

Neurosymbolic AI for Reasoning over Knowledge Graphs: A Survey

Lauren Nicole DeLong, Ramon Fernández Mir, Jacques D. Fleuriot

Abstract—Neurosymbolic AI is an increasingly active area of research that combines symbolic reasoning methods with deep learning to leverage their complementary benefits. As knowledge graphs are becoming a popular way to represent heterogeneous and multi-relational data, methods for reasoning on graph structures have attempted to follow this neurosymbolic paradigm. Traditionally, such approaches have utilized either rule-based inference or generated representative numerical embeddings from which patterns could be extracted. However, several recent studies have attempted to bridge this dichotomy to generate models that facilitate interpretability, maintain competitive performance, and integrate expert knowledge. Therefore, we survey methods that perform neurosymbolic reasoning tasks on knowledge graphs and propose a novel taxonomy by which we can classify them. Specifically, we propose three major categories: (1) logically-informed embedding approaches, (2) embedding approaches with logical constraints, and (3) rule learning approaches. Alongside the taxonomy, we provide a tabular overview of the approaches and links to their source code, if available, for more direct comparison. Finally, we discuss the unique characteristics and limitations of these methods, then propose several prospective directions toward which this field of research could evolve.

Index Terms—Neurosymbolic AI, knowledge graphs, representation learning, hybrid AI, graph neural networks.

I. INTRODUCTION

KNOWLEDGE GRAPHS (KGs) are becoming an increasingly popular way to represent information because they are multi-relational, easily queried, and can store various types of data in a similar, consistent format [148]. Furthermore, numerous methods have been developed to mine information and make novel predictions from KGs [72, 154], leading to advancements in drug discovery [129, 168], improved user recommendation systems [160], and more efficient traffic forecasting [54, 155], amongst other applications. Traditionally, to make novel predictions about a knowledge base, approaches utilized either a set of rules [45, 110] or generated numerical representations from which patterns could be extracted [112, 143]. Recently, however, newer methods for knowledge discovery and reasoning on graphs have attempted to blend these categories into novel neurosymbolic, or hybrid, methods.

Neurosymbolic artificial intelligence (AI), an increasingly active field of research, describes the combination of symbolic AI, which often includes logic and rule-based approaches, with neural networks and deep learning [47, 65]. Often, neurosymbolic AI’s main advantages are described as achieving comparable performance to current deep learning methods while simultaneously fostering inherent interpretability [47,

144, 65]. However, there are many different ways in which this new field has been approached. With this article, we aim to provide a comprehensive overview and structural classification of neurosymbolic methods designed for reasoning over KGs by describing their:

- overarching taxonomy,
- benefits and weaknesses,
- applications for which they are already used, and
- potential research directions.

Specifically, we classify approaches into three major categories: (1) logically-informed embedding approaches (§IV-A), which use symbolic inference and deep learning sequentially, (2) approaches which learn embeddings with logical constraints (§IV-B), and (3) those which learn logical rules for reasoning (§IV-C). Moreover, by distinguishing five *critical characteristics* of these approaches, namely INTERPRETABILITY, GUIDED TRAINING via rules and knowledge, the ability to encode UNDERREPRESENTED TYPES and LONG-RANGE DEPENDENCIES, as well as the efficient aggregation of heterogeneous information (HETEROGENEOUS AGGREGATION), we explore how each category of approaches has unique capabilities for various facets of research.

A. Related Work and Novel Contributions

This survey is the first to provide a full, comprehensive overview of neurosymbolic methods for reasoning over KGs. Because the topic is still relatively young, earlier articles tend to focus largely on background information. For example, an article by Zhang *et al.* [166] discusses general neurosymbolic AI, neural-network-based graph reasoning, and symbolic graph reasoning. However, it only covers a subset of neurosymbolic methods for reasoning on KGs. Similarly, a paper by Chen *et al.* [24] provides an expansive overview of the information present in our *Background* (§II) and *KG Completion* (§III) sections. Previously, a relevant review by Lamb *et al.* [85] described how several types of graph neural networks already possess tools, such as attention mechanisms, to be considered or used for neurosymbolic purposes. Additionally, work by Boschin *et al.* [15] overlaps with our paper. However, we take a different approach toward grouping and describing neurosymbolic methods for KGs. In particular, while Boschin *et al.* focus on grouping methods by the various rules or domain-specific knowledge used, we outline a taxonomy in terms of the structure of the architectures, including the orders in which symbolic and neural modules are placed and integrated.

II. BACKGROUND

A. A Brief Introduction to Knowledge Graphs

A *graph* or *network* structure G is composed of a finite set of vertices or nodes, V , and a finite set of edges, E , connecting the vertices. A graph G can be represented, then, as a tuple of these two sets, (V, E) . Each edge $(u, v) \in E$ connects two *adjacent* or *neighbouring* vertices, u and $v \in V$. The term *triple* denotes a pair of vertices connected by an edge [59].

A *knowledge graph* (KG) uses a graph structure to represent a *knowledge base* (KB), a collection of factual triples denoting relations between real-world entities [72]. KGs have been employed in research domains from biochemistry and medical data [37, 48] to social media platforms [69]. They may be as simple as nodes and edges, or they may additionally contain semantic information, relation and entity types, and node and edge properties [25, 72]. KGs have become one of the most popular ways to represent information, as they capture multi-relational data well, are fast and easy to query when stored in a graph database, and allow the representation of many different types of data in a similar format (a “universal language”). Some of the best known KGs, such as YAGO [138] and DBpedia [8], contain millions of entities and relations representing general information about people, places, and other general items such as movies and music and are widely used as benchmark datasets. Other smaller KGs exist to store information about niche areas, such as the COVID-19 pandemic [37, 129].

KGs are also becoming an increasingly popular way to represent *growing* or *changing* KBs due to their flexible storage schemas [148]. However, because KGs house information that we, as people, know about some given domain, they are inherently incomplete in comparison to the real world. The practice of adding new information to a graph, known as *KG completion*, is, therefore, a common way to make predictions, a topic discussed further within §III-A.

B. A Brief Introduction to Logic-Based Reasoning

While the “symbolic” components of neurosymbolic approaches vary, many of them utilize some form of *logic-based reasoning*. We can think of a logic as a formal system comprised of syntax, semantics and inference rules. Different logics may suit different applications either because of their abilities to represent interesting properties or because of their levels of automation. In this section, we will cover logic programming, probabilistic logic programming, fuzzy logic and description logic, highlighting how they are used for reasoning on graph structures.

Logic programming [10] is one of the most prominent and widely used symbolic reasoning techniques. The fundamental building blocks of this paradigm are *Horn clauses* [68], which are formulas of the form:

$$H \leftarrow B_1 \wedge \dots \wedge B_n.$$

These are meant to be interpreted as rules where the body $B_1 \wedge \dots \wedge B_n$ implies the (positive) head H , or as facts, if the body is empty. In its most basic form, every element is an atom and all the free variables are implicitly universally

quantified. For example, consider the following background knowledge and rule:

$$\begin{aligned} &\text{likes}(alice, bob), \\ &\text{likes}(bob, alice), \\ &\text{friends}(X, Y) \leftarrow \text{likes}(X, Y) \wedge \text{likes}(Y, X). \end{aligned}$$

Intuitively, we can deduce that $\text{friends}(alice, bob)$.

From this, an interesting question arises: given some background knowledge and a set of positive and negative examples, can we come up with consistent rules? This is the approach in *inductive logic programming* (ILP) [103, 172]. The First-Order Inductive Learner (FOIL) [115] is a classical ILP implementation that iteratively creates rules using simple heuristics to pick candidates. The basic approach relies on clean data and a small knowledge base, although there are relevant tools such as Progol [102] or Aleph [136] that improve on this classical approach. There are also many variations and extensions of ILP using advanced statistical methods that refine the search procedure, for instance, SHERLOCK [127] and AMIE [45], which is mentioned again in §III-B. Cropper *et al.* [30] provide a comprehensive survey of ILP, and Zhang *et al.* survey its use in explainable AI [172]; we refer the interested reader to those surveys for further information.

A useful extension to Horn clauses to guide the search is to consider *probabilistic logic* [106] with *soft rules* [104, 57], where a probability $\in [0, 1]$ is attached, e.g.

$$0.75 :: \text{car}(X) \leftarrow \text{wheels}(X, 4) \wedge \text{drivable}(X)$$

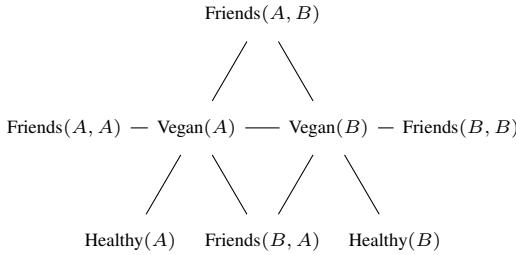
says that we are 75% certain that four-wheeled drivable objects are cars. Many tools take this probabilistic approach as it allows noisy data and uncertainty, and takes advantage of statistical methods. A good example is ProbLog [34] where all the rules are soft and independent, and the resolution algorithm is extended to keep track of probabilities.

Alternatively, we can attach a weight w_i to every first-order formula F_i in a knowledge base, i.e. $(F_i, w_i)_{i=1}^n$. Assuming that our language has finitely many constants, we can now define a graph called a *Markov Logic Network* (MLN) [117]. First, recall that an *atom* is a formula composed of constants, variables and functions but with no logical connectives, and that a formula is *grounded* if it has no variables. The nodes of a MLN are every grounding of every atom appearing in the F_i s, and there is an edge if the ground atoms appear together in at least one of the F_i s. We will illustrate the concept by adapting an example from Domingos *et al.* [38]. Suppose we have constants A and B and the following formulas:

$$\begin{aligned} &\forall x. \text{Vegan}(x) \Rightarrow \text{Healthy}(x) \\ &\forall x. \forall y. \text{Friends}(x, y) \Rightarrow (\text{Vegan}(x) \Leftrightarrow \text{Vegan}(y)) \end{aligned}$$

with weights 1.3 and 0.7 respectively.

The ground Markov network looks as follows:



A possible *world*, x , is a truth value assignment to each node. The probability distribution over the possible worlds, denoted by the random variable X , is given by

$$P(X = x) = \frac{\exp(\sum_{i=1}^n w_i n_i(x))}{\sum_y \exp(\sum_{i=1}^n w_i n_i(y))}$$

where n is the number of formulas and $n_i(x)$ is the number of true groundings of F_i in x . In our example, let x be the world where A and B are vegan, unhealthy and have no friends. We can compute that $P(X = x) \approx 0.00055$. With this machinery in place, we can answer queries such as the probability that A is healthy.

Instead of associating a weight (or probability) to each formula, representing the “degree of belief”, we can also attach a value to the “degree of truth” of the statement. This approach is known as *fuzzy logic*, in which the interpretation of the usual boolean logical connectives is adjusted [163]. For instance, a fuzzy version of $x \wedge y$ could be $\min(x, y)$ where x and y are in the range $[0, 1]$. In general, any binary operation on $[0, 1]$ that is commutative, monotone, associative, and respects the identity is called a *t-norm* and can represent conjunction in fuzzy logic. It is also possible to go back and forth while preserving some of the semantics. Doing so allows us to apply statistical inference to an originally discrete problem and, conversely, logical reasoning to uncertain data. As we will see in §IV-B2, fuzzy inference systems can be used to increase explainability in deep learning systems [13].

Another angle is to consider specific fragments of first-order logic with different levels of expressive power. *Description logics* [82] are *knowledge representation languages* that have a rich syntax to talk about concepts, hierarchies, and cardinalities. Therefore, they provide a convenient semantic framework to reason over KGs, particularly in an ontological sense. In fact, we can think of ontologies as data schemas for KGs. The W3C Web Ontology Language (OWL) [97] is an important implementation of this idea. Originally, OWL was designed to describe the Semantic Web, but now it is integral to several neurosymbolic systems [62]. There are also variations such as OWL 2 EL, designed with biomedical ontologies in mind, which we will see in §IV-A. Here, EL means “existential language”, so in OWL 2 EL we can have existential but not universal quantification. These restrictions enable polynomial time reasoning, which is crucial for large knowledge bases. In this context, reasoners are pieces of software that can infer logical consequences from a set of formulas.

C. A Brief Introduction to Neural Networks

As the name suggests, the “neural” components of neurosymbolic approaches comprise neural network-based structures.

Recently, much state-of-the-art work in machine learning (ML) and AI involves artificial neural networks (NN) [121, 141] due to their impressive abilities to learn functions that capture underlying patterns in data [81, 126]. A NN uses a series of nodes, or *neurons*, with activation functions to act as some *firing threshold*. Similarly to biological neurons, if the input into each activation function exceeds that threshold, then there is an output. Specifically, *deep* NNs are those with two or more layers [141, 134].

In recent years, many deep NN variants have sprouted to handle various data structures and accomplish novel goals. Convolutional NNs (CNNs) [81], for example, are generally used for image data, recurrent NNs (RNNs) with Long Short-Term Memory (LSTM) are ideal for dynamic temporal data [67, 133], and several forms of graph NNs (GNN) have been designed to handle multi-relational data in graph structures, such as factual triples in KGs, as described previously in §II-A [154]. However, many NNs are *black-box* models which lack interpretability [170]. This means that a human cannot inherently understand the internal processes that lead to the model output. With multiple layers and sometimes up to millions or billions of parameters [16, 27], it becomes infeasible for a human to follow the individual actions taken by the model to derive predictions. In some cases, such as biomedical applications in which datasets may contain demographic bias, this may be ethically wrong or even dangerous [109, 108]. As a result, many current efforts centre upon adding some aspect of interpretability to such methods [170].

D. Neurosymbolic AI: A Hybrid Approach

Neurosymbolic AI is the field of research that studies the combination of deep learning and symbolic reasoning [144, 47, 65]. The argument for this hybrid approach is that neural and symbolic systems can complement each other and mitigate their respective weaknesses. For many applications, it is highly desirable to combine accountable and interpretable logic-based modules with effective deep learning ones. While interpretability is defined in various ways across the literature, it is generally viewed as the ability to be understood by a human [101, 170]. Essentially, an end-user might easily comprehend the reasoning processes that led to model predictions.

Based on the classification given by Kautz at AAAI-20¹ (available as a written summary [75]), neurosymbolic integration is often categorized as follows [125]:

- 1) SYMBOLIC NEURO SYMBOLIC: a NN takes a symbolic representation as input and reconstructs another symbolic representation as output.
- 2) SYMBOLIC[NEURO]: systems where a symbolic engine queries a NN during reasoning to, for example, estimate a utility function.
- 3) NEURO;SYMBOLIC: the NN and the symbolic module solve complementary tasks and communicate frequently to guide each other.
- 4) NEURO:SYMBOLIC → NEURO: symbolic and neural components are tightly-coupled, but symbolic knowledge is “compiled” into the training set.

¹<https://roc-hci.com/announcements/the-third-ai-summer/>

5) NEURO_{SYMBOLIC}: symbolic logic rules provide the template for the structure of the NN by representing them as tensor embeddings.

6) NEURO[SYMBOLIC]: neural engines capable of logical reasoning at certain points in the execution.

To connect our survey back to this preliminary classification of neurosymbolic systems, we will refer back to which of these categories, if any, each section of approaches is most aligned. Generally, in §IV-A and §IV-C, the approaches most resemble the NEURO:SYMBOLIC → NEURO and NEURO;SYMBOLIC categories. Systems that train with logical constraints are explained in §IV-B, and fall mostly into the NEURO_{SYMBOLIC} and NEURO[SYMBOLIC] classes.

III. A HISTORY OF KNOWLEDGE GRAPH COMPLETION

While there are various reasoning tasks which could be executed upon KGs, those which perform KG completion (see below) are especially common in the literature, and, more particularly, in the approaches surveyed in the next sections. Thus, we provide a brief background to understand the main body of this article.

A. KG Completion and Reasoning

Since KGs house human knowledge about some given domain, they are notoriously incomplete. KGs could contain false positives based on misinformation and incorrect assumptions. Furthermore, it is often unclear as to whether the absence of details in the KG should indicate missing information or the negation of such information [25]. *KG completion* (KGC) denotes a series of methods which can be used to refine the graphs and uncover novel information. For example, such completion can reveal unknown molecular functions in a biological KG [129, 37] or predict user connections in a social one [69].

In particular, *link prediction* is a popular type of KGC which determines whether an edge exists between a given pair of nodes (*e.g.*, whether two users of a social network might be acquainted in real life). Another variant, known as *relation prediction*, aims to infer the existence of links of specific relation types (*e.g.*, whether the two users are family, friends, or acquaintances) [25]. Link and relation predictions have been used for tasks such as drug repurposing for COVID-19 [168], polypharmacy side effect prediction [19, 176], and user recommendations in social networks [33]. Notably, while other questions may be answered through reasoning over KGs, such as subgraph or graph similarity (*e.g.* representing and comparing chemical structures [42]), a majority of surveyed approaches utilize KGC.

B. Rule-based Methods for KG Completion

Some of the simplest methods toward KGC utilize a set of rules that can be used for logical inference. The source of said rules varies between approaches. As previously stated in §II-B, *ontologies* are formalizations of KB semantics and are often used to represent the unique patterns, relationships, and hierarchies within a specific knowledge domain [39].

Notable examples in biomedicine, for example, include the Gene Ontology [7] and the Disease Ontology [128]. Therefore, ontologies based on description logics can serve as the source of such rules.

Since ontologies often comprise expert-curated knowledge [39], they can be especially beneficial if the application requires domain-specific knowledge [83]. However, ontologies do not exist for every specific domain, and existing ones may impose limitations on the type of predictions made. Alternatively, rules can be mined directly from the KG using ILP methods (discussed in §II-B) such as AMIE and its variants [45]. In these cases, rules are based on association patterns within the KG rather than the general application domain. Notably, Galárraga *et al.* [45] show that association rule mining can correspond to mining Horn clauses in sufficiently large KGs. Therefore, all approaches discussed here do *logical rule mining*.

The major benefit of rule-based approaches is that they are inherently interpretable. One could refer back to the rules which governed the algorithm to get a human-readable understanding of how and why certain predictions were made. For example, SAFRAN [110], which was inspired by AnyBURL [98], provides confidence-weighted, post-hoc explanations based on rules for every prediction. This allows model tuning, adjustment for incorrect predictions, explanations for end-users, and more [110]. Unfortunately, compared to other methodologies, such as those discussed in the next section, rule-based methods do not always achieve the same performance [49, 98]. Although some types of logic, such as probabilistic and fuzzy logic (see §II-B), allow the quantification of uncertainty, programs using them are still limited in their abilities to generalize beyond the patterns encoded in the rules [49]. Additionally, rule-based methods often suffer from scalability issues when faced with large KGs [146, 76], an increasingly pressing issue with the rise of big data and digitalization [72].

C. Graph Embedding-based Methods for KG Completion

In contrast to rule-based methods, those which generate representative *KG embeddings* (KGE) typically scale well to large datasets. A KGE is a numerical vector representation of the KG constituents such that *proximity* in the embedding space approximates some sort of *similarity* in the original KG [167, 87]. Based on the graph notation defined in §II-A, $\text{similarity}(u, v) \approx \text{proximity}(e_u, e_v)$ for nodes u and $v \in V$, in which e_x denotes the embedding for node $x \in V$. From embeddings, one can re-construct parts of the KG to answer some predictive task, such as those mentioned in §III. Many popular KGE methods are based on NNs [14, 126] and show competitive predictive performances in various fields [176, 116]. In particular, GNNs are a popular category of methods to generate KGEs [154]. GNNs account for connections between components; the input is typically a matrix representation of the graph and its constituents, such as node entities, edges and relation types, and node or edge features [154].

The way in which *similarity* is defined and encoded is a major discrepancy between KGE methods [87]. For example,

both DeepWalk [112] and node2vec [52] assess the frequency in which nodes co-occur on random walks through the graph. In contrast, methods like TransE [14] represent relationships as 1-1 translations in the embedding space. Notably, methodologies like RESCAL [105], DistMult [158], and ComplEx [143] were groundbreaking for their stronger focus on relational information. Used alongside a *link prediction* task (§III), for example, these approaches encode similar triples as similar elements in the embedding space. Thereafter, a *decoder* function uses embeddings to compute *scores*. For instance, the DistMult decoder [158, 126], which associates each relation r with a diagonal matrix $R_r \in \mathbb{R}^{d \times d}$, computes the score of each triple with each relation, $r(u, v)$ as:

$$\text{score}_{r(u,v)} = e_u^T R_r e_v$$

Models are trained to assign higher scores to positive triples in the KG and lower scores to negative ones that are not originally present in the data. Following training, the models are then used on some disjoint test set of triples. Negative triples which are scored highly are then regarded as novel positive predictions because they are considered to be similar to the positive samples in the KG [158, 143, 79].

The Graph Convolutional Network (GCN) [78] and its variants (*e.g.*, Relational GCN (R-GCN) [126] and Graph Attention Network (GAT) [145]) have expanded upon this concept with a *message-passing paradigm*. These methods encode each node in the graph as a weighted combination of itself and its surrounding neighbors, thereby incorporating connectivity information. For example, the GCN uses the following equation, where $e_u^{(0)}$ is the node's initial features:

$$e_u^{(l+1)} = \sigma \left(W^{(l)} \sum_{v \in N(u) \cup \{u\}} \frac{e_v^{(l)}}{\sqrt{|N(v)| |N(u)|}} \right)$$

The embedding for each node, u , in the subsequent *layer*, or iteration, $l + 1$, is computed by combining representations in the current layer, l , from itself and all neighboring nodes, $N(u)$. $W^{(l)}$ is a learnable weight matrix, and σ is an activation function [78, 59, 31]. Thus, information is aggregated across edges at each hidden layer. However, this paradigm possesses several limitations as layers increase, including *oversmoothing*, the convergence of node representations to the same value, making them less distinguishable [150, 157, 123], and *oversquashing*, the aggregation of too much information into a single vector, resulting in a less informative representation [9, 50]. Furthermore, GNNs that utilize the message-passing paradigm tend to aggregate information between dissimilar node types (*graph heterophily*) without accounting for the varied behaviors or relationships between them. This typically results in diminished or inconsistent performance [157, 20]. While a number of promising extensions that capture the semantics of heterogeneous graphs have been studied recently [131], there are no strong guarantees that the encoding and aggregation mechanisms will be faithful to the heterogeneous relations in the graph.

Aside from these, KGE methods share other, general limitations. Unlike rule-based approaches, for example, the best performing KGE approaches tend to be black-box models, so no

human-level explanations are generated along with predictions [170]. Additionally, as many KGE methods utilize supervised learning, they require relatively large amounts of labeled data to capture relationships between entities [173, 88]. Several neurosymbolic approaches help to overcome such limitations through characteristics discussed in the next section.

IV. NEUROSYMBOLIC APPROACHES FOR REASONING OVER KGs

As established, we observe a dichotomy between methods for reasoning on graph structures. Symbolic, rule-based methods (§III-B) utilize domain knowledge and logic to infer new information in a naturally interpretable fashion. However, such approaches are limited in their performance and scalability to large KGs. In contrast, state-of-the-art approaches for learning KGEs (§III-C) tend to be black-box methods which aggregate information in domain-agnostic ways. As discussed in §II-D, recent studies in *neurosymbolic AI* often combine aspects from deep learning and symbolic reasoning to mitigate their respective weaknesses and bridge such chasms [144, 47, 65]. Within this section, we survey several such approaches. Since each approach hybridizes the two fields in various ways, we aim to formalize and simplify the language we use by referring to each method's *neural* and *symbolic* modules. A *neural* module refers to the use of a NN, often to produce KGEs, and a *symbolic* module typically involves the use of logical rules. A list of the methods and representations each module could comprise, along with specific examples, is given in Table I.

Amongst the neurosymbolic approaches surveyed [4, 23, 92], we note the following critical characteristics which are unique from symbolic or neural approaches alone:

1) **INTERPRETABILITY:** As mentioned, we often see a tradeoff between interpretability, which symbolic AI naturally possesses (in the form of verbal interpretability [142]), and performance, in which black-box, KGE methods seem to dominate. In particular, enforcing interpretability often comes with a drop in predictive performance [41, 100] and might lead to infeasible computational complexity [11, 41]. The surveyed neurosymbolic approaches foster interpretability and, in many cases, do not sacrifice performance heavily.

2) **GUIDED TRAINING:** Some neurosymbolic approaches can integrate ontological or expert-defined knowledge into an otherwise data-driven approach [4, 153], bypassing the need for the model to learn known patterns. Furthermore, this guides learning toward more domain-congruous patterns. In some contexts, this is ideal for working with limited or small datasets. As described in §III-C, KGE methods tend to require a lot of labeled training data. Often, when a dataset is small, researchers devise relevant pre-training tasks on similar, larger datasets to steer the model's parameters toward a pertinent context [89, 161]. For example, one study pre-trained a GNN to predict atom-level perturbations in molecules to improve the model's overall ability to predict other molecular properties [77]. However, pre-training is an additional and often computationally expensive training step. Neurosymbolic approaches may be a practical alternative.

| | Neural modules | Symbolic modules |
|-------------------------|---|---|
| Methods: | KGE methods (e.g., ComplEx [143], R-GCN [126]) | ILP / Logical rule mining (e.g., AMIE [45], SAFRAN [110]) |
| | Other NN architectures (e.g., RNN [133]) | Knowledge representation reasoners (e.g., OWL reasoner [97]) |
| Representations: | KGEs | Ontology (e.g., GO [7], DO [128]) |
| | Logic embeddings | Logic program (e.g., Horn clauses [103], MLN [117]) |

TABLE I: Methodologies that could serve as neural and symbolic modules in a neurosymbolic approach for KGs. *Methods* are the ways in which the *Representations* may be obtained. Novel predictions can be made using one or more *Representations*. Note that logic embeddings could be considered neural or symbolic.

3) UNDERREPRESENTED TYPES: Even if a labeled dataset is sufficiently large, sparsity or imbalance amongst labels may result in fewer instances of a certain type or subclass within that data. Many KGE approaches struggle to capture such underrepresented patterns [173, 88]. Through KG augmentation, some neurosymbolic approaches pose novel ways to address this.

4) HETEROGENEOUS AGGREGATION: Heterogeneous KGs typically comprise nodes and edges of varying types, and these types may interact in different ways [131]. As stated in §III-C, KGE methods using message-passing may perform inconsistently due to such heterophily. For example, biological KGs like Hetionet [63] comprise drugs, proteins, and diseases as nodes. While drug-protein edges may describe direct, physical interactions (e.g., *binds*), drug-disease edges would constitute a more conceptual relationship (e.g., *palliates*) [63]. Moreover, the features for such nodes would range from chemical structures to lettered sequences [80]. The aggregation of features between dissimilar node types is a challenge which could be targeted via neurosymbolic methods: rather than adding complexity to KGE approaches to learn the relationships between node types, they can be represented through rules.

5) LONG-RANGE DEPENDENCIES: Due to oversquashing and oversmoothing (see §III-C), several GNN methods suffer from local receptive fields [111], experiencing peak performances at two layers [107]. Consequently, they struggle to capture *long-range dependencies* (i.e., relationships between nodes that are several *hops* apart [92, 9, 50]). Rule-based methods, though, can encode such relationships through a series of conjunctions in which each edge in the path is a binary predicate (e.g., $\text{friends}(A, D) \leftarrow \text{likes}(A, B) \wedge \text{likes}(B, C) \wedge \text{likes}(C, D)$), but inference scales poorly to large KGs. Specifically, §IV-C3 describes a series of hybrid approaches which account for long-range dependencies.

Within the remaining sections, we discuss various approaches along with their strengths and weaknesses, referring specifically to these enumerated characteristics as guides for discussion. Importantly, we group the approaches according to the taxonomy in Figure 1 to facilitate easy comparison. It is independent of Kautz’s classification, introduced in §II-D, as we aim for a finer-grained division, specialized to KGs. However, there is naturally some overlap, so we indicate which types, according to Kautz, are present in each of our classes. Note that Kautz’s types 1 and 2 do not appear in our taxonomy, as none of the surveyed approaches align with them. Additionally, we summarize the main points of this survey

within Table II and provide a curated a repository of available code for each work on GitHub².

A. Logically-Informed Embedding Approaches

We begin by introducing some of the most intuitive neurosymbolic approaches for reasoning on graph structures. This section is depicted by Fig. 1-A, and the recently discussed characteristics on GUIDED TRAINING and UNDERREPRESENTED TYPES are most prominent. To combine the benefits of symbolic and neural modules, these approaches modularize the two and then feed the results from the former into the latter. Since symbolic approaches are often based on expert-defined rules, they can be viewed as methods to *extend* the ground truth. Therefore, inference by the symbolic module is used as a preliminary, *KG augmentation* step, which feeds into the neural module (typically a KGE method) for further processing and prediction. In some situations, this KG augmentation step could be exploited to expand small datasets or ameliorate dataset imbalance, thereby harnessing the UNDERREPRESENTED TYPES characteristic. Because the symbolic knowledge is integrated into the training set, these approaches are much like Kautz’s NEURO:SYMBOLIC → NEURO category. We divide these approaches into two subcategories:

1) **Modular, Two-Step Approaches:** This is depicted by the leftmost subcategory of Fig. 1-A and demonstrates the first example of the UNDERREPRESENTED TYPES characteristic. These approaches involve a simple, unidirectional flow of information from the symbolic module to the neural module. For example, Alshahrani *et al.* [4] developed a biological KG from various sources and employed an ontological reasoner to augment and control the scope of the graph. With the reasoner, a fully deduced graph is thus obtained with newly inferred edges. As illustrated in Fig. 2, embeddings are subsequently generated from this augmented KG using the DeepWalk algorithm [112] (see §III-C). Finally, link prediction is performed using embeddings as input. They tested their pipeline, which they coined **Walking RDF and OWL**, on a drug repurposing task to determine whether two drugs share indications, represented by edges between drug and disease nodes. They found that prediction performance improved with the augmented KG. Agibetov and Samwald [1] aimed to improve **Walking RDF and OWL** with an alternative, log-linear embedding method in the neural module, showing that different combinations of symbolic and neural modules

²<https://github.com/NeSymGraphs>

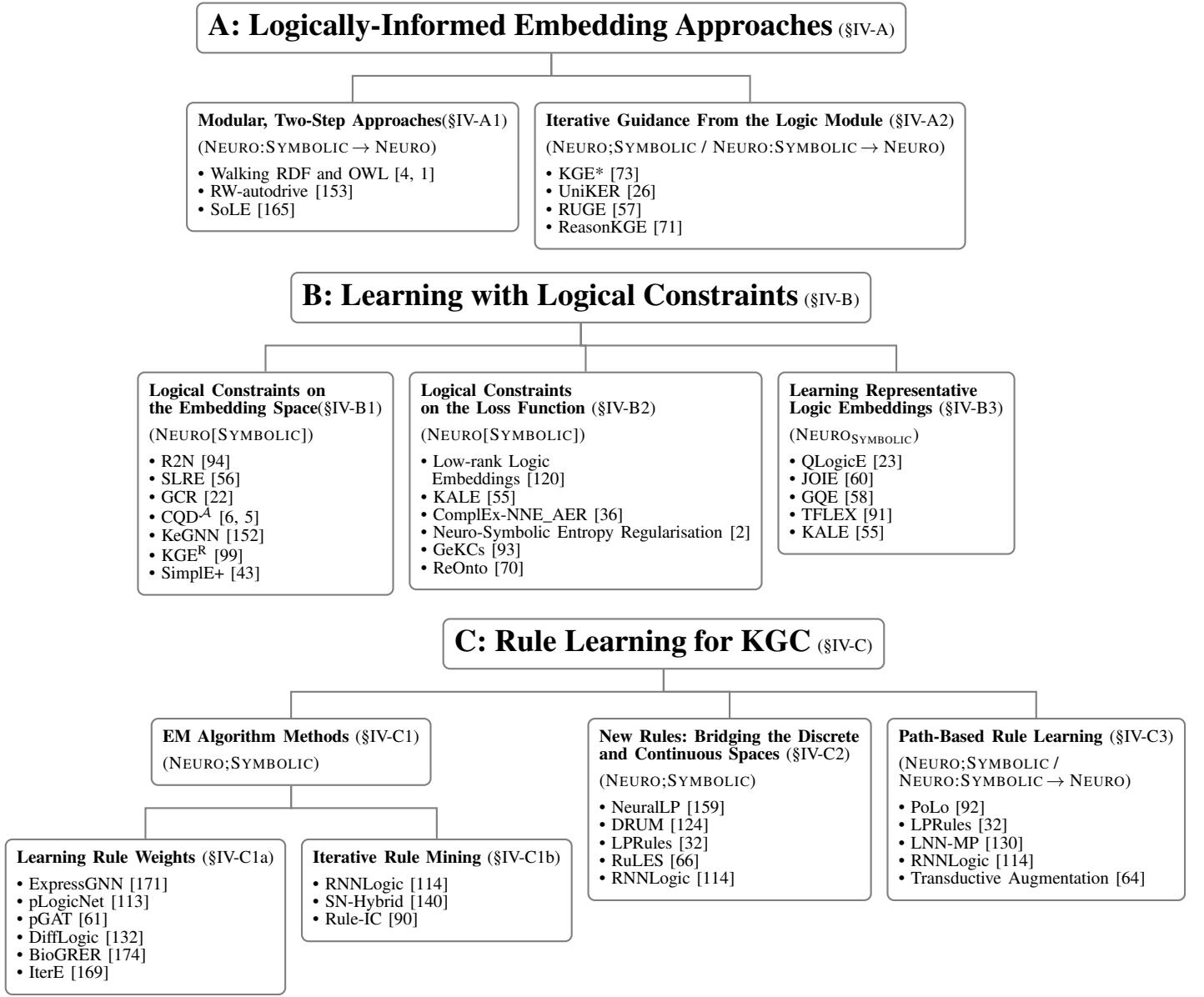


Fig. 1: Taxonomy of neurosymbolic approaches for graph reasoning.

may affect predictions. Consequently, one may wonder how consistently the KG augmentation step improves performance.

Another study by Wickramarachchi *et al.* [153], which we will refer to as **RW-autodrive**, explored this, using varying neural modules in an autonomous driving application. For this study's scene understanding task, data was derived from sensors, such as LIDAR and RADAR, as well as video data taken from the car position. They compared the performances of three different KGE algorithms between both raw KGs and KGs augmented by an ontological reasoner. They showed that, regardless of the KGE algorithm used, better performance was achieved on the augmented KGs, suggesting that incorporation of domain knowledge into KGE methods is especially useful within the autonomous driving domain [153].

While incorporating domain-specific knowledge often improves model performance, the obvious limitation is the re-

quirement for hard rules in the form of an ontology or expert-defined rule set. Thus, there is no opportunity to represent uncertainty or consider new rules. In contrast, **Soft Logical Rules Enhanced Embedding (SoLE)** [165] utilizes soft rules, as discussed in §II-B, alongside corresponding confidences learnt from KGs. Essentially, SoLE's symbolic module acts as a *rule engine* by mining soft rules from the KG. In the KG augmentation step, new groundings, or triples, are inferred by those soft rules and subsequently combined with existent KG triples. However, SoLE does not scale well to larger KGs due to the rule-mining process; an additional rule-pruning step would be useful, a tactic discussed in §III-B.

Overall, two-step approaches are clear and straightforward modular structures to compute KGEs, with room for flexibility on both the symbolic and neural sides. However, their usefulness depends upon the availability of relevant rule sets and face

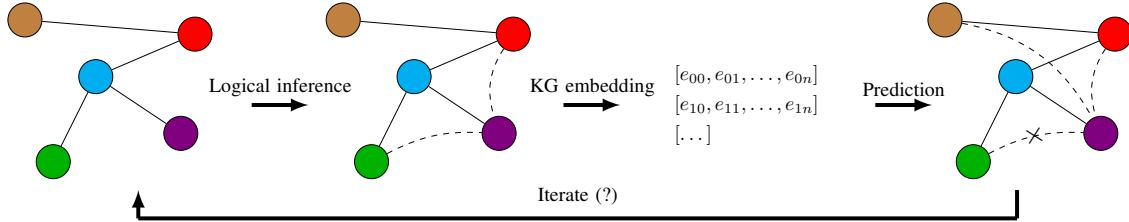


Fig. 2: **Logically-Informed Embedding Approaches** (§IV-A1) These approaches augment the KG with logical inference, then use a KGE method on the augmented KG. As shown by the lowermost arrow, some approaches iterate between logical inference and KGE predictions, using the latter as ground truth for the next iteration’s input (§IV-A2).

scalability issues when rule-mining is involved. Furthermore, these approaches do not take full advantage of the capabilities of either the symbolic or neural modules. In particular, as noted in §IV, many neurosymbolic methods are specifically designed to promote interpretability. However, these two-step methods have the same black-box limitation in their neural modules as any other KGE or GNN approach might since the logic module is only used for KG augmentation.

2) **Iterative Guidance from the Logic Module:** A few approaches extend the two-step pattern to increase interaction between the modules. We classify these approaches into a second subcategory, depicted in the rightmost subsection of Fig. 1-A and characterized by GUIDED TRAINING. In the first subcategory, the directionality of information is one-way: logical inference from the symbolic module informs the KGE method, but the reverse is not true; the output of the embedding method is not used to alter the logical inference step. The minimal interaction between the symbolic and neural modules could, therefore, limit the practicality of these methods. In practice, utilizing the output of logical inference as ground truth labels for the neural module could be fallacious as it assumes that logical inference is monotonic, *i.e.* it always yield true results. In reality, expert-defined rules often have exceptions [47, 44], so regulating the symbolic module could be useful in some cases.

In light of the above, Kaoudi *et al.*’s **KGE*** [73], Cheng *et al.*’s **Unified Framework for Knowledge Graph Inference (UniKER)** [26], and Guo *et al.*’s **Rule-Guided Embedding (RUGE)** [57] all follow similar patterns: the two modules inform one another in an iterative style, as shown in Fig. 2. Both **KGE*** and **UniKER** use Horn rules and forward chaining, a form of logical inference, to augment the KG as in the previous section. Alternatively to Horn rules, **KGE*** also supports using an ontology for the inference step, while **RUGE** uses soft rules. After KG augmentation in the respective symbolic module, a KGE method is trained, and the resulting predictions are used to refine the KG by eliminating the least probable edges and adding the most probable ones. Thereafter, the newly refined KG is passed back to the beginning of the pipeline for the next iteration. One shared shortcoming of the these approaches, however, is that they are designed mainly to generate positive predictions; this can lead to an increased number of false positive predictions. As a solution, another iterative approach, called **ReasonKGE** [71], adds a step in which negative samples are also updated.

By using an iterative process, these methods introduce bidirectionality in which both symbolic and neural modules inform one another by taking turns to refine the KG. One might argue, then, that these approaches fit most closely with Kautz’s third category of NEURO;SYMBOLIC AI, rather than the fourth. However, there is a key caveat: the parameters of the symbolic module are *not* updated as in the neural module; the rules of the symbolic module are static. Therefore, we still view these approaches as primarily neural. Methods which implement dynamic symbolic modules are discussed later within §IV-C1. Ultimately, we consider the static symbolic modules in these approaches as simply *guiding* model training rather than *driving* it. Nevertheless, guidance via the symbolic module allows these approaches to fulfill the GUIDED TRAINING characteristic. For example, if a user chooses to implement domain-specific rules, such as that from an ontology, into the symbolic module, then groundings from such rules will be included as input into the neural module. Therefore, the neural module will learn to encode patterns consistent with such rules. These approaches move conceptually closer, then, to those which apply logical constraints onto neural modules, the topic of the next section.

B. Learning with Logical Constraints

In contrast to using two separate modules, a different neurosymbolic pattern on KGs involves imposing the symbolic module, in the form of logical or rule-based constraints, onto the neural module. This pattern corresponds to Fig. 1-B. In this case, the focus is on training the neural module, but rules are used both to incorporate domain-specific knowledge and to limit the scope of predictions possible. These approaches are, therefore, the epitome of the GUIDED TRAINING characteristic, as their primary goals include guiding neural training via the symbolic module. For reasons discussed toward the end of this section, these approaches tend to be the weakest with regard to INTERPRETABILITY. However, some are particularly useful for integrating heterogeneous information, *i.e.* the HETEROGENEOUS AGGREGATION characteristic. This section fits most closely with Kautz’s sixth category of NEURO[SYMBOLIC] AI because the constraints often act as logical checkpoints within the neural engines.

These approaches fundamentally differ based on where logical constraints are imposed in the learning pipeline. Some approaches, discussed in §IV-B1 below, impose them directly onto the learned embeddings, while other approaches, dis-

cussed in §IV-B2, influence model training by restricting the predictions made from embeddings. §IV-B3, for its part, discusses methods that learn entirely separate embeddings for logical constraints.

1) Logical Constraints on the Embedding Space: One way to approach imposing logical constraints onto KGE methods is to alter the embedding space in some meaningful way, as depicted in Fig. 3 and the leftmost branch of the taxonomy in Fig. 1-B. Such approaches are also particularly useful for the GUIDED TRAINING and HETEROGENEOUS AGGREGATION characteristics. The **Relational Reasoning Network (R2N)** [94], for example, uses structural relational information as constraints on GNN training, sharing quite similar goals to the previously mentioned R-GCN [126] (§III-C). Using the phrase “relational reasoning in the latent space” to describe their process, they transform the latent space embeddings in a way that is dependent upon each node’s neighbors in the original space. They argue that the transformation produces KGEs encoding information about both the individual nodes and the surrounding graph topology. In a similar manner to **R2N**, **Soft Logical Regularity in Embeddings (SLRE)** [56] uses soft logical rules as constraints on *relation* embeddings. These rules are fused into the KGE process by imposing rule-based regularization onto generated embeddings. Consequently, embeddings are potentially more useful for generating predictions aligned with domain-specific or logic-based paradigms.

The emphasis on using relational information seen in **R2N** and **SLRE** is also key to a study by H. Chen *et al.* [22], who propose **Graph Collaborative Reasoning (GCR)**. The **GCR** extends the vanilla GNN to encode whole triples, rather than nodes alone, focusing their analysis on adjacent edges. Conceptually, they transform the KG structure into a series of logical expressions which are used to predict the probability that a novel edge is implied by its adjacent edges. This re-frames the link prediction problem as a neural logic reasoning one. More specifically, they accomplish this by applying logical regularizers to MLPs to simulate logical operations; these MLPs are then applied to triple embeddings to generate predictions. Furthermore, to avoid the requirement of specifying rules by hand, they introduce a method to learn potential rules, formulated as Horn clauses, from the available training triples; this benefit foreshadows an approach presented in §IV-C. Additionally, the authors claim that their method is more scalable to large, complex KGs than other neurosymbolic

methods. However, we note that the scope of applications on which their method might be applicable could be limited. The logical expressions by which they perform link prediction rely on the idea that closely connected neighbors likely share similar interaction patterns. While this is often true for social KGs, it may not suit other domains in which entities with similar patterns might avoid interacting closely.

Two additional studies, **Knowledge Enhanced Graph Neural Networks (KeGNN)** [152] and **Continuous Query Decomposition (CQD)** [6] also view KGs in terms of logical expressions. Both demonstrate that, after using KGE methods for simple link prediction, *t*-norms (§II-B) and *t*-conorms (for disjunctions) can be applied to embeddings to answer more complex queries comprising *multiple* links. In theory, this approach could be useful for LONG-RANGE DEPENDENCIES if the queries involve a series of conjunctions; this idea is elaborated upon in §IV-C3. In particular, **CQD** demonstrates its usefulness for queries of up to eight links. Its extension, **CQD^A** [5], imposes learnable adaptation functions to accommodate for the interactions between parts of a complex query. Such adaptation functions could be a way to cope with graph heterophily, mentioned in §III-C.

Minervini *et al.* also explore the use of logical constraints as regularizers with an approach we will refer to as **KGE^R** [99]. Specifically, they apply constraints to encode information about relations that may mean the same thing, such as *PartOf* and *ComponentOf*, through a model-dependent transformation on the embeddings. They derive these transformations for three different KGE methods and discover that they all yield more accurate link predictions with regularization. As with **SLRE**, applying constraints directly to the embeddings was found to be effective in integrating domain knowledge without impairing scalability [99]. However, unlike **SLRE**, **KGE^R** is not model agnostic, requiring the user to derive a useful transformation for each individual KGE method. We note that this also poses the risk that the regularization method might not generalize well to any given embedding method.

Instead of model-dependent regularization, both **SimplE+** [43] and **SLRE** enforce constraints on embeddings by requiring non-negativity. The authors of **SimplE+** explain that their motivation in using non-negativity, specifically, is to enforce the *subsumption* axiom from OWL’s semantics [97]. Essentially, subsumption determines whether something is a subclass or subproperty of another. By enforcing that generated

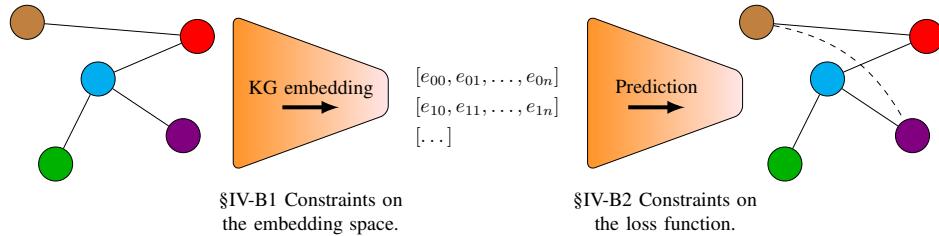


Fig. 3: Learning with Logical Constraints. (§IV-B1) Logical constraints on the embedding space (first filter). These methods drive training by imposing logical constraints onto the embedding space, such as through a transformation. (§IV-B2) Alternatively, logical constraints on the loss function (second filter). These methods drive which predictions are made by encoding logical constraints into the loss function, such as with a penalty term.

entity embeddings must be non-negative, they prove that subsumption is enforced. Furthermore, they also impose a constraint on relations which can be subsumed from one another. Their method outperformed simple logical inference as well as the basic embedding method without non-negativity constraints [43].

Imposing logical constraints onto the embedding space is ideal when one wishes to encode specific information about the relational nature of the graph. In particular, incorporating relational or ontological relationships into KGEs could be considered as a solution to the challenges posed by heterogeneous KGs, as described in §IV. These neurosymbolic approaches take into account the unique and varied relationships between different node types when generating predictions, thus harnessing the HETEROGENEOUS AGGREGATION characteristic. However, depending upon the needs of the researcher, such constraints may be too restrictive. Instead of applying constraints onto the embedding space, thereby limiting the types of patterns encoded, another subcategory of approaches, discussed in the next section, applies constraints to the processes of decoding the embeddings and generating predictions.

2) Logical Constraints on the Loss Function: A drawback of popular KGE algorithms (namely RESCAL, TransE, HolE, and ComplEx, as discussed in §IV-A) is that they only implicitly learn higher-order relationships among triples in the KG, requiring more data to learn its latent logical structure. Moreover, this learned structure is confined to the ground triples available during training. These limitations can be overcome by the next subcategory of approaches, depicted by the centre branch of Fig. 1-B. These use logical constraints to guide training toward predictions which align with some knowledge base or logical entailment (Fig. 3), aligning with the GUIDED TRAINING characteristic. Here, this is accomplished by altering the loss function.

For example, **ReOnto** [70] integrates ontological axioms as a bias term into the loss function of a GNN, coercing the training process to adhere to biomedical knowledge. Despite reporting competitive performance, **ReOnto** requires a domain-specific ontology. Alternatively, an approach called **Low-rank Logic Embeddings (LRLE)** [120] integrates logical constraints as differentiable loss functions to be jointly optimized. Given logical rules of the form

$$\text{relation_a}(X, Y) \rightarrow \text{relation_b}(X, Y),$$

two distinct processes are carried out. First, the rules are applied to the existing training triples to generate new ones, as described in §IV-A. Additionally, this encourages the learned embeddings to encode inter-relation structures. Second, a unified set of differentiable loss functions corresponding to both the set of training triples and the logical rules is created using fuzzy t -norms (see §II-B). This creates a joint optimization problem over capturing factual information while obeying the given constraints. A summation of the individual losses is formulated as a log-likelihood loss so that embeddings which assign a high marginal probability to rule and triple satisfaction are preferred. The authors argue that the framing of constraint

satisfaction as a probabilistic model allows the embeddings to be unaffected by noisy data.

However, **LRLE** is also limited in that it only models existent relations and entity-pairs. Consequently, rules will not be discovered for entity-pairs that do not appear in the training data. This shortcoming is significant as most KGs are inherently incomplete. This drawback is tackled by Guo *et al.*'s method, **Embeddings by jointly modeling Knowledge And Logic (KALE)** [55], which performs the same unification of triples and rules, but also explicitly models entities on their own. This allows for novel relations between entities to be predicted, as demonstrated by a recent study [135] which employed **KALE** as an effective recommender system. **KALE** extends TransE and maintains the same competitive time and space complexity thereof. Moreover, while their approach uses a pairwise ranking loss, the authors note that the generality of their approach facilitates the use of different losses such as with **LRLE**.

Unfortunately, the last two approaches have the additional computational burden of grounding universally quantified rules before training. This requirement hinders their ability to scale to large KGs with many rules, a weakness reminiscent of the methods within §IV-A. In contrast, Ding *et al.* [36] build upon the ComplEx algorithm, creating **ComplEx-NNE_AER**, to include non-negativity constraints on entity representations and approximate entailment constraints on relation representations. These constraints are encoded through a penalty term applied to the objective function. The authors argue that these constraints are sufficient to impose a compact, informative prior structure of the KG on the embedding space, without negatively affecting efficiency or scalability. Most importantly, their approach does not require rule grounding like the previously mentioned ones. Unfortunately, however, constraints are incorporated into the loss in a way that does not support general first-order logic rules like that of **LRLE** or **KALE**. We thus surmise that their approach might perform comparatively worse on more complex KG structures.

More generally, Ahmed *et al.* [2] encode constraints as a loss function that can be applied to any NN model, including GNNs. In an approach termed **Neuro-Symbolic Entropy Regularisation**, the authors use *semantic loss* [156] as a measure of how much a first-order logic formula is satisfied by the output of a model. In contrast with methods which use fuzzy logic [120], the semantic loss is a probabilistic definition measuring the likelihood that the output will satisfy the constraint, given the induced probability distribution over the trained NN's outputs. The semantic loss is then combined with entropy regularization. This encourages NN predictions to conform to a structure which satisfies the specified constraints while also ensuring more distinct decision boundaries in link prediction. Similarly, **Generative KGE Circuits (GeKCs)** [93] take advantage of probabilistic circuits [147] to ensure that hard constraints are met [3]; in other words, unlike semantic loss, which encourages conformity to constraints, GeKCs guarantee that predictions will satisfy constraints.

Within this and the previous subsection, the pipeline in which KGEs are trained is linear (see Fig. 3), with the logical constraints being injected into the pipeline at some point.

Alternatively, however, logical constraints could be treated as an independent source of information to be encoded in parallel to the KG. Within the next subsection, we discuss studies that explore that possibility.

3) Learning Representative Logic Embeddings: Within the last subcategory of learning with logical constraints, represented by the rightmost branch of Fig. 1-B, some approaches work by learning representative embeddings for logical constraints. Then, they systematically combine them with the KGEs before generating predictions. This idea is illustrated in Fig. 4. Similar to the previous subcategories, it helps to drive training toward established and possibly domain-specific knowledge, harnessing the GUIDED TRAINING characteristic. For example, the **Quantum Logic Empowered Embedding (QLogicE)** method [23] uses quantum logic [12] to create logic embeddings while simultaneously computing KGEs. The scoring function for each possible triple is a weighted sum of the scores for the two embedding methods, so the overall loss is also computed as a weighted composition of the two respective losses. The weights of the scoring and loss functions determine what proportion of each embedding method to take into account. Such a tactic is used within the method, **Joint Embedding of Instances and Ontological Concepts (JOIE)** [60]. Here, two separate embedding spaces are generated for the instantiated triples and the underlying ontological structure, so a joint loss is calculated. In **JOIE**, however, they found that performance is improved even further by including a third set of embeddings on a combined graph structure comprising both instantiated triples and ontological relations. Despite improved performance, one might wonder whether **JOIE** moves away from achieving an interpretable model.

In contrast, a study by Hamilton *et al.* [58] combines logic and node embedding spaces in a way that the logic embeddings represent a query. Their method, **Graph Query Embeddings (GQE)**, uses geometric operations which represent logical operators to compute *query embeddings*. Thereafter, the likelihood that a set of nodes satisfy a query is calculated as the cosine similarity between the query embeddings and the respective node embeddings. In a highly similar manner, the **Temporal Feature-Logic Embedding framework (TFLEX)** [91] additionally accounts for temporal dependencies by embedding timestamps. These methods essentially perform complex logical queries on a graph, making them similar to Kautz's second category of **SYMBOLIC[NEURO]** AI.

Similarly to the above approaches, **KALE** [55], mentioned

previously, learns embeddings based on both the triples and logical rules. Essentially, they train a KGE algorithm while computing truth values for logical rules based on the presence of instantiated triples. Training is based on loss encompassing both triples and logical formulae. While **KALE** fits most closely into this category of logical constraints, one could also use the truth values of the logical rules as confidences. In §IV-C, we discuss the possibility of assigning confidence values to rules in order to quantify relative importance. Understanding which rules are most important in generating predictions can be a valuable source of interpretability. However, the authors of **KALE** do not explore this possibility.

In general, since all of the approaches described within this category use the symbolic module to guide training of the neural module, they have the potential to replace pre-training in some circumstances, as suggested in the GUIDED TRAINING characteristic. Specifically, some studies use pre-training to steer a neural module's parameters toward a relevant context; this is typically due to a lack of sufficient data pertaining to the task at hand [89, 161]. In this case, the symbolic module accomplishes a similar function, guiding the neural module's parameters *during* training, as opposed to *before* training.

However, these approaches are weakest with regard to INTERPRETABILITY as none of them necessarily provide insight as to why certain predictions were made over others. In fact, one could argue that the approaches which learn logic embeddings *remove* the inherent interpretability that is typically so fundamental to using logic. To fully leverage this property, the symbolic and neural components of an architecture should be combined in a way that they not only inform one another but also in a way that user-level explanations are generated. In the next section, we review a variety of approaches which aim to modify or learn new logical rules to do just that.

C. Rule Learning for KGC

In the previous sections, all of the described methodologies used predefined rules for KGC. However, quite like the MLN, approaches in this final category attempt to *learn* such rules. Such approaches are represented in Fig. 1-C, and, collectively, they possess *all* of the five characteristics listed in §IV. Overall, most of these methods train some neural module to learn confidences for rules or systematically adjust a rule mining scheme which constitutes the symbolic module. The presence of dynamic symbolic modules makes these

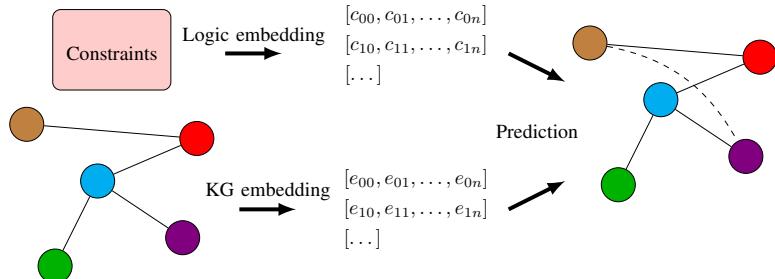


Fig. 4: **Learning Representative Logic Embeddings.** Some methods encode logical constraints as embeddings, then combine them with KGEs to make more informed predictions.

approaches unique from all previously described categories. In particular, it makes this category of approaches strong with regard to the INTERPRETABILITY characteristic. As we will see next, the majority of these approaches fall into one or more subcategories based on similarities in their general architectures as well as the ways in which rules are learned.

a) EM Algorithm Methods: Many rule-learning methods claim to take an Expectation-Maximization algorithm (EM) based approach, alternating between two modules whose outcomes inform each other. These approaches are listed in the leftmost branch of Fig. 1-C and distinguished by the INTERPRETABILITY, GUIDED TRAINING, and UNDERREPRESENTED TYPES characteristics. Put succinctly, the EM algorithm is used to estimate the maximum likelihood of model parameters for situations in which there is incomplete data [35]. KGC, then, is an ideal and obvious setting for such methods. With varying degrees of faithfulness to the original EM algorithm proposed by Dempster *et al.* [35], the approaches in this section all, generally, claim to adopt the iterative nature of the EM algorithm. More specifically, each approach alternates between the prediction of missing components (the “E-step”), typically in the form of KGC, and the optimization of model parameters to account for such predictions (the “M-step”). Normally, the parameters to be optimized belong to the symbolic module, which can be understood in this section as a dynamic framework involving logical rules (*e.g.*, a MLN or rule mining system). These methods vary, however, in their interpretations of how an EM-based algorithm might be executed for KGC.

a) EM Methods – Learning Rule Weights: The most straightforward approach for creating a dynamic symbolic module involves assigning *weights*, *scores*, or *confidences* to logical rules, which denote relative importance, then imposing incremental updates. Similar to approaches described in §IV-A2, several EM methods use two complementary modules to inform one another, encapsulating the GUIDED TRAINING characteristic, but the symbolic module is now dynamic, adjusting the rule base at each iteration (Fig. 5). For instance, ExpressGNN [171], pLogicNet [113], pGAT [61], and DiffLogic [132] all incorporate a MLN into the symbolic module of their algorithms to learn corresponding weights,

or confidences, for logical rules. During the E-step, a KG augmented by logical inference is utilized to train a KGE method. Recall that the primary goal of the KGE method is to distill patterns from the input graph into representative embeddings, from which the KG can be reconstructed along with its missing, or latent, components. Thereafter, during the M-step, the newly predicted triples from the KGE method are used to inform and update the weights of the MLN.

The above methods have only minor differences: for example, while ExpressGNN uses a vanilla GNN as the embedding method [171], pGAT works with a variant of the GAT [145] (see §III-C), utilizing the attention coefficients as an additional source of semantic information [61]. Operating similarly to pLogicNet and pGAT, Zhao *et al.*’s method, **Biomedical KG refinement with Embedding and Rules (BioGRER)** [174], is additionally specialized to operate on a biological KGs. To do so, it incorporates domain-specific knowledge into logical rules. However, unlike the aforementioned approaches, while the rule-learning module is highly similar to a MLN, it is not explicitly based on a MLN.

In contrast, Zhang *et al.*’s **IterE** [169] does not use a MLN. IterE initializes its symbolic module as a pool of randomly selected OWL-based axioms, followed by grounding. Axioms with more than one instance are selected for the final rule set. As in the previous methods, new triples inferred by the symbolic module are used as ground truth for training the neural module. Likewise, scores in the symbolic module are updated based on each iteration’s KGEs. Unlike previous approaches, however, the KGEs are used directly for score updates, skipping the prediction step in between.

b) EM Methods – Iterative Rule Mining: While pLogicNet, pGAT, IterE, and BioGRER learn and update rule *weights*, the rule set itself never changes. In contrast, Suresh and Neville’s **Hybrid Method (SN-Hybrid)** [140], Lin *et al.*’s **Rule-enhanced Iterative Complementation (Rule-IC)** [90], and Qu *et al.*’s **RNNLogic** [114] mine fresh rules over iterations, rather than using the same rules and adjusting confidences. In this sense, these methods are better aligned with the goals of ILP (§II-B). In other words, KGE-derived predictions from the neural modules are used to reduce the rule mining search space and alter the pool of candidate

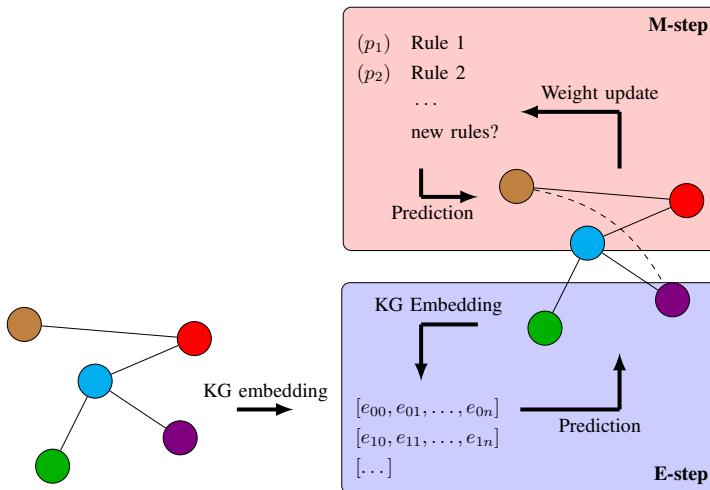


Fig. 5: **EM Algorithm Methods.** These approaches adopt the idea of iterative optimization from the EM algorithm [35] to describe systems which iterate between symbolic and neural modules. Typically, the E-step involves KGC through a KGE method (the neural module), and the M-step involves updating the parameters of the symbolic module. To make the symbolic module (typically in the form of a MLN or rule mining approach) dynamic, some methods update rule confidences (§IV-C1a). Other methods update and alter a pool of candidate rules (§IV-C1b). Note that approaches in this category may deviate from this portrayal.

rules, which are subsequently used for logical inference in the symbolic modules. To illustrate this, we refer the reader once again to Fig. 5 with the notion that the rule set is iteratively altered (hence the “*new rules?*” option within the rule set). Notably, **Rule-IC**’s rule set is ontology-based, leading to richer KG semantics, while **SN-Hybrid** and **RNNLogic** mine general Horn rules.

Another key difference between the rule mining processes used within these approaches is the ways in which they determine the pool of candidate rules or axioms. **Rule-IC**, for example, constructs its rule set similarly to the previous section’s **IterE**, but it uses a computed confidence threshold rather than a frequency threshold [90]. In **SN-Hybrid**, however, a pool of candidate rules is generated using an algorithm akin to AMIE [45], in which partial rules are queued and then extended until they become proper rules. The candidate rules are later pruned via a combination of standard confidence measures, a measure of the ability to explain the existing triples, and the confidences determined through the neural module [140].

This category’s essential quality of back-and-forth guidance between symbolic and neural modules makes it most similar to Kautz’s third category of NEURO;SYMBOLIC AI. Consequently, one major perk of EM-based algorithms is the way they account for *underrepresented* relation types. Often, if a relation type occurs less frequently, KGE methods neglect to account for it, as there are fewer examples upon which to train, and the dataset imbalance creates bias toward more frequent relation types. Various KGE methods attempt to account for this via adjusted sampling or regularization methods [126]. Iterative methods could be used to increase the occurrence of underrepresented relation types during training. Specifically, one might consider the inference step in the symbolic modules as a way to increase the number of positive edges being fed into the embedding module. For example, if the method utilizes rule confidences, such as in §IV-C1a, one could initialize or fix rules regarding rare relation types with high confidences so that the symbolic module predicts more high-probability instances of that type. Therefore, this subcategory of approaches handles small or imbalanced datasets well, capturing the UNDERREPRESENTED TYPES characteristic. We explore this prospective direction further within §V-B5.

2) New Rules: Bridging the Discrete and Continuous Spaces: In the previous subcategory, we introduced approaches with dynamic symbolic modules. Specifically, these approaches either made iterative confidence updates to pre-defined or mined rules (§IV-C1a) or mined a fresh set of rules based on the output of the neural module (§IV-C1b). However, previous approaches mined and filtered rules based on predefined heuristics, including the number or proportion of groundings [140] or alignment with certain OWL axioms [90]. Such heuristics may not be ideal for various applications, thereby limiting the generalizability of such methods. Alternatively, the neural module can be used directly for the task of generating and selecting descriptive rules. This leads to approaches, depicted in the centre branch of the taxonomy (Fig. 1-C), that aim to learn both structural information in a *discrete* space and respective weight parameters in a *continuous* space

[159]. The creation of a method that can do *both* of these and train in an end-to-end fashion is inherently difficult [159, 124]. However, the payoff is an enhanced INTERPRETABILITY characteristic.

Methods have been developed in response to suggest new ways to perform differentiable logic-based reasoning. TensorLog [28], for example, encodes graph entities into vectors and unique relations into respective adjacency matrices. Logical inference can then be imitated as the product of the adjacency matrices for the relations in the body of a rule with the vector of a given entity. This produces a vector in which nonzero entries represent entities for which the rule holds true. **Neural LP** [159] integrates TensorLog into an approach which additionally learns rule confidences in an end-to-end fashion. Specifically, it learns confidence values not just for every rule but for every relation involved in every rule, even accounting for differences in the varying lengths of rule bodies. **DRUM** [124] is a refined version of **Neural LP** in which the authors point out that **Neural LP** learns high confidences for incorrect rules. To overcome this problem and reduce parameters, **DRUM** incorporates a bidirectional RNN to capture forward and backward information about possible pairings of relations, as not all relations can actually coexist in the same body of a rule. As a consequence, **DRUM** outperforms **Neural LP** and even some black-box embedding-based methods at link prediction within various datasets [124].

In contrast, the authors of **LPRules** [32] take a slightly different approach to learning rules and confidences simultaneously. Their approach generates a weighted combination of logical rules for link prediction by iteratively augmenting a small starter pool of candidate rules. It differs from the aforementioned methods which learn confidences for the entire set of possible rules because it only works with a subset of rules at any given time, thereby reducing the search space and optimizing the time it takes to find a set of rules which are highly predictive [32]. Another method, **RULES** [66], operates very similarly, augmenting its rule set iteratively by constructing and extending rules with additional atoms and logical refinement operators. **RULES** is arguably more sophisticated, though, because the quality of candidate rules is checked against pre-computed KGEs and, optionally, text embeddings. Notably, **LPRules** and **RULES** are comparable to the EM-based algorithms because the output of previously generated rules influences the generation of rules in the next iteration, much like that of **RNNLogic** [114]. In fact, similarly to **RNNLogic**, **LPRules** initializes its small pool of starter candidate rules via path-based heuristics. Accordingly, **RNNLogic** could also be classified within this category as it uses a NN variant for rule generation.

Like the EM-based methods, this category is also most similar to Kautz’s third classification of NEURO;SYMBOLIC AI. However, the major benefit of **Neural LP**, **DRUM**, and **LPRules** is that they train end-to-end. Unlike **IterE** [169] and **SN-Hybrid** [140], which use the neural module to guide rule mining, these approaches are not limited to predefined heuristics. As a major limitation, however, **Neural LP** and **DRUM** only train on positive examples and have yet to be tried on a dataset with negative examples. In many applications,

positive predictions are more interesting than negative predictions. However, a lack of negative examples in the training data increases the risk of false positive predictions, which, in some contexts, such as drug-target prediction, could lead to a waste of time, money, and resources or potentially even health hazards.

3) Path-Based Rule Learning: Most of the previous approaches which mine or generate rules do so in a way that treats the latter as statements involving individual relation types. However, there exists another branch of work, the rightmost part of Fig. 1-C, which accomplishes rule-learning through path-based approaches. Path-based methods can make inferences from chains of edges or relations, as shown in Fig. 6. Consequently, the generated rule sets are not only interpretable but often more expressive, as they facilitate understanding of relationships between nodes that are several hops away from each other, leveraging long-range dependencies. Therefore, in addition to leveraging all the other listed characteristics in §IV, this subcategory of approaches is *uniquely* distinguished by the LONG-RANGE DEPENDENCIES characteristic.

Long-range dependency is particularly important in application domains such as biology, in which a chain of interactions between macromolecules connects entities several steps apart [21, 29]. This is why Liu *et al.* [92] design a path-based neurosymbolic method for predicting novel drug indications (*i.e.*, treatment possibilities for pharmaceutical drugs [37]). Specifically, they investigate paths which start and end with *drug* and *disease* nodes, respectively. By exploring indirect drug indications which might operate through other entities, such as genes, they expand the set of novel discoveries possible. However, using path-based exploration expands the search space. Therefore, to search the KG meaningfully, they employ **Policy-guided Walks with Logical Rules (PoLo)**. Essentially, they start by encoding general path patterns, or *metapaths*, as logical rules with pre-computed confidence scores. While metapaths can be manually selected from expert knowledge, they can also be mined, such as in the **PoLo** follow-up study by Drancé *et al.* [40]. In both studies, they train an agent to walk through a biological KG, storing history using a LSTM. To use the rules as guidance, the agent is not only rewarded when it finds a positive drug indication but also when the path it follows corresponds to one of the logically encoded metapaths. Ultimately, the incorporation of path-based rules in the symbolic module guides the neural module toward exploration of long-range dependencies so that indirect relationships between distant nodes can be

predicted. Therefore, this can be seen as another example of the GUIDED TRAINING characteristic. Furthermore, expert-curated metapaths tend to consider the most probable combinations of relationships between various node types, so these approaches have the potential to handle heterogeneous graphs well, fitting the HETEROGENEOUS AGGREGATION characteristic. Alternatively, if one uses rule mining to generate the metapaths, as Drancé *et al.* did, the generated rules and their corresponding confidences facilitate the INTERPRETABILITY characteristic of the model.

LPRules, previously mentioned in §IV-C2, operates in a similar style to **PoLo**. Specifically, one of the heuristics which **LPRules** proposes to generate and update the pool of candidate rules utilizes long-range dependencies in the KG. For rule generation regarding a specific relation, such as drug indications in the **PoLo** study, it iterates through instances of that relation and finds the shortest path between each respective pair of nodes, excluding the current direct edge. From the shortest path, a rule is generated based on its general sequence of relations (which **PoLo** called a metapath) [32]. From the aforementioned studies, we can see that KGC can depend heavily upon information beyond a node’s local neighborhood. However, one major challenge with path-based approaches is that various sequences of relations are often imbalanced, as some relations and patterns thereof are significantly rarer than others.

Dependent upon the application, less frequent path sequences might be less interesting or relevant to the end-goal. Sen *et al.* [130] aim to address this. In their study, they first utilize a Mixture of Paths (MP), in which the body of each generated logical rule contains one of many possible length- k relation sequences in the KG. To generate such rules, they use the so-called Logical NN (LNN) [118], coining their neurosymbolic approach **LNN-MP**. The previously mentioned **RNNLogic** (see §IV-C2) also generates path-based rules in a similar manner to the MP method. However, Sen *et al.* argue that RNN-based approaches tend to be unnecessarily complex. Because LNNs are based on real-valued boolean logic, **LNN-MP** is computationally simpler than RNN-based approaches, and the rules learned are fully interpretable. Thereafter, to represent the most common path patterns in the KG, **LNN-MP** can be used on KGEs pre-trained with bias toward more prevalent relation sequences. By doing so, **LNN-MP** is capable of learning rules most relevant to the paths present in the KG, improving KGC performance.

Similarly, an approach by Hirose *et al.* [64] weighs path-based rules on relation path frequencies in the KG, focus-

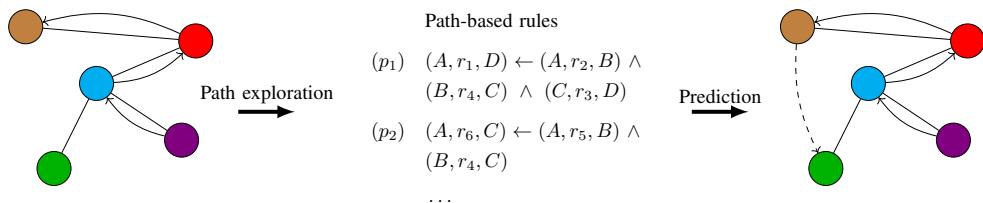


Fig. 6: Path-Based Rule Learning. These methods generate rules based on paths in the KG to encode long-range dependencies. Path exploration is sometimes guided by a NN or KGE method, and confidence scores are often computed for rules.

ing KGC upon the most prevalent relation sequences. Their method, **Transductive Augmentation (TA)**, draws inspiration from **UniKER**, described in §IV-A2. Like the latter, it follows the same general process, illustrated within Fig. 2, of iteratively augmenting a KG on which a KGE method trained. **TA** is made more sophisticated, however, by two major changes. First, relational path rules are mined from the KG using random walks. Then, based on the frequency of relation paths in the KG, confidence scores are computed to permit filtration for the top thousand rules. **TA** improves KGC performance in comparison to KGE methods alone, and it generates weighted, path-based rules which, once again, highlight the path-like patterns that are most relevant to predictions made.

While the benefits of path-based methods have already been abundantly discussed here, we note that these approaches also rely on several assumptions. First, they all operate on heterogeneous KGs containing multiple relation types; single-relation KGs may not benefit from such methods. Second, one might wonder, especially in the biological context of **PoLo**, whether these methods operate equally well on both directed and undirected graphs. Finally, approaches such as **LNN-MP** and **TA** assume that the most relevant path patterns are those which reflect the most prevalent sequences in the KG, but this may not always be the case; such approaches would lack the ability to handle underrepresented classes, therefore falling short of the **UNDERREPRESENTED TYPES** characteristic. On the other hand, Drancé *et al.* point out that more interesting rules might be obtained by considering the goals of the application domain in which one is interested [40], and sometimes, underrepresented relation types are most important or interesting for KGC, a topic explored within §V-B5.

Because the path-based approaches are varied in the way they operate, it is difficult to categorize them into one of Kautz's types. Since these approaches share the most similarity with those in the previous rule-learning sections as well as §IV-A2, we believe that this section could be split between **NEURO:SYMBOLIC → NEURO** and **NEURO;SYMBOLIC** types. Because we focus on a specific area of neurosymbolic AI, we are untroubled by this imperfect division. As this area is still developing, we expect that future approaches will continue to refine and update both our classification and Kautz's.

Within the last decade and particularly within the last six years, one can see that neurosymbolic approaches for reasoning on KGs have gained significant interest. We found that the surveyed approaches fit best within three major categories: (1) logically-informed embedding approaches (§IV-A), which are most similar to Kautz's **NEURO:SYMBOLIC → NEURO** category, (2) embedding approaches with logical constraints (§IV-B), most like **NEURO[SYMBOLIC]** AI, and (3) rule-learning approaches (§IV-C), which fit most closely to **NEURO;SYMBOLIC** AI. Furthermore, we anticipate that this area of work will expand significantly within the next few years to fill out the parts of Kautz's neurosymbolic classifications which we did not frequently mention, such as **SYMBOLIC[NEURO]** AI. Additionally, we expect that these approaches, along with their unique capabilities, will fill in several research gaps. In an attempt to facilitate and inspire

future studies, we discuss prospective uses in the next section.

V. LIMITATIONS & PROSPECTIVE DIRECTIONS

As neurosymbolic reasoning over KGs is still a young field of research, there are plenty of technical and practical areas yet to be fully explored. We next mention several common limitations of these approaches, then suggest a number of prospective directions, which we hope will cultivate a greater interest in this domain. In particular, we note that many of these directions could be useful for biomedical data and applications, so we use a number of examples in this domain to illustrate our points.

A. Limitations

While respective benefits and weaknesses were discussed for each technical category, there are also several *general* limitations of these studies as well as neurosymbolic approaches as a whole.

1) Increased Complexity: While, in theory, neurosymbolic methods could decrease computational complexity by guiding neural training, they may also *increase* it, depending upon their implementation. This is because existing symbolic methods which rely upon grounding rules, like the traditional implementation of the MLN, do not tend to scale well to large datasets [139]. Additionally, as explained in §IV-C2, it is difficult to integrate the symbolic module with the neural module in an end-to-end fashion. Therefore, if a neurosymbolic pipeline fails to integrate the symbolic module in a way that improves scalability, then combining it with a neural module increases model complexity.

2) Stacking-Induced Performance Gains: Several surveyed approaches stack two models, one each for symbolic and neural processes, into one combined model. Examples of such approaches can be found in §IV-A and §IV-C1. Sometimes, studies using these approaches claim to see improved performance in comparison to the use of the neural module alone [112, 153]. However, work such as that by Rivas-Barragan *et al.* [119] have found that *ensemble methods* for KGs tend to see improved performance than the use of a single model. This calls to question whether the observed performance increase is simply due to model stacking, as opposed to the neurosymbolic aspect.

3) Domain Knowledge Availability: Several neurosymbolic approaches discussed, such as those that require an ontology [153, 70], use domain-specific knowledge. Unfortunately, expert-curated knowledge is not readily available or digitized for every field. In such cases, these approaches are not easily applicable.

4) Ill-defined Interpretability: As discussed in §IV, a major characteristic of the surveyed approaches is their facilitation of model interpretability. However, the concept of interpretability is generally ill-defined [101], and it is not captured consistently across the surveyed approaches. Therefore, while there is a general consensus that neurosymbolic AI can improve interpretability, there is no official metric for it, making this a subjective evaluation.

TABLE II: Summary of Surveyed Neurosymbolic Approaches. **D** = incorporates domain knowledge, **L** = augments KG with logical inference, **C** = applies logical constraints, **W** = learns rule weights or confidences, **U** = updates candidate rule pool.

| | Reference | Year | KGs | D | L | C | W | U |
|------------------------------|---------------------------------|------|--|---|---|---|---|---|
| Logically-Informed Embedding | Walking RDF and OWL [4, 1] | 2017 | paper-specific biological KG | ✓ | ✓ | ✗ | ✗ | ✗ |
| | UniKER [26] | 2020 | WN18RR, FB15k-237, Kinship | ✗ | ✓ | ✗ | ✗ | ✗ |
| | RUGE [57] | 2020 | FB15K, YAGO37 | ✗ | ✓ | ✗ | ✗ | ✗ |
| | RW-autodrive [153] | 2020 | NuScenes, Lyft-Level5 | ✗ | ✓ | ✗ | ✗ | ✗ |
| | SoLE [165] | 2019 | FB15K, DB100K | ✗ | ✓ | ✗ | ✗ | ✗ |
| | KGE* [73] | 2022 | DBpedia20k, LUBM | ✓ | ✓ | ✗ | ✗ | ✗ |
| | ReasonKGE [71] | 2021 | DBpedia15k, LUBM3U, Yago3-10 | ✓ | ✓ | ✗ | ✗ | ✗ |
| Logical Constraints | Neuro-Symbolic Entropy Reg. [2] | 2022 | ACE05, SciERC | ✗ | ✗ | ✓ | ✗ | ✗ |
| | GeKCs [93] | 2023 | FB15k-237, WN18RR, ogbl-biokg | ✓ | ✗ | ✓ | ✗ | ✗ |
| | ReOnto [70] | 2023 | BioRel, ADE | ✓ | ✗ | ✓ | ✗ | ✗ |
| | R2N [95] | 2021 | Countries dataset, Nations dataset, UMLS, Kinship, Cora | ✗ | ✗ | ✓ | ✗ | ✗ |
| | SLRE [56] | 2020 | FB15K, DB100K | ✗ | ✗ | ✓ | ✗ | ✗ |
| | GCR [22] | 2022 | Amazon e-commerce, FB15k-237 | ✗ | ✗ | ✓ | ✗ | ✗ |
| | QLogicE [23] | 2022 | FB15k, FB15k-237, YAGO3-10, UMLS, Kinship, WN18RR | ✗ | ✗ | ✓ | ✗ | ✗ |
| | ComplEx-NNE_AER [36] | 2018 | FB15K, DBpedia, WN18 | ✗ | ✗ | ✓ | ✗ | ✗ |
| | KALE [55] | 2016 | WN18, FB122 | ✗ | ✗ | ✓ | ✓ | ✗ |
| | JOIE [60] | 2019 | YAGO26K-906, DB111K-174 | ✓ | ✗ | ✓ | ✗ | ✗ |
| | Graph Query Embeddings [58] | 2018 | paper-specific biological KG, Reddit | ✗ | ✗ | ✓ | ✗ | ✗ |
| | TFLEX [91] | 2024 | paper-original: ICEWS14, ICEWS05-15, GDELT-500 | ✗ | ✗ | ✓ | ✗ | ✗ |
| | CQD ^A [6, 5] | 2023 | FB15K, FB15K-237, NELL995 | ✗ | ✗ | ✓ | ✗ | ✗ |
| | KeGNN [152] | 2023 | Cora, Citeseer, PubMed, Flickr | ✗ | ✗ | ✓ | ✗ | ✗ |
| | Low-rank Logic Embeddings [120] | 2015 | Freebase | ✗ | ✓ | ✓ | ✗ | ✗ |
| | SimplE+ [43] | 2019 | FB15K, WN18 | ✗ | ✗ | ✓ | ✗ | ✗ |
| | KGE ^R [99] | 2017 | DBpedia, YAGO, WordNet | ✗ | ✗ | ✓ | ✗ | ✗ |
| Rule Learning | LPRules [32] | 2021 | FB15k-237, YAGO3-10, DB111K-174, WN18RR, UMLS, Kinship | ✗ | ✗ | ✗ | ✓ | ✓ |
| | RuLES [66] | 2018 | FB15K, Wiki44K, Family Dataset | ✗ | ✗ | ✗ | ✓ | ✓ |
| | pLogicNet [113] | 2019 | FB15k, FB15k-237, WN18, WN18RR | ✗ | ✓ | ✗ | ✓ | ✗ |
| | PoLo [92, 40] | 2021 | Hetionet, OREGANO KG | ✓ | ✗ | ✗ | ✓ | ✗ |
| | pGAT [61] | 2020 | WN18RR, FB15K-237 | ✗ | ✓ | ✗ | ✓ | ✗ |
| | DiffLogic [132] | 2023 | CodeX, YAGO3-10, WN18, WN18RR, Kinship | ✗ | ✓ | ✗ | ✓ | ✗ |
| | Neural LP [159] | 2017 | WordNet18, Freebase15K, Freebase15KSelected, UMLS, Kinship, WikiMovies | ✗ | ✗ | ✗ | ✓ | ✓ |
| | RNNLogic [114] | 2020 | WN18RR, FB15k-237, UMLS, Kinship | ✗ | ✗ | ✗ | ✓ | ✓ |
| | DRUM [124] | 2019 | UMLS, Kinship, Family Dataset, WN18RR, FB15K-237 | ✗ | ✗ | ✗ | ✓ | ✓ |
| | ExpressGNN [171] | 2020 | Cora, UW-CSE, Kinship, FB15K-237 | ✗ | ✓ | ✗ | ✓ | ✗ |
| | BioGRER [174] | 2020 | kg-covid-19 | ✓ | ✗ | ✗ | ✓ | ✗ |
| | SN-Hybrid [140] | 2020 | YAGO3-10, FB15K-237, WN18RR | ✗ | ✗ | ✗ | ✓ | ✓ |
| | Rule-IC [90] | 2021 | FB15k, FB15k-237, WN18, WN18RR | ✗ | ✗ | ✗ | ✓ | ✓ |
| | IterE [169] | 2019 | FB15k, FB15k-237, WN18, WN18RR | ✗ | ✗ | ✗ | ✓ | ✗ |
| | LNN-MP [130] | 2021 | UMLS, Kinship, WN18RR, FB15k-237 | ✗ | ✗ | ✗ | ✓ | ✓ |
| | Transductive Augmentation [64] | 2021 | WN18RR, FB15k-237 | ✗ | ✓ | ✗ | ✓ | ✗ |

B. Prospective Directions

1) *Underexplored Application Areas:* The selection of application domains chosen in the surveyed papers was largely driven by a preference for commonly used and openly available KGs, including YAGO [138] and DBpedia [8], which comprise general knowledge. While this is ideal for benchmarking, few studies demonstrated their method for a specific task or within a certain domain. Now, with our proposed taxonomy at hand and a general understanding of each category’s features (see Table II), prospective studies could aim to do so. For instance, surveyed methods which foster increased interpretability, such as those within §IV-C, are likely to be especially helpful for areas that closely affect peoples’ health or livelihoods, such as biomedicine, for reasons previously established, or autonomous driving, to ensure user and bystander safety [100]. Additionally, models which affect financial decisions, such as loan applications or customer turnover, might also have significant impacts

on business success and benefit from interpretability [100]. Alternatively, if a field is already well investigated and ample expert knowledge is available, this expertise could be used in a KG augmentation step as in §IV-A or as training constraints as in §IV-B. Example applications of this could include tasks within the natural sciences in which there are several openly available databases of information (*e.g.*, biology [7, 128] or chemistry [48]), language-oriented tasks which rely on human-defined rules [51, 175], and multimedia computing, which is built on human-made features [175]. Because this field is still developing, there are plenty of new territories that researchers might explore.

2) *Multimodal Data Integration:* Multimodal data describes a dataset or a combination of datasets that contain various forms, such as images, videos, text or long sequences and numerical measurements. For example, a news article might contain both text and images to describe a story [46, 84]. Specifically, multimodal data tends to be particularly

popular within the biomedical and autonomous driving domains, among others [84, 162]. While using a single type of data at a time is much simpler, various data formats can complement and complete one another. However, fusing the various data modalities into a unified input is challenging: one must find a way to represent all modalities similarly, such as numerically, while avoiding adding bias. Existing approaches tend to either (1) concatenate modalities into one, input vector, in which modalities are no longer distinguishable, (2) project modalities into representative embeddings via an autoencoder, or (3) use entirely separate models for each modality [46, 137]. However, a neurosymbolic approach could exploit the HETEROGENEOUS AGGREGATION characteristic by incorporating logic or domain-specific knowledge about the relationships between the modalities. Since multimodal data is valuable for the way its various forms complement one another, it is important to consider what kind of information these modalities contribute to one another, and how we can define these relationships through expert knowledge and logic.

3) Conditional or Interdependent Edge Types: In some cases, certain relations or edge types might be dependent upon one another in a way that one edge does not occur unless other edges exist. An example of this includes molecular signalling cascades, in which one protein will not interact with another unless contact is made with another protein or molecule upstream [21, 29]. Such dependence could also be frequency-dependent. In another molecular signalling example, if an upstream inhibitor interacts with some protein, the less that protein may take part in other interactions [53, 151]. A similar example can be observed in traffic forecasting, in which traffic on certain roads can increase or decrease that on other roads [54, 155]. While such dependencies are not unique to neurosymbolic reasoning, it might be ideal for addressing them. For instance, with the incorporation of rules, neurosymbolic methods could encode dependency information between relation types. Potentially, a method such as **GCR** [22], which uses logic to assess the probability that an edge is implied from its adjacent edges, could be ideal for interdependent edge types. Furthermore, since many neurosymbolic methods learn confidences for rules, there also exists the potential to learn confidences for which rules might coexist or influence one another. Neurosymbolic methods could, therefore, provide unique ways to interpret such dynamic KGs and the relationships between heterogeneous edge types.

4) Spatiotemporal Reasoning: KGs might also have spatial or temporal dependencies to consider. This is highly relevant for traffic forecasting [155], biological domains [74], and scene understanding [18]. Amongst previous approaches for reasoning over spatiotemporal KGs, many of which are available through the *Pytorch Geometric Temporal* package [122], there is still a division between those which take rule-based approaches and those which use deep learning approaches [149], with only one of the surveyed approaches accounting for temporal dependencies [91]. Thus, spatiotemporal applications are a promising direction for neurosymbolic hybridization.

Similarly to the proposed approach for interdependent edge types, neurosymbolic approaches could learn rules which describe the relationships between the various edge types and

time, for example. Alternatively, rules could be learned for each specific time range or spatial location. Gene expression, for instance, is the process by which our genetic code is transcribed and translated into functional products; the set of genetic codes which are expressed as well as the magnitude to which they are expressed varies completely between bodily tissue types [155, 17]. One could model the interactions between the functional products as a graph, but the levels to which they affect one another depend directly upon their existence and abundance, factors controlled by the tissue type [17]. This is an example of a network, therefore, which is spatially dependent, and learning rules specific to tissue types could unveil how processes differ between tissues. Additionally, short-term traffic forecasting, the prediction of traffic flows within a small time frame, relies heavily on spatial context, such as the layout and style of the roads being considered, as well as temporal context, such as the time of day or week [86]. Such rules could add another layer of interpretability to informs researchers on how underlying processes differ across tissues, locations, time periods, and more.

Alternatively, a neurosymbolic architecture could encode time- or space-dependent rules as constraints. For example, in a chemical or biological network, the half-lives of molecules such as messenger ribonucleic acid (mRNA) not only act as a time limit on potential interactions but also determine the levels to which various functional proteins are expressed, and therefore the extent to which those proteins can have effects on other players in the network [96, 17]. Such constraints can more realistically model spatiotemporal dependencies by adding logic or domain-specific rules.

5) Few Shot Learning: As mentioned under the UNDERREPRESENTED TYPES characteristic, some methods increase the number of rare relation types through logical inference. Such approaches could, therefore, be used as a solution for few shot learning problems. In few shot learning, there exist a small number of instances of some class within the training data [173]. In KGs, this could include node classes or relation types. Oftentimes, KGE methods neglect rare relation types [164]. This is often accounted for through sampling or regularization methods [126], meta-learning [164, 173], and learned attention coefficients or relation-specific parameters, [164] but neurosymbolic methods that exploit rule-based deduction could pose an alternative solution. Methods such as **pLogicNet** [113] and **pGAT** [61] augment the KG with logically inferred triples and feed the it into the neural module. The KG augmentation step may also increase the instances of rare relation types, so the neural module has a higher sample size on which to train. This poses potential to address few shot learning on KGs.

VI. CONCLUSION

Methods for reasoning on KGs are popular and widely applicable across domains [174, 40, 159, 22]. Therefore, it is unsurprising that there are already such a varied range of neurosymbolic methods, despite neurosymbolic AI being a young area of research. In this article, we introduce a taxonomy by which to classify these novel approaches based

on the ways they contribute toward balancing interpretability, knowledge-integration, and improved predictive performance within the context of KGC. Specifically, we found that the surveyed methods fit quite well into three major categories: (1) logically-informed embedding approaches, (2) embedding approaches with logical constraints, and (3) rule-learning approaches. Throughout the article, we not only compare the various approaches but also delve into deeper subcategories of classification. For easy reference, we summarized our findings in a tabular view and compiled the available code repositories on one GitHub page³. Finally, we propose prospective application-based and technical directions toward which this field might steer. Through this survey, we provide a comprehensive overview of existing methods for neurosymbolic reasoning on KGs with hopes to guide future research.

ACKNOWLEDGMENTS

LND is funded by the University of Edinburgh (UoE) Informatics Graduate School through the Global Informatics Scholarship. RFM is funded by the UoE Institute for Academic Development through the Principal's Career Development PhD Scholarship. We thank these institutions for their support. We also thank Emile van Krieken, Paola Galdi, Filip Smola, Matthew Whyte, and Zonglin Ji for their thoughtful contributions.

REFERENCES

- [1] Asan Agibetov and Matthias Samwald. “Fast and scalable learning of neuro-symbolic representations of biomedical knowledge”. In: *arXiv preprint arXiv:1804.11105* (2018).
- [2] Kareem Ahmed et al. “Neuro-symbolic entropy regularization”. In: *Uncertainty in Artificial Intelligence*. PMLR. 2022, pp. 43–53.
- [3] Kareem Ahmed et al. “Semantic probabilistic layers for neuro-symbolic learning”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 29944–29959.
- [4] Mona Alshahrani et al. “Neuro-symbolic representation learning on biological knowledge graphs”. In: *Bioinformatics* 33.17 (2017), pp. 2723–2730.
- [5] Erik Arakelyan et al. “Adapting Neural Link Predictors for Data-Efficient Complex Query Answering”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023.
- [6] Erik Arakelyan et al. “Complex Query Answering with Neural Link Predictors”. In: *International Conference on Learning Representations*. 2021, pp. 1–14.
- [7] Michael Ashburner et al. “Gene ontology: tool for the unification of biology”. In: *Nature Genetics* 25.1 (2000), pp. 25–29.
- [8] Sören Auer et al. “DBpedia: A nucleus for a web of open data”. In: *The Semantic Web*. Springer. 2007, pp. 722–735.
- [9] Pradeep Kr Banerjee et al. “Oversquashing in GNNs through the lens of information contraction and graph expansion”. In: *2022 58th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2022, pp. 1–8.
- [10] Chitta Baral and Michael Gelfond. “Logic programming and knowledge representation”. In: *The Journal of Logic Programming* 19 (1994), pp. 73–148.
- [11] Dimitris Bertsimas et al. “The price of interpretability”. In: *arXiv preprint arXiv:1907.03419* (2019).
- [12] Garrett Birkhoff and John Von Neumann. “The logic of quantum mechanics”. In: *Annals of mathematics* (1936), pp. 823–843.
- [13] David Bonanno et al. “An approach to explainable deep learning using fuzzy inference”. In: *Next-Generation Analyst V*. Vol. 10207. SPIE. 2017, pp. 132–136.
- [14] Antoine Bordes et al. “Translating embeddings for modeling multi-relational data”. In: *Advances in Neural Information Processing Systems* 26 (2013).
- [15] Armand Boschin et al. “Combining Embeddings and Rules for Fact Prediction”. In: *International Research School in Artificial Intelligence in Bergen (AIB 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2022.
- [16] Tom Brown et al. “Language models are few-shot learners”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1877–1901.
- [17] Christopher Buccellati and Matthias Selbach. “mRNAs, proteins and the emerging principles of gene expression control”. In: *Nature Reviews Genetics* 21.10 (2020), pp. 630–644.
- [18] Chengzhi Cao et al. “Discovering Intrinsic Spatial-Temporal Logic Rules to Explain Human Actions”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023.
- [19] Vincenzo Carletti et al. “Predicting Polypharmacy Side Effects Through a Relation-Wise Graph Attention Network”. In: *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer. 2021, pp. 119–128.
- [20] Andrea Cavallo et al. “GCNH: A Simple Method For Representation Learning On Heterophilous Graphs”. In: *arXiv preprint arXiv:2304.10896* (2023).
- [21] Lufen Chang and Michael Karin. “Mammalian MAP kinase signalling cascades”. In: *Nature* 410.6824 (2001), pp. 37–40.
- [22] Hanxiong Chen et al. “Graph Collaborative Reasoning”. In: *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, 2022, pp. 75–84.
- [23] Panfeng Chen et al. “QLogicE: Quantum Logic Empowered Embedding for Knowledge Graph Completion”. In: *Knowledge-Based Systems* 239 (2022), p. 107963.
- [24] Xiaojun Chen, Shengbin Jia, and Yang Xiang. “A review: Knowledge reasoning over knowledge graph”. In: *Expert Systems with Applications* 141 (2020), p. 112948.
- [25] Zhe Chen et al. “Knowledge graph completion: A review”. In: *IEEE Access* 8 (2020), pp. 192435–192456.
- [26] Kewei Cheng et al. “UniKER: A unified framework for combining embedding and Horn rules for knowledge graph inference”. In: *ICML Workshop on Graph Representation Learning and Beyond*. 2020.
- [27] Aakanksha Chowdhery et al. “Palm: Scaling language modeling with pathways”. In: *arXiv preprint arXiv:2204.02311* (2022).
- [28] William W Cohen. “Tensorlog: A differentiable deductive database”. In: *arXiv preprint arXiv:1605.06523* (2016).
- [29] Peter B Crino. “The mTOR signalling cascade: paving new roads to cure neurological disease”. In: *Nature Reviews Neurology* 12.7 (2016), pp. 379–392.
- [30] Andrew Cropper, Sebastijan Dumančić, and Stephen H Muggleton. “Turning 30: New ideas in inductive logic programming”. In: *arXiv preprint arXiv:2002.11002* (2020).
- [31] Ameya Daigavane, Balaraman Ravindran, and Gaurav Aggarwal. “Understanding convolutions on graphs”. In: *Distill* 6.9 (2021), e32.
- [32] Sanjeeb Dash and Joao Goncalves. “LPRules: Rule Induction in Knowledge Graphs Using Linear Programming”. In: *arXiv preprint arXiv:2110.08245* (2021).
- [33] Nur Nasuha Daud et al. “Applications of link prediction in social networks: A review”. In: *Journal of Network and Computer Applications* 166 (2020), p. 102716.
- [34] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. “ProbLog: A Probabilistic Prolog and Its Application in Link Discovery.” In: *IJCAI*. Vol. 7. Hyderabad. 2007, pp. 2462–2467.
- [35] Arthur P Dempster, Nan M Laird, and Donald B Rubin. “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the royal statistical society: series B (methodological)* 39.1 (1977), pp. 1–22.
- [36] Boyang Ding et al. “Improving knowledge graph embedding using simple constraints”. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics. 2018.
- [37] Daniel Domingo-Fernández et al. “COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology”. In: *Bioinformatics* 37.9 (2021), pp. 1332–1334.
- [38] Pedro Domingos et al. “Markov Logic”. In: *Probabilistic inductive logic programming*. Springer. 2008, pp. 92–117.
- [39] Dejing Dou, Hao Wang, and Haishan Liu. “Semantic data mining: A survey of ontology-based approaches”. In: *Proceedings of the 2015*

³<https://github.com/NeSymGraphs>

- IEEE 9th international conference on semantic computing (IEEE ICSC 2015)*. IEEE. 2015, pp. 244–251.
- [40] Martin Drancé et al. “Neuro-symbolic XAI for Computational Drug Repurposing.” In: *KEOD*. 2021, pp. 220–225.
- [41] Gintare Karolina Dziugaite, Shai Ben-David, and Daniel M Roy. “Enforcing Interpretability and its Statistical Impacts: Trade-offs between Accuracy and Interpretability”. In: *arXiv preprint arXiv:2010.13764* (2020).
- [42] Yin Fang et al. “Knowledge graph-enhanced molecular contrastive learning with functional prompt”. In: *Nature Machine Intelligence* 5.5 (2023), pp. 542–553.
- [43] Bahare Fatemi, Siamak Ravanbakhsh, and David Poole. “Improved knowledge graph embedding using background taxonomic information”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 01. 2019, pp. 3526–3533.
- [44] Benjamin LS Furman et al. “Sex chromosome evolution: so many exceptions to the rules”. In: *Genome biology and evolution* 12.6 (2020), pp. 750–763.
- [45] Luis Galárraga et al. “Fast rule mining in ontological knowledge bases with AMIE+”. In: vol. 24. 6. Springer, 2015, pp. 707–730.
- [46] Jing Gao et al. “A survey on deep learning for multimodal data fusion”. In: *Neural Computation* 32.5 (2020), pp. 829–864.
- [47] Artur d’Avila Garcez and Luis C Lamb. “Neurosymbolic AI: The 3 rd wave”. In: *Artificial Intelligence Review* (2023), pp. 1–20.
- [48] Anna Gaulton et al. “The ChEMBL database in 2017”. In: *Nucleic Acids Research* 45.D1 (2017), pp. D945–D954.
- [49] Aryo Pradipta Gemma et al. “Knowledge Graph Embeddings in the Biomedical Domain: Are They Useful? A Look at Link Prediction, Rule Learning, and Downstream Polypharmacy Tasks”. In: *arXiv preprint arXiv:2305.19979* (2023).
- [50] Francesco Di Giovanni et al. “How does over-squashing affect the power of GNNs?” In: *Transactions on Machine Learning Research* (2024).
- [51] Sujatha Das Gollapalli, Xiao-Li Li, and Peng Yang. “Incorporating expert knowledge into keyphrase extraction”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. 1. 2017.
- [52] Aditya Grover and Jure Leskovec. “node2vec: Scalable feature learning for networks”. In: *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 2016, pp. 855–864.
- [53] David Gubb et al. “Protease inhibitors and proteolytic signalling cascades in insects”. In: *Biochimie* 92.12 (2010), pp. 1749–1759.
- [54] Shengnan Guo et al. “Attention based spatial-temporal graph convolutional networks for traffic flow forecasting”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 922–929.
- [55] Shu Guo et al. “Jointly Embedding Knowledge Graphs and Logical Rules”. In: *EMNLP*. 2016.
- [56] Shu Guo et al. “Knowledge Graph Embedding Preserving Soft Logical Regularity”. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020, pp. 425–434.
- [57] Shu Guo et al. “Knowledge graph embedding with iterative guidance from soft rules”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1. 2018.
- [58] Will Hamilton et al. “Embedding logical queries on knowledge graphs”. In: *Advances in Neural Information Processing Systems* 31 (2018).
- [59] William L Hamilton. “The Graph Neural Network Model”. In: *Graph Representation Learning*. Springer, 2020, pp. 51–70.
- [60] Junheng Hao et al. “Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, pp. 1709–1719.
- [61] L Vivek Harsha Vardhan, Guo Jia, and Stanley Kok. “Probabilistic logic graph attention networks for reasoning”. In: *Companion Proceedings of the Web Conference 2020*. 2020, pp. 669–673.
- [62] David Herron, Ernesto Jiménez-Ruiz, and Tillman Weyde. “On the Benefits of OWL-based Knowledge Graphs for Neural-Symbolic Systems”. In: *Proceedings of the 17th International Workshop on Neural-Symbolic Learning and Reasoning*. Vol. 3432. CEUR Workshop Proceedings. 2023, pp. 327–335.
- [63] Daniel Scott Himmelstein et al. “Systematic integration of biomedical knowledge prioritizes drugs for repurposing”. In: *eLife* 6 (2017).
- [64] Yushi Hirose, Masashi Shimbo, and Taro Watanabe. “Transductive Data Augmentation with Relational Path Rule Mining for Knowledge Graph Embedding”. In: *2021 IEEE International Conference on Big Knowledge (ICBK)*. IEEE. 2021, pp. 377–384.
- [65] Pascal Hitzler et al. “Neuro-symbolic approaches in Artificial Intelligence”. In: *National Science Review* 9.6 (2022), nwac035.
- [66] Vinh Thinh Ho et al. “Rule learning from knowledge graphs guided by embedding models”. In: *The Semantic Web-ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I* 17. Springer. 2018, pp. 72–90.
- [67] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural computation* 9.8 (1997), pp. 1735–1780.
- [68] Alfred Horn. “On sentences which are true of direct unions of algebras!”. In: *The Journal of Symbolic Logic* 16.1 (1951), pp. 14–21.
- [69] Jui-Ting Huang et al. “Embedding-based retrieval in facebook search”. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020, pp. 2553–2561.
- [70] Monika Jain, Kuldeep Singh, and Raghava Mutharaju. “ReOnto: A Neuro-Symbolic Approach for Biomedical Relation Extraction”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2023, pp. 230–247.
- [71] Nitisha Jain et al. “Improving knowledge graph embeddings with ontological reasoning”. In: *The Semantic Web-ISWC 2021: 20th International Semantic Web Conference, ISWC 2021, Virtual Event, October 24–28, 2021, Proceedings*. Springer. 2021, pp. 410–426.
- [72] Shaoxiong Ji et al. “A survey on knowledge graphs: Representation, acquisition, and applications”. In: *IEEE Transactions on Neural Networks and Learning Systems* 33.2 (2021), pp. 494–514.
- [73] Zoi Kaoudi, Abelardo Carlos Martinez Lorenzo, and Volker Markl. “Towards Loosely-Coupling Knowledge Graph Embeddings and Ontology-based Reasoning”. In: *arXiv preprint arXiv:2202.03173* (2022).
- [74] Amol Kapoor et al. “Examining covid-19 forecasting using spatio-temporal graph neural networks”. In: *arXiv preprint arXiv:2007.03113* (2020).
- [75] Henry A. Kautz. “The third AI summer: AAAI Robert S. Engelmore Memorial Lecture”. In: *AI Magazine* 43.1 (2022), pp. 105–125. DOI: <https://doi.org/10.1002/aaai.12036>.
- [76] Mishal Kazmi, Peter Schüller, and Yücel Saygin. “Improving scalability of inductive logic programming via pruning and best-effort optimisation”. In: *Expert Systems with Applications* 87 (2017), pp. 291–303.
- [77] Tal Kiani et al. “Utilizing Perturbation of Atoms’ Positions for Equivariant Pre-Training in 3D Molecular Analysis”. In: *2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE. 2023, pp. 1–6.
- [78] Thomas N. Kipf and Max Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In: *International Conference on Learning Representations*. 2017.
- [79] Xiang Kong, Xianyang Chen, and Eduard Hovy. “Decompressing knowledge graph representations for link prediction”. In: *arXiv preprint arXiv:1911.04053* (2019).
- [80] Sophia Krix et al. “MultiGML: multimodal graph machine learning for prediction of adverse drug events”. In: *Heliyon* 9.9 (2023).
- [81] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in Neural Information Processing Systems* 25 (2012).
- [82] Markus Krötzsch, Frantisek Simancik, and Ian Horrocks. “A description logic primer”. In: *arXiv preprint arXiv:1201.4089* (2012).
- [83] Maxat Kulmanov et al. “Semantic similarity and machine learning with ontologies”. In: *Briefings in bioinformatics* 22.4 (2021).
- [84] Dana Lahat, Tülay Adali, and Christian Jutten. “Multimodal data fusion: an overview of methods, challenges, and prospects”. In: *Proceedings of the IEEE* 103.9 (2015), pp. 1449–1477.
- [85] Luis C Lamb et al. “Graph neural networks meet neural-symbolic computing: A survey and perspective”. In: *arXiv preprint arXiv:2003.00330* (2020).
- [86] Ibai Lana et al. “Road traffic forecasting: Recent advances and new challenges”. In: *IEEE Intelligent Transportation Systems Magazine* 10.2 (2018), pp. 93–109.
- [87] Bentian Li and Dechang Pi. “Network representation learning: a systematic literature review”. In: *Neural Computing and Applications* 32.21 (2020), pp. 16647–16679.
- [88] Jing Li, Shanshan Feng, and Billy Chiu. “Few-Shot Relation Extraction With Dual Graph Neural Network Interaction”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2023).

- [89] Jun Li et al. “Sparseness analysis in the pretraining of deep neural networks”. In: *IEEE Transactions on Neural Networks and Learning Systems* 28.6 (2016), pp. 1425–1438.
- [90] Qika Lin et al. “Rule-enhanced iterative complementation for knowledge graph reasoning”. In: *Information Sciences* 575 (2021), pp. 66–79.
- [91] Xueyuan Lin et al. “TFLEX: Temporal Feature-Logic Embedding Framework for Complex Reasoning over Temporal Knowledge Graph”. In: *Advances in Neural Information Processing Systems* 36 (2024).
- [92] Yushan Liu et al. “Neural multi-hop reasoning with logical rules on biomedical knowledge graphs”. In: *European Semantic Web Conference*. Springer. 2021, pp. 375–391.
- [93] Lorenzo Loconte et al. “How to Turn Your Knowledge Graph Embeddings into Generative Models”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023.
- [94] Giuseppe Marra, Michelangelo Diligenti, and Francesco Giannini. “Learning Representations for Sub-Symbolic Reasoning”. In: *arXiv preprint arXiv:2106.00393* (2021).
- [95] Giuseppe Marra and Ondřej Kuželka. “Neural markov logic networks”. In: *Uncertainty in Artificial Intelligence*. PMLR. 2021, pp. 908–917.
- [96] David M Mauger et al. “mRNA structure regulates protein expression through changes in functional half-life”. In: *Proceedings of the National Academy of Sciences* 116.48 (2019), pp. 24075–24083.
- [97] Deborah L McGuinness, Frank Van Harmelen, et al. “OWL web ontology language overview”. In: *W3C recommendation* 10.10 (2004), p. 2004.
- [98] Christian Meilicke et al. “Anytime Bottom-Up Rule Learning for Knowledge Graph Completion.” In: *IJCAI*. 2019, pp. 3137–3143.
- [99] Pasquale Minervini et al. “Regularizing knowledge graph embeddings via equivalence and inversion axioms”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2017, pp. 668–683.
- [100] Christoph Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. 2nd ed. 2022. URL: <https://christophm.github.io/interpretable-ml-book>.
- [101] Christoph Molnar, Giuseppe Casalicchio, and Bernd Bischl. “Interpretable machine learning—a brief history, state-of-the-art and challenges”. In: *Joint European conference on machine learning and knowledge discovery in databases*. Springer. 2020, pp. 417–431.
- [102] Stephen Muggleton. “Inverse entailment and Progol”. In: *New generation computing* 13.3 (1995), pp. 245–286.
- [103] Stephen Muggleton and Luc De Raedt. “Inductive logic programming: Theory and methods”. In: *The Journal of Logic Programming* 19 (1994), pp. 629–679.
- [104] Ndapandula Nakashole et al. “Query-Time Reasoning in Uncertain RDF Knowledge Bases with Soft and Hard Rules.” In: *VLDS* 884 (2012), pp. 15–20.
- [105] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. “A three-way model for collective learning on multi-relational data”. In: *International Conference on Machine Learning*. 2011.
- [106] Nils J Nilsson. “Probabilistic logic”. In: *Artificial intelligence* 28.1 (1986), pp. 71–87.
- [107] Humaira Noor et al. “Determining the Optimal Number of GAT and GCN Layers for Node Classification in Graph Neural Networks”. In: *2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS)*. IEEE. 2023, pp. 111–116.
- [108] Natalia Norori et al. “Addressing bias in big data and AI for health care: A call for open science”. In: *Patterns* 2.10 (2021), p. 100347.
- [109] Jacob P Olejarczyk and Michael Young. *Patient Rights And Ethics*. StatPearls Publishing, Treasure Island (FL). 2022. URL: <http://europepmc.org/books/NBK538279>.
- [110] Simon Ott, Christian Meilicke, and Matthias Samwald. “SAFRAN: An interpretable, rule-based link prediction method outperforming embedding models”. In: *arXiv preprint arXiv:2109.08002* (2021).
- [111] Luca Pasa et al. “Empowering simple graph convolutional networks”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2023).
- [112] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. “DeepWalk: Online Learning of Social Representations”. In: *KDD ’14*. New York, New York, USA: Association for Computing Machinery. 2014, pp. 701–710. ISBN: 9781450329569. DOI: 10.1145/2623330.2623732. URL: <https://doi.org/10.1145/2623330.2623732>.
- [113] Meng Qu and Jian Tang. “Probabilistic logic neural networks for reasoning”. In: *Advances in Neural Information Processing Systems* 32 (2019).
- [114] Meng Qu et al. “RNNLogic: Learning Logic Rules for Reasoning on Knowledge Graphs”. In: *International Conference on Learning Representations*.
- [115] J. Ross Quinlan. “Learning logical definitions from relations”. In: *Machine learning* 5.3 (1990), pp. 239–266.
- [116] Neal Ravindra et al. “Disease state prediction from single-cell data using graph attention networks”. In: *Proceedings of the ACM conference on health, inference, and learning*. 2020, pp. 121–130.
- [117] Matthew Richardson and Pedro Domingos. “Markov Logic Networks”. In: *Machine learning* 62.1 (2006), pp. 107–136.
- [118] Ryan Riegel et al. “Logical neural networks”. In: *arXiv preprint arXiv:2006.13155* (2020).
- [119] Daniel Rivas-Barragan et al. “Ensembles of knowledge graph embedding models improve predictions for drug discovery”. In: *Briefings in Bioinformatics* 23.6 (2022).
- [120] Tim Rocktäschel, Sameer Singh, and Sebastian Riedel. “Injecting logical background knowledge into embeddings for relation extraction”. In: *Proceedings of the 2015 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2015, pp. 1119–1129.
- [121] Frank Rosenblatt. “The perceptron: a probabilistic model for information storage and organization in the brain.” In: *Psychological review* 65.6 (1958), p. 386.
- [122] Benedek Rozemberczki et al. “Pytorch geometric temporal: Spatiotemporal signal processing with neural machine learning models”. In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021, pp. 4564–4573.
- [123] T Konstantin Rusch, Michael M Bronstein, and Siddhartha Mishra. “A survey on oversmoothing in graph neural networks”. In: *arXiv preprint arXiv:2303.10993* (2023).
- [124] Ali Sadeghian et al. “Drum: End-to-end differentiable rule mining on knowledge graphs”. In: *Advances in Neural Information Processing Systems* 32 (2019).
- [125] Md Kamruzzaman Sarker et al. “Neuro-symbolic artificial intelligence: Current trends”. In: *arXiv preprint arXiv:2105.05330* (2021).
- [126] Michael Schlichtkrull et al. “Modeling relational data with graph convolutional networks”. In: *European semantic web conference*. Springer. 2018, pp. 593–607.
- [127] Stefan Schoenmakers et al. “Learning first-order horn clauses from web text”. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. 2010, pp. 1088–1098.
- [128] Lynn M Schriml et al. “The Human Disease Ontology 2022 update”. In: *Nucleic Acids Research* 50.D1 (2022), pp. D1255–D1261.
- [129] Bruce Schultz et al. “The COVID-19 PHARMACOME: A method for the rational selection of drug repurposing candidates from multimodal knowledge harmonization”. In: *bioRxiv* (2021), pp. 2020–09.
- [130] Prithviraj Sen et al. “Combining Rules and Embeddings via Neuro-Symbolic AI for Knowledge Base Completion”. In: *arXiv preprint arXiv:2109.09566* (2021).
- [131] Zehui Shao et al. “Heterogeneous Graph Neural Network With Multi-View Representation Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 35.11 (2023), pp. 11476–11488. DOI: 10.1109/TKDE.2022.3224193.
- [132] Chen Shengyuan et al. “Differentiable Neuro-Symbolic Reasoning on Large-Scale Knowledge Graphs”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023.
- [133] Alex Sherstinsky. “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network”. In: *Physica D: Nonlinear Phenomena* 404 (2020), p. 132306.
- [134] Sandro Skansi. *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*. 1st. Springer Publishing Company, Incorporated. 2018. ISBN: 3319730037.
- [135] Giuseppe Spillo et al. “Knowledge-aware recommendations based on neuro-symbolic graph embeddings and first-order logical rules”. In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 2022, pp. 616–621.
- [136] Ashwin Srinivasan. *The Aleph Manual*. 2001. URL: <http://web.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/>.
- [137] Sören Richard Stahlhmidt, Benjamin Ulfenborg, and Jane Synnergren. “Multimodal deep learning for biomedical data fusion: a review”. In: *Briefings in Bioinformatics* 23.2 (2022).
- [138] 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*. 2007, pp. 697–706.

- [139] Zhengya Sun et al. “Scalable learning and inference in Markov logic networks”. In: *International Journal of Approximate Reasoning* 82 (2017), pp. 39–55.
- [140] Susheel Suresh and Jennifer Neville. “A hybrid model for learning embeddings and logical rules simultaneously from knowledge graphs”. In: *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2020, pp. 1280–1285.
- [141] Peter Tino, Lubica Benuskova, and Alessandro Sperduti. “Artificial neural network models”. In: *Springer Handbook of Computational Intelligence*. Springer, 2015, pp. 455–471.
- [142] Erico Tjoa and Cuntai Guan. “A survey on explainable Artificial Intelligence (XAI): Toward medical XAI”. In: *IEEE Transactions on Neural Networks and Learning Systems* 32.11 (2020), pp. 4793–4813.
- [143] Théo Trouillon et al. “Complex embeddings for simple link prediction”. In: *International conference on machine learning*. PMLR, 2016, pp. 2071–2080.
- [144] Eftymia Tsamoura, Timothy Hospedales, and Loizos Michael. “Neural-symbolic integration: A compositional perspective”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 2021, pp. 5051–5060.
- [145] Petar Veličković et al. “Graph attention networks”. In: *arXiv preprint arXiv:1710.10903* (2017).
- [146] Deepak Venugopal and Vibhav G Gogate. “Scaling-up importance sampling for Markov logic networks”. In: *Advances in Neural Information Processing Systems* 27 (2014).
- [147] Antonio Vergari et al. “Probabilistic circuits: Representations, inference, learning and applications”. In: *AAAI Tutorial* (2020).
- [148] Chad Vicknair et al. “A comparison of a graph database and a relational database: a data provenance perspective”. In: *Proceedings of the 48th annual Southeast regional conference*. 2010, pp. 1–6.
- [149] Jiapu Wang et al. “A Survey on Temporal Knowledge Graph Completion: Taxonomy, Progress, and Prospects”. 2023. arXiv: 2308.02457.
- [150] Jie Wang et al. “GUIDE: Training Deep Graph Neural Networks via Guided Dropout Over Edges”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [151] Yifei Wang et al. “RPI-GGCN: Prediction of RNA–Protein Interaction Based on Interpretability Gated Graph Convolution Neural Network and Co-Regularized Variational Autoencoders”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2024).
- [152] Luisa Werner et al. “Knowledge Enhanced Graph Neural Networks for Graph Completion”. In: *International Joint Conference on Artificial Intelligence 2023 Workshop on Knowledge-Based Compositional Generalization*. 2023.
- [153] Ruwan Wickramarachchi, Cory Henson, and Amit Sheth. “An evaluation of knowledge graph embeddings for autonomous driving data: Experience and practice”. In: *Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice (AAAI-MAKE 2020)*. 2020.
- [154] Zonghan Wu et al. “A comprehensive survey on graph neural networks”. In: *IEEE Transactions on Neural Networks and Learning Systems* 32.1 (2020), pp. 4–24.
- [155] Zonghan Wu et al. “Traversenet: Unifying space and time in message passing for traffic forecasting”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [156] Jingyi Xu et al. “A Semantic Loss Function for Deep Learning with Symbolic Knowledge”. In: *Proceedings of the 35th International Conference on Machine Learning*. Vol. 80. Proceedings of Machine Learning Research. PMLR, July 2018, pp. 5502–5511.
- [157] Yujun Yan et al. “Two sides of the same coin: Heterophily and oversmoothing in graph convolutional neural networks”. In: *2022 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2022, pp. 1287–1292.
- [158] Bishan Yang et al. “Embedding entities and relations for learning and inference in knowledge bases”. In: *arXiv preprint arXiv:1412.6575* (2014).
- [159] Fan Yang, Zhilin Yang, and William W Cohen. “Differentiable learning of logical rules for knowledge base reasoning”. In: *Advances in Neural Information Processing Systems* 30 (2017).
- [160] Liangwei Yang et al. “ConsisRec: Enhancing GNN for social recommendation via consistent neighbor aggregation”. In: *Proceedings of the 44th international ACM SIGIR conference on Research and development in information retrieval*. 2021, pp. 2141–2145.
- [161] Yu Yao et al. “Understanding how pretraining regularizes deep learning algorithms”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2021).
- [162] Jia-Li Yin et al. “Automatic dangerous driving intensity analysis for advanced driver assistance systems from multimodal driving signals”. In: *IEEE Sensors Journal* 18.12 (2017), pp. 4785–4794.
- [163] Lotfi A Zadeh. “Fuzzy logic”. In: *Computer* 21.4 (1988), pp. 83–93. DOI: 10.1109/2.53.
- [164] Chuxu Zhang et al. “Few-shot knowledge graph completion”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 03. 2020, pp. 3041–3048.
- [165] Jindou Zhang and Jing Li. “Enhanced knowledge graph embedding by jointly learning soft rules and facts”. In: *Algorithms* 12.12 (2019), p. 265.
- [166] Jing Zhang et al. “Neural, symbolic and neural-symbolic reasoning on knowledge graphs”. In: *AI Open* 2 (2021), pp. 14–35.
- [167] Muhan Zhang and Yixin Chen. “Link prediction based on graph neural networks”. In: *Advances in Neural Information Processing Systems*. 2018, pp. 5165–5175.
- [168] Rui Zhang et al. “Drug repurposing for COVID-19 via knowledge graph completion”. In: *Journal of biomedical informatics* 115 (2021), p. 103696.
- [169] Wen Zhang et al. “Iteratively learning embeddings and rules for knowledge graph reasoning”. In: *The World Wide Web Conference*. 2019, pp. 2366–2377.
- [170] Yu Zhang et al. “A survey on neural network interpretability”. In: *IEEE Transactions on Emerging Topics in Computational Intelligence* 5.5 (2021), pp. 726–742.
- [171] Yuyu Zhang et al. “Efficient Probabilistic Logic Reasoning with Graph Neural Networks”. In: *International Conference on Learning Representations*.
- [172] Zheng Zhang, Levent Yilmaz, and Bo Liu. “A Critical Review of Inductive Logic Programming Techniques for Explainable AI”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2023).
- [173] Feng Zhao, Tiancheng Huang, and Donglin Wang. “Graph few-shot learning via restructuring task graph”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [174] Sendong Zhao et al. “Biomedical knowledge graph refinement with embedding and logic rules”. In: *arXiv preprint arXiv:2012.01031* (2020).
- [175] Yue-ting Zhuang et al. “Challenges and opportunities: from big data to knowledge in AI 2.0”. In: *Frontiers of Information Technology & Electronic Engineering* 18.1 (2017), pp. 3–14.
- [176] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. “Modeling polypharmacy side effects with graph convolutional networks”. In: *Bioinformatics* 34.13 (2018), pp. i457–i466.