, as well as additional annotation of the images of object-part classes.

The use of KGs for intepretable ML remains an open problem, either to devise a generalizable method of infusing KG in ML training that demonstrably improves interpretability, or to determine that such a method is not feasible. At the moment, there is interest in the use of KGs for interpretable ML, but we do not have a KG-based method that demonstrably improves interpretability in ML. This gap in the research was also noted by [38]. Moreoever, in order for KGs to be of significant help for explainability, we contend that they must be involved internally in the model itself. Using machine learning to generate KGs means that this generation process itself is not interpretable, and invoking the KG after the operation of the machine learning model means that it cannot distinguish between models that produced the same output, even if by very different means.

3.5 Benchmark datasets, and metrics

The ever-expanding number of available methods targeting KG construction, refinement, or usage in ML approaches entails a need for appropriate benchmark datasets and metrics to evaluate their capabilities. Some datasets are considered as *de facto* standards to evaluate approaches developed for KG-related tasks such as FB15k-237 and WN18RR for link prediction, or Citeseer for node classification. **However, we claim that current datasets do not suffice for a sound and complete evaluation of the capabilities of developed approaches. Indeed, they present several issues such as:**

- unwanted leakages between train and test sets;
- absence of shared patterns between train and test sets;
- lack of necessary characteristics to support the use of background knowledge in ML models (*e.g.*, presence of inverse axioms, hierarchy of classes or properties).
- scattering of datasets across several repositories hindering their discovery and re-usage

In the following, we briefly illustrate and discuss each of these issues and propose possible ways to overcome them.

Several datasets have been made available to the community over the past few years, *e.g.*, by using (fragments of) open KGs [17, 150, 157]. At first, the presence of patterns in train and test sets was regarded with a concern for unwanted leakages. For example, the two datasets FB15k and WN18 were previously widely adopted to evaluate link prediction approaches. It was later discovered that both datasets present data leakage between train and test sets due to inverse relations [43, 184]. A link prediction approach can then easily learn to predict a test triple (t, r^{-1}, h) if triple (h, r, t) is in the train set, where r^{-1} denotes the inverse relation of r. Two filtered versions named FB15k-237 [184] and WN18RR [43] were thus created by filtering such triples, to avoid spurious performance measures. Nevertheless, patterns such as inversion, symmetry, hierarchy or composition and their capture by KGE models are now argued to be of interest, especially if adequately considered in the experimental and evaluation setting [118, 24]. In particular, some authors claim that test triples should be inferrable from patterns learned and premises existing in the train set. This imposes additional constraints when constituting datasets but enables to evaluate the ability of KGE models to efficiently model, capture, and implement those patterns [118, 24]. In this view, train sets should contain samples of premises and conclusions of the considered patterns to learn. Test sets should contain conclusions that can be inferred from patterns learned and premises in train sets. This empirical evaluation is of interest to substantiate some theoretical guarantees of model design or, conversely, to outline some unexpected abilities. For instance, some KGE models such as RotatE [177] are theoretically designed to capture patterns such as symmetry, antisymmetry, inversion, and composition and should be evaluated accordingly. It follows that detecting (and potentially removing) some patterns is an important step of dataset preprocessing. For now, detection (and removal) of inverses is performed statistically, as featured in the AYNEC/AYNEXT system [10, 170]. They detect whether two relations r_1 and r_2 are inverses of each other if some proportions of triples involving r_1 have their counterpart involving r_2. The identification of other patterns also relies on statistical approaches such as rule mining for their detection [118]. It is noteworthy that ontologies provide definitions of inverses, symmetric predicates and hierarchies of properties and classes. Hence, besides statistical approaches, ontological axioms should be taken into account to detect or implement patterns. Indeed, train sets could be completely based on ontological axioms and deductive reasoning to feature the needed patterns to learn or remove some unwanted ones.

Also, we previously outlined the interest in studying the role and usage of background knowledge in ML models. For now, datasets are often regarded as simple graph data without consideration for (or association with) additional knowledge potentially provided by ontologies. Beside improving datasets by adding triples respecting patterns or removing unwanted ones, the association of ontological axioms with datasets could support the development of learning techniques, settings, and models that consider them, following our claim for further consideration of knowledge in KGE models. To illustrate, instead of enriching datasets with triples respecting patterns, models could be evaluated on their ability to consider patterns stated by ontological axioms to predict missing triples in the test set. It is noteworthy that knowledge is already leveraged to enrich the training process in some proposals. For instance, Type-Constrained Negative Sampling [98] replaces the head or the tail of a triple with an entity of the same type when generating negative triples. d'Amato *et al.* [36] use a reasoner to deduce additional triples from axioms defining equivalent classes, equivalent properties, inverses, or subclasses. Similarly, Iana and Paulheim [79] test whether materializing all triples induced by transitive properties, symmetric properties, and sub-properties leads to improved embeddings. Ontological information is also needed to evaluate the semantics captured by KGE models. In this view, Jain *et al.* [80] relies on the existence of types of entities. They learn embeddings on the YAGO3-10 and FB15k-237 datasets with various KGE models and then use these embeddings to predict entity types with classification or clustering approaches. Their analysis shows that semantic representation in the embedding space is not universal across models. In a similar fashion, the DLCC node classification benchmark was introduced to evaluate the capability of classification approaches to reproduce classes defined by Description Logic Constructors [150]. For example, the constructor $\exists r.\top$ is used to group nodes having a particular outgoing relation. Interestingly, they propose two gold standards: one based on the real graph DBpedia and another synthetic standard that is generated by a gold standard generator publicly available. The analysis of ontological information captured by KGE models also motivates new metrics besides traditional metrics such as precision, recall, Hits@K, or Mean Reciprocal Rank. For example, Hubert *et al.* introduced the Sem@K metric [76, 75, 77] to measure the number of predicted triples that respect domain and range of relations among the top-K predicted triples. This metric can thus be seen as measuring the ability of KGE models to capture the semantic profiles of relations. The aforementioned work highlights an interest in using ontological information in KGE model design, learning process, or evaluation. Consequently, we advocate for the further development of benchmark datasets that include various ontological axioms, separately or combined. The availability of such datasets would in turn encourage and support the development of neuro-symbolic methods leveraging such axioms. However, it is noteworthy that not all current benchmarks offer the ontological information that is needed by particular approaches. That is why some authors resort to synthetic KG generators [125, 150], sometimes with a fixed ontology. To further this research direction, synthetic KG generators should be enriched with the synthetic generation of schemas with different levels of expressiveness and constructs. This would allow an on-demand generation of specific ontologies and knowledge graphs featuring the needed ontological axioms.

To further support the research community, we also call for a more systematic approach in the development, characterisation, and collection of benchmark datasets. For now, benchmark datasets (or versions of) are scattered across several repositories such as GitHub or Zenodo. This leads to some of them being widely adopted (*e.g.*, FB15k-237) and some other to be only re-used in a few papers. A unified repository, similar to the UCI Machine Learning repository[19], is needed to encourage their reuse and adoption by the community. Constituting such a repository first requires to crawl (semi-)automatically several sources, including GitHub or Zenodo, and links in papers available in digital libraries, `arXiv`, or `PapersWithCode`. Additionally, given that different approaches may leverage different characteristics of datasets (*e.g.*, DL constructors [150], sub-properties [36, 79], domain and range of predicates [78], patterns in train and test sets [118]), datasets should be qualified w.r.t. the presence or absence of these characteristics. This would help researchers and developers to select suitable datasets to evaluate their approaches. To this aim, scalable automatic methods need to be developed to crawl and analyse KG-based datasets in the wild and detect a broad range of characteristics including those aforementioned. This qualification process will produce metadata that enrich usual dataset metadata such as providers, or licence. To represent these new dataset metadata, an additional perspective thus lies in extending existing ontologies describing datasets (*e.g.*, VoID, DCAT). Ontologies introduced to describe mining processes and their features such as DMOP [89] could offer sources of inspiration in this matter.

4 Conclusion

The interrelation between knowledge graphs and machine learning has been supporting advances in both fields. Machine learning methods have indeed allowed efficient construction and refinement of large knowledge graphs. Conversely, knowledge graphs have been leveraged in various machine learning tasks to improve performance, *e.g.*, in question answering, or image classification.

However, this interrelation still does not consider parts of knowledge graphs and ML methods summarised in Table 1 that we deem important and offering promising research directions. In particular, we believe KGs constitute a major structure for prompting Large Languages Models and could allow researchers to formalise interactions (*e.g.*, providing contexts in prompts, or deciding prompt sequencing). Additionally, rich semantics of KGs and knowledge actionable by various forms of reasoning capabilities could benefit KGE models through a deeper integration. This could lead to improved performance, or a better handling or generation of informative negatives which are essential in model learning. Regarding informative negatives, we also believe that symbol-based ML, which is particularly adapted to the symbolic structure of KGs, could provide an interesting perspective, especially with the mining of disjointness axioms. KGs are human- and machine-interpretable, and thus are a promising structure on which construct in-model interpretable ML models. Nevertheless, the infusion of KGs directly within ML models and an

[19] `https://archive.ics.uci.edu/`

actual demonstration of the production of more interpretable and reliable explanations are open challenges. To assess improved performance or interpretability of ML models thanks to KGs, extensive experimental evaluations are needed, which require datasets showcasing different levels of semantics, or schema constructs to assess their individual impacts. That is why, we also call for a more systematic collection and characterization of datasets, as well as the creation of synthetic KG generators to enrich the collection of available benchmarks.

In our view, such integrations and interactions open promising challenges to foster both fields of research. We believe these directions to be stepping stones to place KGs as central assets towards neuro-symbolic and explainable AI.

References

1 Ralph Abboud, İsmail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. BoxE: A box embedding model for knowledge base completion. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020.*, 2020. URL: `https://proceedings.neurips.cc/paper/2020/hash/6dbbe6abe5f14af882ff977fc3f35501-Abstract.html`.

2 Leonard Adolphs, Shehzaad Dhuliawala, and Thomas Hofmann. How to query language models?, 2021. `arXiv:2108.01928`.

3 Aishwarya Agrawal, Jiasen Lu, Stanislaw Antol, Margaret Mitchell, C. Lawrence Zitnick, Devi Parikh, and Dhruv Batra. VQA: visual question answering - www.visualqa.org. *International Journal of Computer Vision*, 123(1):4–31, 2017. `doi:10.1007/S11263-016-0966-6`.

4 Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, May 26-28, 1993*, pages 207–216. ACM Press, 1993. `doi:10.1145/170035.170072`.

5 Muhammad Aurangzeb Ahmad, Carly Eckert, and Ankur Teredesai. Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB 2018, Washington, DC, USA, August 29 - September 01, 2018*, pages 559–560. ACM, 2018. `doi:10.1145/3233547.3233667`.

6 Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Sahand Sharifzadeh, Volker Tresp, and Jens Lehmann. Pykeen 1.0: A python library for training and evaluating knowledge graph embeddings. *Journal of Machine Learning Research*, 22:82:1–82:6, 2021. URL: `http://jmlr.org/papers/v22/20-825.html`.

7 Safa Alsaidi, Amandine Decker, Puthineath Lay, Esteban Marquer, Pierre-Alexandre Murena, and Miguel Couceiro. A neural approach for detecting morphological analogies. In *8th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2021, Porto, Portugal, October 6-9, 2021*, pages 1–10. IEEE, 2021. `doi:10.1109/DSAA53316.2021.9564186`.

8 Hiba Arnaout, Simon Razniewski, and Gerhard Weikum. Enriching knowledge bases with interesting negative statements. In *Conference on Automated Knowledge Base Construction, AKBC 2020, Virtual, June 22-24, 2020*, 2020. `doi:10.24432/C5101K`.

9 Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. DBpedia: A nucleus for a web of open data. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735. Springer, 2007. `doi:10.1007/978-3-540-76298-0_52`.

10 Daniel Ayala, Agustín Borrego, Inma Hernández, Carlos R. Rivero, and David Ruiz. AYNEC: all you need for evaluating completion techniques in knowledge graphs. In *The Semantic Web - 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2-6, 2019, Proceedings*, volume 11503 of *Lecture Notes in Computer Science*, pages 397–411. Springer, 2019. `doi:10.1007/978-3-030-21348-0_26`.

11 Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *Description Logic Handbook, 2nd edition.* Cambridge University Press, 2010. `doi:10.1017/CBO9780511711787`.

12 Samy Badreddine, Artur S. d'Avila Garcez, Luciano Serafini, and Michael Spranger. Logic tensor networks. *Artificial Intelligence*, 303:103649, 2022. `doi:10.1016/J.ARTINT.2021.103649`.

13 Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. Multi-relational poincaré graph embeddings. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019.*, pages 4465–4475, 2019. URL: `https://proceedings.neurips.cc/paper/2019/hash/f8b932c70d0b2e6bf071729a4fa68dfc-Abstract.html`.

14 Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2019. `doi:10.1109/TPAMI.2018.2798607`.

15 Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. A multitask, multilingual, multimodal evaluation of chatgpt on

reasoning, hallucination, and interactivity, 2023. arXiv:2302.04023.

16 Shreyansh P. Bhatt, Amit P. Sheth, Valerie L. Shalin, and Jinjin Zhao. Knowledge graph semantic enhancement of input data for improving AI. *IEEE Internet Computing*, 24(2):66–72, 2020. doi:10.1109/MIC.2020.2979620.

17 Peter Bloem, Xander Wilcke, Lucas van Berkel, and Victor de Boer. kgbench: A collection of knowledge graph datasets for evaluating relational and multimodal machine learning. In *The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings*, volume 12731 of *Lecture Notes in Computer Science*, pages 614–630. Springer, 2021. doi:10.1007/978-3-030-77385-4_37.

18 Kurt D. Bollacker, Robert P. Cook, and Patrick Tufts. Freebase: A shared database of structured general human knowledge. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*, pages 1962–1963. AAAI Press, 2007. URL: http://www.aaai.org/Library/AAAI/2007/aaai07-355.php.

19 Piero Andrea Bonatti, Stefan Decker, Axel Polleres, and Valentina Presutti. Knowledge graphs: New directions for knowledge representation on the semantic web (dagstuhl seminar 18371). *Dagstuhl Reports*, 8(9):29–111, 2018. doi:10.4230/DAGREP.8.9.29.

20 Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana 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*, pages 2787–2795, 2013. URL: https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html.

21 Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Velickovic. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021. arXiv:2104.13478.

22 Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

23 Hongyun Cai, Vincent W. Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018. doi:10.1109/TKDE.2018.2807452.

24 Yixin Cao, Xiang Ji, Xin Lv, Juanzi Li, Yonggang Wen, and Hanwang Zhang. Are missing links predictable? an inferential benchmark for knowledge graph completion. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 6855–6865. Association for Computational Linguistics, 2021. URL: https://doi.org/10.18653/v1/2021.acl-long.534, doi:10.18653/V1/2021.ACL-LONG.534.

25 Shruthi Chari, Oshani Seneviratne, Mohamed Ghalwash, Sola Shirai, Daniel M. Gruen, Pablo Meyer, Prithwish Chakraborty, and Deborah L McGuinness. Explanation ontology: A general-purpose, semantic representation for supporting user-centered explanations. *Semantic Web*, (to appear), 2023. doi:10.3233/SW-233282.

26 Michel Chein and Marie-Laure Mugnier. *Graph-based Knowledge Representation - Computational Foundations of Conceptual Graphs*. Advanced Information and Knowledge Processing. Springer, 2009. doi:10.1007/978-1-84800-286-9.

27 Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Ian Horrocks, Jeff Z. Pan, and Huajun Chen. Knowledge-aware zero-shot learning: Survey and perspective. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4366–4373. ijcai.org, 2021. doi:10.24963/IJCAI.2021/597.

28 Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Jeff Z. Pan, Yuan He, Wen Zhang, Ian Horrocks, and Huajun Chen. Zero-shot and few-shot learning with knowledge graphs: A comprehensive survey. *Proceedings of the IEEE*, 111(6):653–685, 2023. doi:10.1109/JPROC.2023.3279374.

29 Xiaojun Chen, Shengbin Jia, and Yang Xiang. A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, 141, 2020. doi:10.1016/J.ESWA.2019.112948.

30 Zhuo Chen, Yufeng Huang, Jiaoyan Chen, Yuxia Geng, Yin Fang, Jeff Z. Pan, Ningyu Zhang, and Wen Zhang. Lako: Knowledge-driven visual question answering via late knowledge-to-text injection. In *Proceedings of the 11th International Joint Conference on Knowledge Graphs, IJCKG 2022, Hangzhou, China, October 27-28, 2022*, pages 20–29. ACM, 2022. doi:10.1145/3579051.3579053.

31 Michael Cochez, Petar Ristoski, Simone Paolo Ponzetto, and Heiko Paulheim. Global RDF vector space embeddings. In *The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part I*, volume 10587 of *Lecture Notes in Computer Science*, pages 190–207. Springer, 2017. doi:10.1007/978-3-319-68288-4_12.

32 Roberto Confalonieri and Giancarlo Guizzardi. On the multiple roles of ontologies in explainable AI, 2023. arXiv:2311.04778.

33 Claudia d'Amato. Machine learning for the semantic web: Lessons learnt and next research directions. *Semantic Web*, 11(1):195–203, 2020. doi:10.3233/SW-200388.

34 Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito. Query answering and ontology population: An inductive approach. In *The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008, Proceedings*, volume 5021 of *Lecture Notes in Computer Science*, pages 288–302. Springer, 2008. doi:10.1007/978-3-540-68234-9_23.

35 Claudia d'Amato, Nicola Flavio Quatraro, and Nicola Fanizzi. Embedding models for knowledge graphs induced by clusters of relations and background knowledge. In *Inductive Logic Programming - 30th International Conference, ILP 2021, Virtual Event, October 25-27, 2021, Proceedings*, volume 13191 of *Lecture Notes in Computer Science*, pages 1–16. Springer, 2021. doi:10.1007/978-3-030-97454-1_1.

36 Claudia d'Amato, Nicola Flavio Quatraro, and Nicola Fanizzi. Injecting background knowledge into embedding models for predictive tasks on knowledge graphs. In *The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings*, volume 12731 of *Lecture Notes in Computer Science*, pages 441–457. Springer, 2021. doi:10.1007/978-3-030-77385-4_26.

37 Claudia d'Amato, Andrea G. B. Tettamanzi, and Duc Minh Tran. Evolutionary discovery of multi-relational association rules from ontological knowledge bases. In *Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings*, volume 10024 of *Lecture Notes in Computer Science*, pages 113–128, 2016. doi:10.1007/978-3-319-49004-5_8.

38 Tirtharaj Dash, Sharad Chitlangia, Aditya Ahuja, and Ashwin Srinivasan. A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Scientific Reports*, 12(1):1040, 2022.

39 Jia Deng, Nan Ding, Yangqing Jia, Andrea Frome, Kevin Murphy, Samy Bengio, Yuan Li, Hartmut Neven, and Hartwig Adam. Large-scale object classification using label relation graphs. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I*, volume 8689 of *Lecture Notes in Computer Science*, pages 48–64. Springer, 2014. doi:10.1007/978-3-319-10590-1_4.

40 Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, pages 248–255. IEEE Computer Society, 2009. doi:10.1109/CVPR.2009.5206848.

41 Jia Deng, Jonathan Krause, Alexander C. Berg, and Li Fei-Fei. Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, pages 3450–3457. IEEE Computer Society, 2012. doi:10.1109/CVPR.2012.6248086.

42 Edmund Dervakos, Konstantinos Thomas, Giorgos Filandrianos, and Giorgos Stamou. Choose your data wisely: A framework for semantic counterfactuals. In *International Joint Conference on Artificial Intelligence*, 2023.

43 Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1811–1818. AAAI Press, 2018. URL: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17366.

44 Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 601–610. ACM, 2014. doi:10.1145/2623330.2623623.

45 Alaaeldin El-Nouby, Shikhar Sharma, Hannes Schulz, R. Devon Hjelm, Layla El Asri, Samira Ebrahimi Kahou, Yoshua Bengio, and Graham W. Taylor. Tell, draw, and repeat: Generating and modifying images based on continual linguistic instruction. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 10303–10311. IEEE, 2019. doi:10.1109/ICCV.2019.01040.

46 Nicola Fanizzi, Claudia d'Amato, and Floriana Esposito. DL-FOIL concept learning in description logics. In *Inductive Logic Programming, 18th International Conference, ILP 2008, Prague, Czech Republic, September 10-12, 2008, Proceedings*, volume 5194 of *Lecture Notes in Computer Science*, pages 107–121. Springer, 2008. doi:10.1007/978-3-540-85928-4_12.

47 Nicola Fanizzi, Claudia d'Amato, and Floriana Esposito. Metric-based stochastic conceptual clustering for ontologies. *Information Systems*, 34(8):792–806, 2009. doi:10.1016/J.IS.2009.03.008.

48 Shangbin Feng, Vidhisha Balachandran, Yuyang Bai, and Yulia Tsvetkov. Factkb: Generalizable factuality evaluation using language models enhanced with factual knowledge, 2023. arXiv:2305.08281.

49 Yanwei Fu, Tao Xiang, Yu-Gang Jiang, Xiangyang Xue, Leonid Sigal, and Shaogang Gong. Recent advances in zero-shot recognition: Toward data-efficient understanding of visual content. *IEEE Signal Processing Magazine*, 35(1):112–125, 2018. doi:10.1109/MSP.2017.2763441.

50 Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. Fast rule mining in ontological knowledge bases with AMIE+. *The VLDB Journal*, 24(6):707–730, 2015. doi:10.1007/S00778-015-0394-1.

51 Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, pages 413–422. International World Wide Web Conferences Steering Committee / ACM, 2013. doi:10.1145/2488388.2488425.

52 Yang Gao, Yi-Fan Li, Yu Lin, Hang Gao, and Latifur Khan. Deep learning on knowledge graph for recommender system: A survey, 2020. arXiv:2004.00387.

53 Yuxia Geng, Jiaoyan Chen, Zhiquan Ye, Zonggang Yuan, Wei Zhang, and Huajun Chen. Explainable zero-shot learning via attentive graph convolutional network and knowledge graphs. *Semantic Web*, 12(5):741–765, 2021. doi:10.3233/SW-210435.

54 Lise Getoor and Ben Taskar, editors. *Introduction to Statistical Relational Learning.* MIT Press, 2007.

55 Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael A. Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *5th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2018, Turin, Italy, October 1-3, 2018*, pages 80–89. IEEE, 2018. doi:10.1109/DSAA.2018.00018.

56 Olga Golovneva, Moya Chen, Spencer Poff, Martin Corredor, Luke Zettlemoyer, Maryam Fazel-Zarandi, and Asli Celikyilmaz. ROSCOE: A suite of metrics for scoring step-by-step reasoning. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL: https://openreview.net/pdf?id=xYlJRpzZtsY.

57 Ramanathan V. Guha. Towards A model theory for distributed representations. In *2015 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 22-25, 2015.* AAAI Press, 2015. URL: http://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10220.

58 Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5):93:1–93:42, 2019. doi:10.1145/3236009.

59 Lin Guo and Qun Dai. Graph clustering via variational graph embedding. *Pattern Recognition*, 122:108334, 2022. doi:10.1016/J.PATCOG.2021.108334.

60 Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 192–202. The Association for Computational Linguistics, 2016. doi:10.18653/V1/D16-1019.

61 Wenzhong Guo, Jianwen Wang, and Shiping Wang. Deep multimodal representation learning: A survey. *IEEE Access*, 7:63373–63394, 2019. doi:10.1109/ACCESS.2019.2916887.

62 Víctor Gutiérrez-Basulto and Steven Schockaert. From knowledge graph embedding to ontology embedding? an analysis of the compatibility between vector space representations and rules. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*, pages 379–388. AAAI Press, 2018. URL: https://aaai.org/ocs/index.php/KR/KR18/paper/view/18013.

63 Ayoub Harnoune, Maryem Rhanoui, Mounia Mikram, Siham Yousfi, Zineb Elkaimbillah, and Bouchra El Asri. BERT based clinical knowledge extraction for biomedical knowledge graph construction and analysis, 2023. arXiv:2304.10996.

64 Frederick Hayes-Roth, Donald A Waterman, and Douglas B Lenat. *Building expert systems.* Addison-Wesley Longman Publishing Co., Inc., 1983.

65 Bin He, Daoyuan Zhou, Jingjing Xiao, Xiangyang Jiang, Qun Liu, Nianwen J. Yuan, and Tao Xu. Bert-mk: Integrating graph contextualized knowledge into pre-trained language models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2281–2290, Online, November 2020. Association for Computational Linguistics. URL: https://aclanthology.org/2020.findings-emnlp.207.

66 Feijuan He, Yaxian Wang, Xianglin Miao, and Xia Sun. Interpretable visual reasoning: A survey. *Image and Vision Computing*, 112:104194, 2021. URL: https://doi.org/10.1016/j.imavis.2021.104194, doi:10.1016/J.IMAVIS.2021.104194.

67 Shizhu He, Kang Liu, Guoliang Ji, and Jun Zhao. Learning to represent knowledge graphs with gaussian embedding. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, pages 623–632. ACM, 2015. doi:10.1145/2806416.2806502.

68 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. *Knowledge Graphs.* Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers, 2021. doi:10.2200/S01125ED1V01Y202109DSK022.

69 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys*, 54(4):1–37, 2021.

70 Ian Horrocks, Peter. F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. SWRL: A semantic web rule language combining OWL and RuleML, 2004.

71 Yang Hu, Adriane Chapman, Guihua Wen, and Wendy Hall. What can knowledge bring to machine learning? - A survey of low-shot learning for structured data. *ACM Transactions on Intelligent*

Systems and Technology, 13(3):48:1–48:45, 2022. doi:10.1145/3510030.

72 Zhiwei Hu, Víctor Gutiérrez-Basulto, Zhiliang Xiang, Xiaoli Li, Ru Li, and Jeff Z. Pan. Type-aware embeddings for multi-hop reasoning over knowledge graphs. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 3078–3084. ijcai.org, 2022. doi:10.24963/IJCAI.2022/427.

73 Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. Knowledge graph embedding based question answering. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 105–113, 2019.

74 Zhicheng Huang, Zhaoyang Zeng, Yupan Huang, Bei Liu, Dongmei Fu, and Jianlong Fu. Seeing out of the box: End-to-end pre-training for vision-language representation learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 12976–12985. Computer Vision Foundation / IEEE, 2021. doi:10.1109/CVPR46437.2021.01278.

75 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. Knowledge graph embeddings for link prediction: Beware of semantics! In *Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), Virtual Conference, online, October 24, 2022*, volume 3342 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3342/paper-4.pdf.

76 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. New strategies for learning knowledge graph embeddings: The recommendation case. In *Knowledge Engineering and Knowledge Management - 23rd International Conference, EKAW 2022, Bolzano, Italy, September 26-29, 2022, Proceedings*, volume 13514 of *Lecture Notes in Computer Science*, pages 66–80. Springer, 2022. doi:10.1007/978-3-031-17105-5_5.

77 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. Sem@k: Is my knowledge graph embedding model semantic-aware? *Semantic Web*, (to appear), 2023. doi:10.48550/arXiv.2301.05601.

78 Nicolas Hubert, Pierre Monnin, Armelle Brun, and Davy Monticolo. Treat different negatives differently: Enriching loss functions with domain and range constraints for link prediction, 2023. arXiv:2303.00286.

79 Andreea Iana and Heiko Paulheim. More is not always better: The negative impact of a-box materialization on rdf2vec knowledge graph embeddings. In *Proceedings of the CIKM 2020 Workshops co-located with 29th ACM International Conference on Information and Knowledge Management (CIKM 2020), Galway, Ireland, October 19-23, 2020*, volume 2699 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020. URL: https://ceur-ws.org/Vol-2699/paper05.pdf.

80 Nitisha Jain, Jan-Christoph Kalo, Wolf-Tilo Balke, and Ralf Krestel. Do embeddings actually capture knowledge graph semantics? In *The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings*, volume 12731 of *Lecture Notes in Computer Science*, pages 143–159. Springer, 2021. doi:10.1007/978-3-030-77385-4_9.

81 Lucas Jarnac, Miguel Couceiro, and Pierre Monnin. Relevant entity selection: Knowledge graph bootstrapping via zero-shot analogical pruning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, pages 934–944. ACM, 2023. doi:10.1145/3583780.3615030.

82 Mirantha Jayathilaka, Tingting Mu, and Uli Sattler. Visual-semantic embedding model informed by structured knowledge. In *Proceedings of the 9th European Starting AI Researchers' Symposium 2020 co-located with 24th European Conference on Artificial Intelligence (ECAI 2020), Santiago Compostela, Spain, August, 2020*, volume 2655 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020. URL: https://ceur-ws.org/Vol-2655/paper23.pdf.

83 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514, 2022. doi:10.1109/TNNLS.2021.3070843.

84 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514, 2022. doi:10.1109/TNNLS.2021.3070843.

85 Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):248:1–248:38, 2023. doi:10.1145/3571730.

86 Ziwei Ji, Zihan Liu, Nayeon Lee, Tiezheng Yu, Bryan Wilie, Min Zeng, and Pascale Fung. RHO: reducing hallucination in open-domain dialogues with knowledge grounding. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 4504–4522. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.FINDINGS-ACL.275.

87 J. Józefowska, A Lawrynowicz, and T. Lukaszewski. The role of semantics in mining frequent patterns from knowledge bases in description logics with rules. *Theory and Practice of Logic Programming*, 10(3):251–289, 2010. doi:10.1017/S1471068410000098.

88 S Kandul, V Micheli, J Beck, M Kneer, T Burri, F Fleuret, and M Christen. Explainable ai: A review of the empirical literature. *SSRN 4325219*, 2023.

89 C. Maria Keet, Agnieszka Lawrynowicz, Claudia d'Amato, Alexandros Kalousis, Phong Nguyen, Raúl Palma, Robert Stevens, and Melanie Hilario. The data mining optimization ontology. *Journal of Web Semantics*, 32:43–53, 2015. doi:10.1016/J.WEBSEM.2015.01.001.

90 Mayank Kejriwal. *Domain-specific knowledge graph construction*. Springer, 2019.

91 Wonjae Kim, Bokyung Son, and Ildoo Kim. Vilt: Vision-and-language transformer without convolution or region supervision. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 5583–5594. PMLR, 2021. URL: `http://proceedings.mlr.press/v139/kim21k.html`.

92 Daphne Koller and Nir Friedman, editors. *Probabilistic graphical models: principles and techniques*. MIT Press, 2009.

93 Wei Kun Kong, Xin Liu, Teeradaj Racharak, Guanqun Sun, Jianan Chen, Qiang Ma, and Le-Minh Nguyen. Weext: A framework of extending deterministic knowledge graph embedding models for embedding weighted knowledge graphs. *IEEE Access*, 11:48901–48911, 2023. `doi:10.1109/ACCESS.2023.3276319`.

94 Lili Kotlerman, Ido Dagan, Bernardo Magnini, and Luisa Bentivogli. Textual entailment graphs. *Natural Language Engineering*, 21(5):699–724, 2015.

95 Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123(1):32–73, 2017. `doi:10.1007/s11263-016-0981-7`.

96 Anastasia Kritharoula, Maria Lymperaiou, and Giorgos Stamou. Large language models and multimodal retrieval for visual word sense disambiguation, 2023. `arXiv:2310.14025`, `doi:10.48550/ARXIV.2310.14025`.

97 Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. `doi:10.1145/3065386`.

98 Denis Krompaß, Stephan Baier, and Volker Tresp. Type-constrained representation learning in knowledge graphs. In *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part I*, volume 9366 of *Lecture Notes in Computer Science*, pages 640–655. Springer, 2015. `doi:10.1007/978-3-319-25007-6_37`.

99 Abhijeet Kumar, Abhishek Pandey, Rohit Gadia, and Mridul Mishra. Building knowledge graph using pre-trained language model for learning entity-aware relationships. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 310–315. IEEE, 2020.

100 Ugur Kursuncu, Manas Gaur, and Amit P. Sheth. Knowledge infused learning (K-IL): towards deep incorporation of knowledge in deep learning. In *Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice, AAAI-MAKE 2020, Palo Alto, CA, USA, March 23-25, 2020, Volume I*, volume 2600 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020. URL: `https://ceur-ws.org/Vol-2600/paper14.pdf`.

101 Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, pages 951–958. IEEE Computer Society, 2009. `doi:10.1109/CVPR.2009.5206594`.

102 Philippe Langlais, François Yvon, and Pierre Zweigenbaum. Improvements in analogical learning: Application to translating multi-terms of the medical domain. In *EACL 2009, 12th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, Athens, Greece, March 30 - April 3, 2009*, pages 487–495. The Association for Computer Linguistics, 2009. URL: `https://aclanthology.org/E09-1056/`.

103 Jens Lehmann, Sören Auer, Lorenz Bühmann, and Sebastian Tramp. Class expression learning for ontology engineering. *Journal of Web Semantics*, 9(1):71–81, 2011. `doi:10.1016/J.WEBSEM.2011.01.001`.

104 Jens Lehmann and Lorenz Bühmann. ORE - A tool for repairing and enriching knowledge bases. In *The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part II*, volume 6497 of *Lecture Notes in Computer Science*, pages 177–193. Springer, 2010. `doi:10.1007/978-3-642-17749-1_12`.

105 Douglas B. Lenat, Alan Borning, David W. McDonald, Craig Taylor, and Steven Weyer. Knoesphere: Building expert systems with encyclopedic knowledge. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence. Karlsruhe, FRG, August 1983*, pages 167–169. William Kaufmann, 1983. URL: `http://ijcai.org/Proceedings/83-1/Papers/034.pdf`.

106 Adam Lerer, Ledell Wu, Jiajun Shen, Timothée Lacroix, Luca Wehrstedt, Abhijit Bose, and Alex Peysakhovich. Pytorch-biggraph: A large scale graph embedding system. In *Proceedings of Machine Learning and Systems 2019, MLSys 2019, Stanford, CA, USA, March 31 - April 2, 2019*. mlsys.org, 2019. URL: `https://proceedings.mlsys.org/book/282.pdf`.

107 Paul Lerner, Olivier Ferret, and Camille Guinaudeau. Multimodal inverse cloze task for knowledge-based visual question answering. In *Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2-6, 2023, Proceedings, Part I*, volume 13980 of *Lecture Notes in Computer Science*, pages 569–587. Springer, 2023. `doi:10.1007/978-3-031-28244-7_36`.

108 Guohao Li, Xin Wang, and Wenwu Zhu. Boosting visual question answering with context-aware knowledge aggregation. In *MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020*, pages 1227–1235. ACM, 2020. `doi:10.1145/3394171.3413943`.

109 Shaobo Li, Xiaoguang Li, Lifeng Shang, Zhenhua Dong, Chengjie Sun, Bingquan Liu, Zhenzhou Ji, Xin Jiang, and Qun Liu. How pre-trained language models capture factual knowledge? A causal-inspired analysis. In *Findings of the Association*

for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 1720–1732. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.FINDINGS-ACL.136.

110 Shaobo Li, Xiaoguang Li, Lifeng Shang, Chengjie Sun, Bingquan Liu, Zhenzhou Ji, Xin Jiang, and Qun Liu. Pre-training language models with deterministic factual knowledge. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 11118–11131. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.EMNLP-MAIN.764.

111 Xuhong Li, Haoyi Xiong, Xingjian Li, Xuanyu Wu, Xiao Zhang, Ji Liu, Jiang Bian, and Dejing Dou. Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, 64(12):3197–3234, 2022. doi:10.1007/S10115-022-01756-8.

112 Jason Liartis, Edmund Dervakos, Orfeas Menis-Mastromichalakis, Alexandros Chortaras, and Giorgos Stamou. Searching for explanations of black-box classifiers in the space of semantic queries. *Semantic Web*, (to appear), 2023. doi:10.3233/SW-233469.

113 Yankai Lin, Xu Han, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Knowledge representation learning: A quantitative review, 2018. arXiv:1812.10901.

114 Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, pages 2181–2187. AAAI Press, 2015. doi:10.1609/AAAI.V29I1.9491.

115 Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1):18, 2021. doi:10.3390/E23010018.

116 Hanxiao Liu, Yuexin Wu, and Yiming Yang. Analogical inference for multi-relational embeddings. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 2168–2178. PMLR, 2017. URL: http://proceedings.mlr.press/v70/liu17d.html.

117 Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. Generated knowledge prompting for commonsense reasoning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 3154–3169. Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.ACL-LONG.225.

118 Shuwen Liu, Bernardo Cuenca Grau, Ian Horrocks, and Egor V. Kostylev. Revisiting inferential benchmarks for knowledge graph completion. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, KR 2023, Rhodes, Greece, September 2-8, 2023*, pages 461–471, 2023. doi:10.24963/KR.2023/45.

119 Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. K-BERT: enabling language representation with knowledge graph. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 2901–2908. AAAI Press, 2020. doi:10.1609/AAAI.V34I03.5681.

120 Maria Lymperaiou and Giorgos Stamou. The contribution of knowledge in visiolinguistic learning: A survey on tasks and challenges. In *Proceedings of the AAAI 2023 Spring Symposium on Challenges Requiring the Combination of Machine Learning and Knowledge Engineering (AAAI-MAKE 2023), Hyatt Regency, San Francisco Airport, California, USA, March 27-29, 2023*, volume 3433 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3433/paper18.pdf.

121 Louis Mahon, Eleonora Giunchiglia, Bowen Li, and Thomas Lukasiewicz. Knowledge graph extraction from videos. In *19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020, Miami, FL, USA, December 14-17, 2020*, pages 25–32. IEEE, 2020. doi:10.1109/ICMLA51294.2020.00014.

122 Alex Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Daniel Khashabi, and Hannaneh Hajishirzi. When not to trust language models: Investigating effectiveness of parametric and non-parametric memories. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 9802–9822. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.ACL-LONG.546.

123 Kenneth Marino, Mohammad Rastegari, Ali Farhadi, and Roozbeh Mottaghi. OK-VQA: A visual question answering benchmark requiring external knowledge. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 3195–3204. Computer Vision Foundation / IEEE, 2019. doi:10.1109/CVPR.2019.00331.

124 Christian Meilicke, Melisachew Wudage Chekol, Manuel Fink Patrick Betz, and Heiner Stuckeschmidt. Anytime bottom-up rule learning for large-scale knowledge graph completion. *The VLDB Journal*, 2023. doi:10.1007/s00778-023-00800-5.

125 André Melo and Heiko Paulheim. Synthesizing knowledge graphs for link and type prediction benchmarking. In *The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, volume 10249 of *Lecture Notes in Computer Science*, pages 136–151, 2017. doi:10.1007/978-3-319-58068-5_9.

126 Zaiqiao Meng, Fangyu Liu, Ehsan Shareghi, Yixuan Su, Charlotte Collins, and Nigel Collier. Rewire-then-probe: A contrastive recipe for probing biomedical knowledge of pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 4798–4810.

Association for Computational Linguistics, 2022. doi:10.18653/V1/2022.ACL-LONG.329.

127 Laurent Miclet, Sabri Bayoudh, and Arnaud Delhay. Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *Journal of Artificial Intelligence Research*, 32:793–824, 2008. doi:10.1613/JAIR.2519.

128 George A. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995. doi:10.1145/219717.219748.

129 Pasquale Minervini, Luca Costabello, Emir Muñoz, Vít Novácek, and Pierre-Yves Vandenbussche. Regularizing knowledge graph embeddings via equivalence and inversion axioms. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18-22, 2017, Proceedings, Part I*, volume 10534 of *Lecture Notes in Computer Science*, pages 668–683. Springer, 2017. doi:10.1007/978-3-319-71249-9_40.

130 Pasquale Minervini, Thomas Demeester, Tim Rocktäschel, and Sebastian Riedel. Adversarial sets for regularising neural link predictors. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, August 11-15, 2017*. AUAI Press, 2017. URL: http://auai.org/uai2017/proceedings/papers/306.pdf.

131 Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018.

132 Melanie Mitchell. Abstraction and analogy-making in artificial intelligence, 2021. arXiv:2102.10717.

133 Aditya Mogadala, Marimuthu Kalimuthu, and Dietrich Klakow. Trends in integration of vision and language research: A survey of tasks, datasets, and methods. *Journal of Artificial Intelligence Research*, 71:1183–1317, 2021. doi:10.1613/JAIR.1.11688.

134 Sebastian Monka, Lavdim Halilaj, and Achim Rettinger. A survey on visual transfer learning using knowledge graphs. *Semantic Web*, 13(3):477–510, 2022. doi:10.3233/SW-212959.

135 Pierre Monnin and Miguel Couceiro. Interactions between knowledge graph-related tasks and analogical reasoning: A discussion. In *Workshop Proceedings of the 30th International Conferece on Case-Based Reasoning co-located with the 30th International Conference on Case-Based Reasoning (ICCBR 2022), Nancy (France), September 12-15th, 2022*, volume 3389 of *CEUR Workshop Proceedings*, pages 57–67. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3389/ICCBR_2022_Workshop_paper_75.pdf.

136 Diego Moussallem, Mihael Arcan, Axel-Cyrille Ngonga Ngomo, and Paul Buitelaar. Augmenting neural machine translation with knowledge graphs, 2019. arXiv:1902.08816.

137 Daniel Neil, Joss Briody, Alix Lacoste, Aaron Sim, Páidí Creed, and Amir Saffari. Interpretable graph convolutional neural networks for inference on noisy knowledge graphs, 2018. arXiv:1812.00279.

138 Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Q. Phung. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*, pages 327–333. Association for Computational Linguistics, 2018. doi:10.18653/V1/N18-2053.

139 Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016. doi:10.1109/JPROC.2015.2483592.

140 Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016. doi:10.1109/JPROC.2015.2483592.

141 Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pages 809–816. Omnipress, 2011. URL: https://icml.cc/2011/papers/438_icmlpaper.pdf.

142 Vicente Ordonez, Jia Deng, Yejin Choi, Alexander C. Berg, and Tamara L. Berg. From large scale image categorization to entry-level categories. In *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013*, pages 2768–2775. IEEE Computer Society, 2013. doi:10.1109/ICCV.2013.344.

143 Mark Palatucci, Dean Pomerleau, Geoffrey E. Hinton, and Tom M. Mitchell. Zero-shot learning with semantic output codes. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009.*, pages 1410–1418. Curran Associates, Inc., 2009. URL: https://proceedings.neurips.cc/paper/2009/hash/1543843a4723ed2ab08e18053ae6dc5b-Abstract.html.

144 Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap, 2023. arXiv:2306.08302.

145 Namyong Park, Andrey Kan, Xin Luna Dong, Tong Zhao, and Christos Faloutsos. Estimating node importance in knowledge graphs using graph neural networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pages 596–606. ACM, 2019. doi:10.1145/3292500.3330855.

146 Heiko Paulheim. Make embeddings semantic again! In *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - to - 12th, 2018*, volume 2180 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018. URL: https://ceur-ws.org/Vol-2180/ISWC_2018_Outrageous_Ideas_paper_4.pdf.

147 Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*, 56(11):13071–13102, 2023. doi:10.1007/S10462-023-10465-9.

148 Matthew E. Peters, Mark Neumann, Robert L. Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. Knowledge enhanced contextual word representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 43–54. Association for Computational Linguistics, 2019. doi:10.18653/V1/D19-1005.

149 Jan Portisch, Nicolas Heist, and Heiko Paulheim. Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction - two sides of the same coin? *Semantic Web*, 13(3):399–422, 2022. doi:10.3233/SW-212892.

150 Jan Portisch and Heiko Paulheim. The DLCC node classification benchmark for analyzing knowledge graph embeddings. In *The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings*, volume 13489 of *Lecture Notes in Computer Science*, pages 592–609. Springer, 2022. doi:10.1007/978-3-031-19433-7_34.

151 Chen Qu, Hamed Zamani, Liu Yang, W. Bruce Croft, and Erik G. Learned-Miller. Passage retrieval for outside-knowledge visual question answering. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 1753–1757. ACM, 2021. doi:10.1145/3404835.3462987.

152 Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 2021. URL: http://proceedings.mlr.press/v139/radford21a.html.

153 Luc De Raedt, editor. *Logical and Relational Learning: From ILP to MRDM (Cognitive Technologies)*. Springer-Verlag, 2008.

154 Enayat Rajabi and Kobra Etminani. Knowledge-graph-based explainable ai: A systematic review. *Journal of Information Science*, page 01655515221112844, 2022.

155 Achim Rettinger, Matthias Nickles, and Volker Tresp. Statistical relational learning with formal ontologies. In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II*, volume 5782 of *Lecture Notes in Computer Science*, pages 286–301. Springer, 2009. doi:10.1007/978-3-642-04174-7_19.

156 P. Ristoski and H. Paulheim. RDF2vec: RDF graph embeddings for data mining. In *The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Proceedings, Part I*, volume 9981 of *LNCS*, pages 498–514. Springer, 2016. doi:10.1007/978-3-319-46523-4_30.

157 Petar Ristoski, Gerben Klaas Dirk de Vries, and Heiko Paulheim. A collection of benchmark datasets for systematic evaluations of machine learning on the semantic web. In *The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part II*, volume 9982 of *Lecture Notes in Computer Science*, pages 186–194, 2016. doi:10.1007/978-3-319-46547-0_20.

158 Giuseppe Rizzo, Claudia d'Amato, and Nicola Fanizzi. An unsupervised approach to disjointness learning based on terminological cluster trees. *Semantic Web*, 12(3):423–447, 2021. doi:10.3233/SW-200391.

159 Giuseppe Rizzo, Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito. Terminological cluster trees for disjointness axiom discovery. In *The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, volume 10249 of *Lecture Notes in Computer Science*, pages 184–201, 2017. doi:10.1007/978-3-319-58068-5_12.

160 Giuseppe Rizzo, Nicola Fanizzi, and Claudia d'Amato. Class expression induction as concept space exploration: From DL-Foil to DL-Focl. *Future Generation Computing Systems*, 108:256–272, 2020. doi:10.1016/J.FUTURE.2020.02.071.

161 Giuseppe Rizzo, Nicola Fanizzi, Claudia d'Amato, and Floriana Esposito. A framework for tackling myopia in concept learning on the web of data. In *Knowledge Engineering and Knowledge Management - 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings*, volume 11313 of *Lecture Notes in Computer Science*, pages 338–354. Springer, 2018. doi:10.1007/978-3-030-03667-6_22.

162 Natalia Díaz Rodríguez, Alberto Lamas, Jules Sanchez, Gianni Franchi, Ivan Donadello, Siham Tabik, David Filliat, Policarpo Cruz, Rosana Montes, and Francisco Herrera. Explainable neural-symbolic learning (*X-NeSyL*) methodology to fuse deep learning representations with expert knowledge graphs: The monumai cultural heritage use case. *Information Fusion*, 79:58–83, 2022. doi:10.1016/J.INFFUS.2021.09.022.

163 Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson, 2020. URL: http://aima.cs.berkeley.edu/.

164 Babak Shahbaba and Radford M. Neal. Improving classification when a class hierarchy is available using a hierarchy-based prior, 2005. arXiv:math/0510449.

165 Amit P. Sheth, Manas Gaur, Ugur Kursuncu, and Ruwan Wickramarachchi. Shades of knowledge-infused learning for enhancing deep learning. *IEEE Internet Computing*, 23(6):54–63, 2019. doi:10.1109/MIC.2019.2960071.

166 Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. In *Proceedings of the 2020 Conference on Empirical Methods*

in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020, pages 4222–4235. Association for Computational Linguistics, 2020. doi:10.18653/V1/2020.EMNLP-MAIN.346.

167 Prashant Shiralkar, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. Finding streams in knowledge graphs to support fact checking. In *2017 IEEE International Conference on Data Mining, ICDM 2017, New Orleans, LA, USA, November 18-21, 2017*, pages 859–864. IEEE Computer Society, 2017. doi:10.1109/ICDM.2017.105.

168 Chenglei Si, Zhe Gan, Zhengyuan Yang, Shuohang Wang, Jianfeng Wang, Jordan L. Boyd-Graber, and Lijuan Wang. Prompting GPT-3 to be reliable. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL: https://openreview.net/pdf?id=98p5x51L5af.

169 Vivian Dos Santos Silva, André Freitas, and Siegfried Handschuh. Exploring knowledge graphs in an interpretable composite approach for text entailment. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 7023–7030. AAAI Press, 2019. doi:10.1609/AAAI.V33I01.33017023.

170 Fernando Sola, Daniel Ayala, Rafael Ayala, Inma Hernández, Carlos R. Rivero, and David Ruiz. AYNEXT - tools for streamlining the evaluation of link prediction techniques. *SoftwareX*, 23:101474, 2023. doi:10.1016/J.SOFTX.2023.101474.

171 Matteo Stefanini, Marcella Cornia, Lorenzo Baraldi, Silvia Cascianelli, Giuseppe Fiameni, and Rita Cucchiara. From show to tell: A survey on deep learning-based image captioning. *IEEE Transactions on Patterns Analysis and Machine Intelligence*, 45(1):539–559, 2023. doi:10.1109/TPAMI.2022.3148210.

172 Fenglong Su, Chengjin Xu, Han Yang, Zhongwu Chen, and Ning Jing. Neural entity alignment with cross-modal supervision. *Information Processing and Management*, 60(2):103174, 2023. doi:10.1016/J.IPM.2022.103174.

173 Yan Su, Xu Han, Zhiyuan Zhang, Yankai Lin, Peng Li, Zhiyuan Liu, Jie Zhou, and Maosong Sun. Cokebert: Contextual knowledge selection and embedding towards enhanced pre-trained language models. *AI Open*, 2:127–134, 2021. URL: https://www.sciencedirect.com/science/article/pii/S2666651021000188.

174 Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 697–706. ACM, 2007. doi:10.1145/1242572.1242667.

175 Tianxiang Sun, Yunfan Shao, Xipeng Qiu, Qipeng Guo, Yaru Hu, Xuanjing Huang, and Zheng Zhang. CoLAKE: Contextualized language and knowledge embedding. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 3660–3670. International Committee on Computational Linguistics, 2020. doi:10.18653/V1/2020.COLING-MAIN.327.

176 Yu Sun, Shuohuan Wang, Yu-Kun Li, Shikun Feng, Hao Tian, Hua Wu, and Haifeng Wang. ERNIE 2.0: A continual pre-training framework for language understanding. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8968–8975. AAAI Press, 2020. doi:10.1609/AAAI.V34I05.6428.

177 Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL: https://openreview.net/forum?id=HkgEQnRqYQ.

178 Vinitra Swamy, Angelika Romanou, and Martin Jaggi. Interpreting language models through knowledge graph extraction, 2021. arXiv:2111.08546.

179 Niket Tandon, Gerard de Melo, and Gerhard Weikum. Acquiring comparative commonsense knowledge from the web. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 166–172. AAAI Press, 2014. doi:10.1609/AAAI.V28I1.8735.

180 Duyu Tang, Bing Qin, and Ting Liu. Knowsentiment: Conceptualizing and learning sentiment knowledge. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 45–55, Berlin, Germany, August 2016. Association for Computational Linguistics. URL: https://aclanthology.org/P16-1004.

181 Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. SKEP: sentiment knowledge enhanced pre-training for sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 4067–4076. Association for Computational Linguistics, 2020. URL: https://doi.org/10.18653/v1/2020.acl-main.374, doi:10.18653/V1/2020.ACL-MAIN.374.

182 Ilaria Tiddi, Mathieu d'Aquin, and Enrico Motta. Dedalo: Looking for clusters explanations in a labyrinth of linked data. In *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings*, volume 8465 of *Lecture Notes in Computer Science*, pages 333–348. Springer, 2014. doi:10.1007/978-3-319-07443-6_23.

183 Ilaria Tiddi and Stefan Schlobach. Knowledge graphs as tools for explainable machine learning: A survey. *Artificial Intelligence*, 302:103627, 2022. doi:10.1016/J.ARTINT.2021.103627.

184 Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on*

Continuous Vector Space Models and their Compositionality, CVSC 2015, Beijing, China, July 26-31, 2015, pages 57–66. Association for Computational Linguistics, 2015. doi:10.18653/V1/W15-4007.

185 An C. Tran, Jens Dietrich, Hans W. Guesgen, and Stephen Marsland. An approach to parallel class expression learning. In *Rules on the Web: Research and Applications - 6th International Symposium, RuleML 2012, Montpellier, France, August 27-29, 2012. Proceedings*, volume 7438 of *Lecture Notes in Computer Science*, pages 302–316. Springer, 2012. doi:10.1007/978-3-642-32689-9_25.

186 An C. Tran, Jens Dietrich, Hans W. Guesgen, and Stephen Marsland. Parallel symmetric class expression learning. *Journal of Machine Learning Research*, 18:64:1–64:34, 2017. URL: http://jmlr.org/papers/v18/14-317.html.

187 Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2071–2080. JMLR.org, 2016. URL: http://proceedings.mlr.press/v48/trouillon16.html.

188 Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha P. Talukdar. Composition-based multi-relational graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL: https://openreview.net/forum?id=BylA_C4tPr.

189 Johanna Völker, Daniel Fleischhacker, and Heiner Stuckenschmidt. Automatic acquisition of class disjointness. *Journal of Web Semantics*, 35:124–139, 2015. doi:10.1016/J.WEBSEM.2015.07.001.

190 Denny Vrandecic and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014. doi:10.1145/2629489.

191 Jindong Wang, Xixu Hu, Wenxin Hou, Hao Chen, Runkai Zheng, Yidong Wang, Linyi Yang, Haojun Huang, Wei Ye, Xiubo Geng, Binxing Jiao, Yue Zhang, and Xing Xie. On the robustness of chatgpt: An adversarial and out-of-distribution perspective, 2023. arXiv:2302.12095, doi:10.48550/ARXIV.2302.12095.

192 Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng Lim. Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 2609–2634. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.ACL-LONG.147.

193 Meihong Wang, Linling Qiu, and Xiaoli Wang. A survey on knowledge graph embeddings for link prediction. *Symmetry*, 13(3):485, 2021. doi:10.3390/SYM13030485.

194 Minjie Wang, Lingfan Yu, Da Zheng, Quan Gan, Yu Gai, Zihao Ye, Mufei Li, Jinjing Zhou, Qi Huang, Chao Ma, Ziyue Huang, Qipeng Guo, Hao Zhang, Haibin Lin, Junbo Zhao, Jinyang Li, Alexander J. Smola, and Zheng Zhang. Deep graph library: Towards efficient and scalable deep learning on graphs, 2019. arXiv:1909.01315.

195 Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017. doi:10.1109/TKDE.2017.2754499.

196 Taowei David Wang, Bijan Parsia, and James A. Hendler. A survey of the web ontology landscape. In *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, volume 4273 of *Lecture Notes in Computer Science*, pages 682–694. Springer, 2006. doi:10.1007/11926078_49.

197 Wei Wang, Vincent W. Zheng, Han Yu, and Chunyan Miao. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2):13:1–13:37, 2019. doi:10.1145/3293318.

198 Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. KGAT: knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pages 950–958. ACM, 2019. doi:10.1145/3292500.3330989.

199 Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 53(3):63:1–63:34, 2021. doi:10.1145/3386252.

200 Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 1112–1119. AAAI Press, 2014. doi:10.1609/AAAI.V28I1.8870.

201 Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. Cross-lingual knowledge graph alignment via graph convolutional networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 349–357. Association for Computational Linguistics, 2018. doi:10.18653/V1/D18-1032.

202 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022*, 2022.

203 Jialin Wu and Raymond J. Mooney. Entity-focused dense passage retrieval for outside-knowledge visual question answering. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 8061–8072. Association for Computational Linguistics, 2022. doi:10.18653/v1/2022.emnlp-main.551.

204 Lingfei Wu, Peng Cui, Jian Pei, and Liang Zhao, editors. *Graph Neural Networks: Foundations,*

Frontiers, and Applications. Springer, 2022. doi:10.1007/978-981-16-6054-2.

205 Xiayu Xiang, Zhongru Wang, Yan Jia, and Binxing Fang. Knowledge graph-based clinical decision support system reasoning: A survey. In *Fourth IEEE International Conference on Data Science in Cyberspace, DSC 2019, Hangzhou, China, June 23-25, 2019*, pages 373–380. IEEE, 2019. doi:10.1109/DSC.2019.00063.

206 Han Xiao, Minlie Huang, and Xiaoyan Zhu. TransG : A generative model for knowledge graph embedding. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics, 2016. doi:10.18653/V1/P16-1219.

207 Ning Xie, Farley Lai, Derek Doran, and Asim Kadav. Visual entailment task for visually-grounded language learning, 2018. arXiv:1811.10582.

208 Chenyan Xiong, Russell Power, and Jamie Callan. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 1271–1279. ACM, 2017. doi:10.1145/3038912.3052558.

209 Canran Xu and Ruijiang Li. Relation embedding with dihedral group in knowledge graph. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 263–272. Association for Computational Linguistics, 2019. doi:10.18653/V1/P19-1026.

210 Chengjin Xu, Fenglong Su, and Jens Lehmann. Time-aware graph neural networks for entity alignment between temporal knowledge graphs, 2022. arXiv:2203.02150, doi:10.48550/ARXIV.2203.02150.

211 Da Xu, Chuanwei Ruan, Evren Körpeoglu, Sushant Kumar, and Kannan Achan. Product knowledge graph embedding for e-commerce. In *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020*, pages 672–680. ACM, 2020. doi:10.1145/3336191.3371778.

212 Bishan Yang and Tom M. Mitchell. Leveraging knowledge bases in lstms for improving machine reading. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1436–1446. Association for Computational Linguistics, 2017. doi:10.18653/V1/P17-1132.

213 Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li 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*, 2015. URL: http://arxiv.org/abs/1412.6575.

214 Zhen Yao, Wen Zhang, Mingyang Chen, Yufeng Huang, Yi Yang, and Huajun Chen. Analogical inference enhanced knowledge graph embedding. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 4801–4808. AAAI Press, 2023. doi:10.1609/AAAI.V37I4.25605.

215 Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. QA-GNN: reasoning with language models and knowledge graphs for question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 535–546. Association for Computational Linguistics, 2021. doi:10.18653/V1/2021.NAACL-MAIN.45.

216 Zi Ye, Yogan Jaya Kumar, Goh Ong Sing, Fengyan Song, and Junsong Wang. A comprehensive survey of graph neural networks for knowledge graphs. *IEEE Access*, 10:75729–75741, 2022. doi:10.1109/ACCESS.2022.3191784.

217 Jason Youn and Ilias Tagkopoulos. KGLM: integrating knowledge graph structure in language models for link prediction. In *Proceedings of the The 12th Joint Conference on Lexical and Computational Semantics, *SEM@ACL 2023, Toronto, Canada, July 13-14, 2023*, pages 217–224. Association for Computational Linguistics, 2023. doi:10.18653/V1/2023.STARSEM-1.20.

218 Donghan Yu, Chenguang Zhu, Yiming Yang, and Michael Zeng. JAKET: joint pre-training of knowledge graph and language understanding. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 11630–11638. AAAI Press, 2022. doi:10.1609/AAAI.V36I10.21417.

219 Wenhao Yu, Chenguang Zhu, Zaitang Li, Zhiting Hu, Qingyun Wang, Heng Ji, and Meng Jiang. A survey of knowledge-enhanced text generation. *ACM Computing Surveys*, 54(11s):227:1–227:38, 2022. doi:10.1145/3512467.

220 Rowan Zellers, Yonatan Bisk, Ali Farhadi, and Yejin Choi. From recognition to cognition: Visual commonsense reasoning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 6720–6731. Computer Vision Foundation / IEEE, 2019. URL: http://openaccess.thecvf.com/content_CVPR_2019/html/Zellers_From_Recognition_to_Cognition_Visual_Commonsense_Reasoning_CVPR_2019_paper.html, doi:10.1109/CVPR.2019.00688.

221 Kunli Zhang, Linkun Cai, Yu Song, Tao Liu, and Yueshu Zhao. Combining external medical knowledge for improving obstetric intelligent diagnosis: model development and validation. *JMIR medical informatics*, 9(5):e25304, 2021.

222 Tong Zhang, Cheng Wang, Ning Hu, Minlie Qiu, Chen Tang, Xiaodong He, and Jian Huang. Dkplm: Decomposable knowledge-enhanced pretrained language model for natural language understanding. In *Thirty-Sixth AAAI Conference on Ar-*

tificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelfth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, pages 11 703–11 711. AAAI Press, February 22 - March 1 2022. URL: `https://ojs.aaai.org/index.php/AAAI/article/view/21425`.

223 Yingying Zhang, Shengsheng Qian, Quan Fang, and Changsheng Xu. Multi-modal knowledge-aware hierarchical attention network for explainable medical question answering. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1089–1097, 2019.

224 Yang Zhao, Lu Xiang, Junnan Zhu, Jiajun Zhang, Yu Zhou, and Chengqing Zong. Knowledge graph enhanced neural machine translation via multi-task learning on sub-entity granularity. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 4495–4505. International Committee on Computational Linguistics, 2020. `doi:10.18653/V1/2020.COLING-MAIN.397`.

225 Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc V. Le, and Ed H. Chi. Least-to-most prompting enables complex reasoning in large language models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL: `https://openreview.net/pdf?id=WZH7099tgfM`.

226 Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81, 2020.

227 Xiaohan Zhou, Yunhui Yi, and Geng Jia. Path-rotate: Knowledge graph embedding by relational rotation of path in complex space. In *10th IEEE/CIC International Conference on Communications in China, ICCC 2021, Xiamen, China, July 28-30, 2021*, pages 905–910. IEEE, 2021. `doi:10.1109/ICCC52777.2021.9580273`.

228 Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are human-level prompt engineers. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL: `https://openreview.net/pdf?id=92gvk82DE-`.

229 Yu Zhou, Haixia Zheng, Xin Huang, Shufeng Hao, Dengao Li, and Jumin Zhao. Graph neural networks: Taxonomy, advances, and trends. *ACM Transactions on Intelligent Systems and Technology*, 13(1):15:1–15:54, 2022. `doi:10.1145/3495161`.

230 Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal A. C. Xhonneux, and Jian Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 29476–29490, 2021.

231 Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021. `doi:10.1109/JPROC.2020.3004555`.

232 Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. Exploring AI ethics of chatgpt: A diagnostic analysis, 2023. `arXiv:2301.12867`, `doi:10.48550/ARXIV.2301.12867`.