IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 November 2023, accepted 27 December 2023, date of publication 18 January 2024, date of current version 26 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3355457

RESEARCH ARTICLE

Restricting the Spurious Growth of Knowledge Graphs by Using Ontology Graphs

KINA TATCHUKOVA AND YANZHEN QU, (Senior Member, IEEE)
Department of Computer Science, Colorado Technical University, Colorado Springs, CO 80907, USA

Corresponding author: Yanzhen Qu (yqu@coloradotech.edu)

ABSTRACT Knowledge Graphs have demonstrated a real advantage in knowledge representation, leveraging graphs NoSQL structures and schema-less technology, which offers superior comprehension, knowledge representation, interpretation, and reasoning. The problem is that current methods for Knowledge Graph embedding rely on the topology of the graph, and essential information about entities and relations has not been fully employed, failing to utilize the graph's ontology to limit the spurious growth of edges, leading to inaccurate, misleading, and fabricated knowledge. This research aims to establish a method to restrict the spurious growth of host graph by imposing an upper bound on edge embedding using the claim's and the host's ontology graph. Through this research, a claim-ontology signature artifact was designed to facilitate open-environment KG completion. This artifact establishes the upper bound for the type of edges predicted by the link prediction algorithm, thus preventing the spurious growth of edges within the Knowledge Graph. Furthermore, the artifact was evaluated in the context of three use cases: host-guided embedding, claim-guided embedding, and topic-guided embedding, using a quantitative framework for design science evaluation. The main finding is that the spurious growth of edges can be limited by imposing an upper bound on the possible edge embedding using the claim's graph and the host ontology graph. A secondary finding is that the artifact could serve as an instrument to manage the ontology-topology tradeoff in Knowledge Graphs.

INDEX TERMS Knowledge graphs, knowledge graphs embedding, knowledge graph quality, link prediction.

I. INTRODUCTION

Knowledge Graphs (KGs) have demonstrated a real advantage in knowledge representation, leveraging graphs NoSQL structures and schema-less technology, which offers superior comprehension, knowledge representation, interpretation, and reasoning. Consequently, KGs are adopted as an essential part of Knowledge Management Systems across various domains, solving many real-life problems, including information retrieval, recommender systems, question-answering, and Artificial Intelligence (AI) [1], [2]. Knowledge Graphs are constructs of nodes that represent entities and edges that illustrate the relationships between these entities. This information is encoded using the Resource Description Framework (RDF), a W3C standard used for modeling semantic web objects [3], [4]. Under the RDF construct, KGs store knowledge or facts in triples ≪head, relation, tail≫ [5]. Industry examples of KGs include Google's KG, comprising over one billion entities and 70 billion assertions, and Microsoft Bing, containing over two billion entities and fifty-five billion facts. Facebook has the world's largest social graph, which includes information about people, places, movies, and music. For example, while eBay owns the largest Product Knowledge Graph, containing knowledge about products and manufacturers [6], [7].

With the maturation and standardization of semantic web technologies, resulting in a remarkable amount of data being shared on the web as linked data (LD), various stakeholders within the larger community of researchers and practitioners in Computer Science and Big Data Analytics have identified the importance of LD quality. In the semantic web context, KGs are perceived as essential knowledge systems products of LD, and their quality is commonly accepted as fitness for use [2], [8], [9]. The two commonly used dimensions

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao.

associated with KG's fitness for use are completeness and correctness. The former is defined as the degree to which the KG covers the topic of interest, and the latter is defined as the degree of freedom from error [10].

Fitness-for-use is not a fixed-value property, and to be fit for use, a KG must be continuously maintained by (a) ensuring completeness by updating the KG with new facts (claims) and (b) ensuring correctness by refining the KG by using graph mining techniques to create new knowledge by predicting new nodes and edges or pruning them, altering the KG structure [11], [12]. Tang et al. [11] indicated three approaches for updating KGs, manual, semi-automated, and automated, each comprising two essential steps. The first step is to extract information from a new claim, and the second step involves modifying the KG according to rules predefined by domain experts. Furthermore, Tang et al. [11] indicated that new claims contain explicit and implicit information, the latter being the information not mentioned in the claim but can be inferred from it. They claimed that the implicit information causes changes in the KG structure [11]. Moreover, by exploring the KG structure for application in Industry 4.0 applications, Yahya et al. [13] discussed the refinement of KGs as a process that runs on two levels: ontology and facts. According to Yahya et al. [13], large-scale KGs have two layers in their graph structure. The first layer comprises an ontology graph that, like schema, stores all the entities and properties used to label the facts. The second contains a facts graph that holds all the facts and the relationships between them. Using these two layers, Yahya et al. described the insertion of new claims into KGs by mapping the claims into ontological terms, thus converting them into a knowledge subgraph [13].

This research is concerned with inserting new claims into KGs and the link prediction process triggered by the claim insertion. Such insertion alters the host KG's topology and, in turn, its quality, which affects the KG's fitness for use. A quantitative Design Science Research (DSR) method was used to design and evaluate an artifact that, when embedded in the insertion process, organically regulates the evolution of the KG topology and the quality of the residing knowledge. The original contributions of this study are two-fold. First, it facilitates effective and reliable machine knowledge curation, discovery, and management by limiting the spurious growth of the KGs. Second, it promotes the use of ontology as an efficient KG knowledge management tool throughout the KG lifecycle. While a significant body of research explores KG updating and refining through multi-label classification and link prediction, no study has explored the connection between ontology and topology.

The rest of this article is structured as follows: Section II provides an overview of related work on KG completion, knowledge representation, link prediction, and artificial reasoning. Section III highlights the problem statement, hypotheses, and research question. Section IV discusses the research method and proposed artifact design and evaluation. Section V reports the findings and discusses the interpretation of the results and artifact applications, along with limitations and recommendations for future research. The final section provides the summary and conclusions.

II. RELATED WORK

A. KNOWLEDGE GRAPH COMPLETION

Knowledge Graph completion is the process of completing the structure of KGs by predicting the missing entities and relationships in the graphs and mining unknown facts [14], [15]. Large-scale KGs contain millions of entities and relationships. However, they still suffer from incompleteness due to missing (a) explicit observable facts and (b) implicit non-observable facts, resulting in an incomplete KG structure and content. Chen et al. attributed the incompleteness of KG's to manual or semi-automated KG construction methods, missing many implicit entities and relationships [14]. Wang et al. observed two critical aspects of successful KG completion, capturing the relational context and multi-faceted relational paths [15]. Researchers describe the KG completion problem as a "fill-in-the-blank" task. Chen et al. [14] suggested three types of KG completion tasks based on which component is missing from the RDF ≪h, r, t≫ representation. Each task type assumes that the other two components exist and aims to predict the missing component. For example, for the triple ≪?, r, t≫, the head is missing, but the relationship and tail are given; the goal is to predict the head entity [14].

B. KNOWLEDGE REPRESENTATION LEARNING

Knowledge representation learning is a branch of KG completion that mines the semantic information of KGs by using ML techniques. He et al. referred to this branch of KG completion methods as KG embedding and stated that this branch of models aims to represent entities and relationships in a low-dimensional continuous vector space [16]. According to Chen et al., it comprises three tasks, including (a) representing relationships and entities in a continuous space, (b) defining a score function that determines the probabilities of established triples, and (c) learning the representation of entities and relationships by solving the optimization problem of maximizing the rationality of observable facts [14]. This branch of methods has expanded in recent years and comprises four subgroups: translation models, semantic matching models, network representation learning models, and neural network (NN) models.

In a comprehensive review of these methods, Chen et al. [14] demonstrated that the translation models predict new entity relationship triples by embedding entities and relationships in the vector space using existing KG structured information. However, these models ignore the semantic information contained in the graph. In contrast, semantic matching models mine the semantic association between entities and relationships, leading to better accuracy and improved performance in natural language processing tasks. The Network representation learning models move further by fusing the information from the network topology structure and the attribute information of nodes and edges, thereby implementing the KG completion task using ML. Finally, neural network models for KG completion are based on various neural network (NN) methods, such as deep, recurrent, convolutional NN, and attention mechanisms. The main benefits of these models are mitigating the problem of data sparsity, handling multi-relational characteristics, and handling sequence-based data [14].

C. LINK PREDICTION

In Knowledge Graph completion and specifically representation learning methods, link prediction has been one of the most widely researched and applied tasks in recent years [17], [18], [19]. According to Zhou, link prediction is a paradigmatic problem in network science that aims to uncover missing relationships or predict future relationships [19]. Martinez et al. [18] stated that link prediction is a method that infers the behavior of the network link formation process based on observed connections. In both cases, the researchers indicated that link prediction methods operate on the topology of the network and the properties of its elements [18], [19]. Rossi et al. classified the link prediction methods into three categories: tensor decomposition, geometric learning, and deep learning, emphasizing that all models in the review learn from the KG structure [20]. Martinez-Rodriguez et al. [21], whose study encompassed complex networks, proposed a different taxonomy, including similarity-based, probabilistic, statistical, algorithmic, and pre-processing methods. They also differentiated between local and global structure approaches [21]. Ferrari et al. [17], completed a benchmarking analysis and provided another classification, including translation, semantic information, and neural network. Presenting their results, they stated that ‘‘the effectiveness of these models strictly depends on the KG properties’’ [17] (.p21).

Amongst the many classes of link prediction methods, one is of particular interest, namely the probabilistic and statistical approaches and, more specifically, the hierarchical structure models. This is because the method was used in this research and requires special attention. Clauset et al. [22] proposed a hierarchical structure model called a Hierarchical Random Graph. Provided an incompletely observed network, the model generates a set of HRGs with associated probabilities that fit the network. Then, it searches for pairs of nodes that are unobserved in the network and have a high average likelihood of connecting within these HRGs [22]. Alternative link prediction methods for link prediction are based on characteristics such as common neighbors, shortest paths between nodes, or the largest product of the node degree [18], [22].

D. ARTIFICIAL REASONING

The topic of artificial reasoning is extensive, encompassing the formation of technological, cognitive reasoning, and artificial argumentation. The current review is offered in the context of Knowledge Graphs as systems used in knowledge-grounded models and applications and their ability to extract and recycle information, creating models of the real world. Bryndin [23] pointed out that such capabilities require an ‘‘artificial mind,’’ defining them as an artificial modeling system. The main advantage of such a system is the cost and the speed of making technical decisions [23]. The benefits of such capabilities are demonstrated in applications such as transfer learning and generative AI, which are considered in this review.

Knowledge graphs enable transfer learning by providing explicit and interpretable mapping between domain spaces [24], [25]. This capability, observed independently by several researchers, including Ammanabrolu and Riedl [24], as a result of studying the efficiency of transfer between deep reinforcement learning agents designed to play text-adventure games, reveals two advantages of KGs in transfer learning. The first was reducing the policy training time between similar games. The second was increasing the quality of the learned control policy. Both are critical in text adventure games because learning in this context requires significant training and simulations [26]. More broadly, Petroni et al. emphasized the importance of knowledge transfer for a wide range of state-of-the-art tasks, such as latent context representation [27]. With transfer learning, researchers have identified two challenges. The first challenge is negative transfer, in which an arbitrary transfer decreases performance in the target domain. The second challenge is the interpretability of the transfer learning [25], [28], [29]. Zhuang et al. explained that the occurrence of negative transfer depends on several factors, including (a) relevance between source and target domains and (b) learners' ability to find transferable parts of knowledge across domains [29].

Generative AI refers to artificial intelligence that generates novel content through generative methods involving the distribution hypothesis, parameter estimation, and sampling new data from estimated models. According to Gozalo-Brizuela and Garrido-Merchan's review of state-of-the-art large generative AI, such models comprise a transformer and generator trained on a massive corpus in an enormous architecture [30]. Although, advanced computing technology enables the advancement of such models with the complexity of rich data without suffering from underfitting, generative AI models still face several limitations, including time-consuming training, bias, overconfidence, and lack of transparency [31]. The most significant limitation, according to Gozalo-Brizuela and Garrido-Merchan, is that ‘‘the models do not understand exactly what they are doing’’ [31](p.20). This limitation points to a lack of a justification for the model output of a knowledge-grounded ontology.

III. PROBLEM STATEMENT, HYPOTHESIS, AND RESEARCH QUESTION

A. PROBLEM STATEMENT

The problem is that current methods for Knowledge Graph embedding rely on the graph's topology, treating attribute triples as relation triples, and essential information about entities and relations has not been fully employed [32], [33], [34] failing to utilize the graph's ontology to limit the spurious growth of edges leading to false, misleading, and fabricated knowledge.

B. HYPOTHESIS

H_0: The spurious growth of KG cannot be limited by imposing an upper bound on the possible edge embedding using the claim graph and the host KG ontology graph.

H_1: The spurious growth of KG can be limited by imposing an upper bound on the possible edge embedding using the claim graph and host KG ontology graph.

C. RESEARCH QUESTION

How can the spurious growth of KG be limited by imposing an upper bound on the possible edge embedding using the claim and host's ontology graphs?

IV. METHODOLOGY

A. METHOD

The research method selected for this study is Design Science (DS) research following the research traditions in Computer Science and Information Technology, where practitioners design and develop artifacts [35] in response to specific requirements using experiential knowledge and practical reasoning [36]. This research offers a construct for limiting the spurious growth of the host KG and requires evaluating the construct using hypothesis testing. These two elements represent a process and a product, and as noted by Bisandu [37], comprise the signature of the design science research method. In addition, through Bisandu [37], following Hevner's guidelines for producing viable artifacts in Information Science research helped to ensure that DSR is the optimal methodological fit for such research [37]. Finally, Elragal and Haddara recommended several evaluation strategies, including hypothesis testing when a new theory, artifact, system, or method is designed to demonstrate the ex-ante vs. ex-post state of the factors in the context of the problem [38]. The data analysis software used for this quantitative analysis was R, a programming language for statistical computing with an integrated development environment (IDE) RStudio.

B. POPULATION AND SAMPLE

The study's target population was the RDF statements of a small KG created from the bibliography of the explored literature, including scholarly and electronic articles and book sections. The Knowledge Graph of these references was constructed using the igraph package for network analysis and visualization. The sampling methods included purposive and random sampling, which are typically used in quantitative research. The sample size was determined using the Handshaking Lemma to compute the total number of edges in a connected graph with n nodes. On average, a reference had seven nodes and 16 edges. Thus, a bibliography KG comprising ten references can result in a graph with approximately 70 nodes, including ten unique nodes representing titles, fewer than ten nodes representing years of publication, and more than ten authors since most papers are co-authored. By the Handshaking Lemma, such KG can have a maximum of 4,830 edges in total. After adding a further reference, seven nodes and 16 edges enter the KG, triggering a link prediction algorithm.

C. ARTIFACT DESIGN

The purpose of creating the artifact is to establish a method to limit the spurious growth of the host's edges by imposing an upper bound on the possible edge embedding using the ontology graphs of the claim and the host. In the design of this artifact, two issues were considered, including where and how this artifact is (a) generated and (b) applied. Before addressing each issue, a description of the problem context, including the process, the scope, and the parameters is provided. This process is depicted in Fig. 1, which describes the formation of the artifact as a transaction between the host and claim ontology graphs. Once the transaction is complete, the claim is inserted in the host KG along with the Claim-Ontology Signature (COS) artifact and the link prediction algorithm (LPA) is executed. The role of the COS is to filter the newly predicted edges before they are embedded in the host, thus limiting the growth of edges in the KG using the established ontology. The scope of the problem includes a host KG and a claim subgraph. The host KG comprises an ontology graph (OG) and a facts graph (FG). The OG defines the generalized concepts and their relationships, while the FG represents the facts and relationships. The claim subgraph defines the new facts and their relationships.

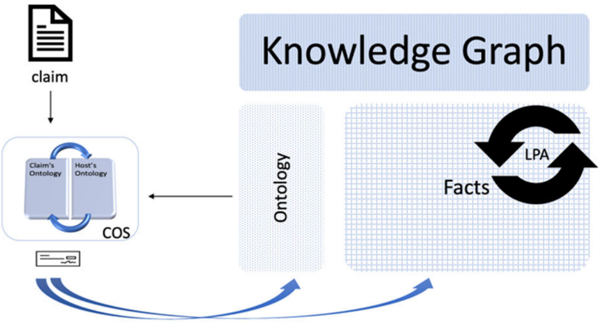

FIGURE 1. Artifact application process flow. The claim and host ontology graphs are the key components in the formation of the artifact.

The ontology layer comprises ontology triples and is the data model used to represent the semantics of the domain concepts through ontological terms. In contrast, the facts layer comprises only facts triples. The ontology layer defines the domain schema, where classes are entities and relationships are properties [13]. For instance, the ontology layer of a bibliography KG provides generalized terms of the classes, including authors, articles, journals, publishers, and the relationships between them. Fig. 2 illustrates that the ontology layer connects the author and article directly through a "wrote" relationship. However, there is no direct connection between the author and the journal, and such connections can be established through a path author-reference-journal because an author is a human, and they are not an entity that can be published in a journal. The KG's Ontology layer is represented using the RDF model where claims (facts) are expressed as a set of triples (subject, predicate, object), denoted as ⟨s, p, o⟩. Table 1 shows the RDF representation of the KG ontology.

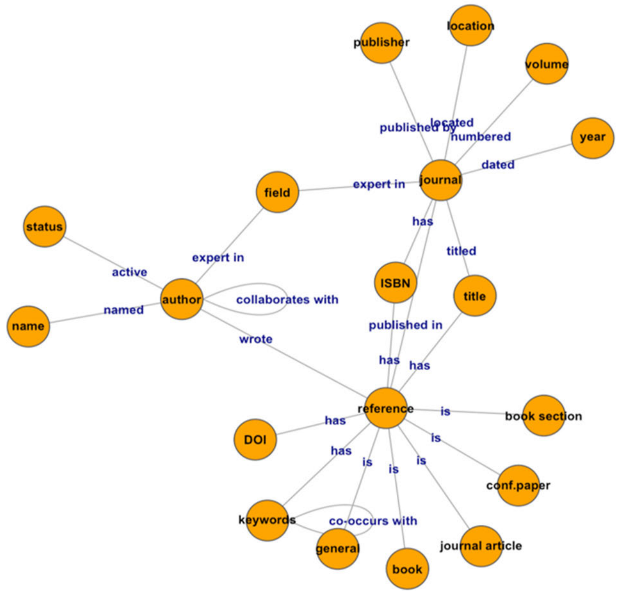

FIGURE 2. Host's ontology graph.

TABLE 1. Host ontology triples.

Triple ID	Subject	Predicate	Object
T1	Author/s	Wrote	Ref Type
T2	Author/s	Have	Name(s)
T3	Author/s	Is (expert in)	Expert Field
T4	Author/s	Is (active)	Research Status
T5	Ref Type	Has	Title
T6	Ref Type	Published by	Journal
T7	Ref Type	Has	Keyword/s
T8	Ref Type	Has	DOI
T9	Ref Type	Has	ISBN/ISSN
T10	Journal	Has	Title
T11	Journal	Published by	Publisher
T12	Journal	Identified	ISBN/ISSN
T13	Journal	Dated	Year
T14	Journal	Located	Place
T15	Journal	Numbered	Volume
T16	Journal	Named	Title
T17	Journal	Is in	Field

Next, the claim structure is also represented using the RDF framework. Fig. 3 illustrates an example of a claim, and the claim ontology footprint (COF) shown in Table 2. The claim ontology footprint contains five RDF ontology triples, represented by the ontology used to express the explicit facts and implicit information that is knowledge.

S. Tiwari, I. Bansal, and C. R. Rivero, "Revisiting the Evaluation Protocol of Knowledge Graph Completion Methods for Link Prediction," presented at the Proceedings of the Web Conference 2021. Published by Association for Computing Machinery

FIGURE 3. Textual representation of a claim.

TABLE 2. Claim ontology triples.

Triple ID	Subject	Predicate	Object
co1	Author(s)	Wrote	Reference Type
co2	Author(s)	Have	Name(s)
co3	Reference Type	Has	Title
co4	Reference Type	Is published by	Journal
co5	Journal	Has	Title (name)

Furthermore, ontology triples are the vehicle for knowledge discovery as these triples can trace the paths to connecting nodes within the existing network through validated relationships. This footprint serves as the foundation of the COS, establishing the upper bounds of each node entering the host KG. Once the claim ontology footprint is set, it is ready for mapping onto KG ontology. The mapping requires searching for a complete triple <s, p, o> that matches the claim triple. This results in a subset of RDF triples from the KG Ontology layer, as listed in Table 3, which forms the ontology projection of the claim. The claim triples cf1 and cf2, through their corresponding ontology footprint vectors <co1, co2> and <co3, co4, co5> map onto the KG's ontology vectors <T1, T2> and <T5, T6, T10>, each describing explicit relationships between the entities. These two vectors describe the explicit relationships between the entities identified in the claim and serve as the basis for constructing the claim ontology signature.

TABLE 3. COS matching triples.

Claim's Fact RDF	Claim's Ontology RDF	KG Ontology RDF	Predicate Type
cf1	co1	T1	Explicit
	co2	T2	Explicit
cf2	co3	T5	Explicit
	co4	T6	Explicit
	co5	T10	Explicit

Finally, constructing the claim ontology signature requires tracing the remaining ontology triples describing the explicit and implicit relationships between the claim entities and KG entities. This process searches for incomplete ontology triples of the form <s, %, %> where only the subject is known. In the example of the bibliography KG, the search for such triples where the subject is Reference results in ontology vector of a form <T7, T8>, representing explicit relationships, and no vector representing implicit relationships. Consequently, the combined number of edges (degrees) attributed to this node can be computed as the sum of the relationships carried by these vectors. In other words, the claim's entity Reference has a claim ontology signature that is a vector of size five with the following components: <T2, T5, T6, T7, T8>. This ontology signature represents the upper bound of the potential number of edges linked to this node once it enters the KG. This number remains fixed unless there is a change in the KG ontology, allowing for alteration of the COS. Table 4 lists the COS for each node of the example claim.

TABLE 4. Example of claim-ontology signature (COS).

Nodes	COS vector	Degree
Reference	<T5, T1, T6, T7, T8>	5
Author	<T2, T1, T3, T4>	4
Journal	<T6, T10, T11, T12, T13, T14, T15, T17>	8

Updating the KG using COS permits embedding only a limited number of edges, with each claim node bounded by the COS vector determined by the KG ontology graph. In addition, the type of each newly embedded edge is validated by the KG ontology, allowing only a specific ontology-validated relationship between the nodes. In other words, the attributes of the edges are always known, and only their value remains to be identified by a subsequent KG fact-completion process. The claim ontology signature allows the embedding of a limited number of ontology-validated edges, thus preventing the formation of false or misleading information.

D. ARTIFACT EVALUATION

The artifact was evaluated in the context of three use cases, including (a) Host-Guided embedding (HGE), (b) Claim-Guided embedding (CGE), and (c) Topic-Guided embedding (TGE). The motivation for selecting and developing the three use cases was two-fold. First, from a KG update perspective, it is known that KGs are updated with external information by synchronizing with existing encyclopedia sources (most likely structured) or extracting data from news streams (typically unstructured) [11]. Regardless of whether the information comes from structured or unstructured data sources, the methods for KG updating rely on elaborate information extraction systems and carefully formulated rules that are often domain-specific, difficult to maintain, and generalize [11]. These obstacles lead to two essential problems with KGs: inefficiency and brittleness. The inefficiency is due to the computational complexity of reasoning. The brittleness is due to the large set of handcrafted logical rules that need to be mined, especially in a dynamic environment [39]. Stakeholders may not be able to solve these two problems, but they can be alleviated by controlling the type and limits of information inserted into the host KG.

V. RESULTS, INTERPRETATIONS, AND APPLICATIONS

A. RESULTS

The artifact evaluation was focused on answering the main research question: How can the spurious growth of KG be limited by imposing an upper bound on the possible edge embedding using the claim and the host's ontology graph? The evaluation environment and protocol were identical across all three use cases while the inserted claims differed based on the use case objectives, driving the difference in the COS artifact limit values. The numerical results are presented in Tables 5 and 6, respectively.

TABLE 5. Tradeoff matrix by use case.

	HGE		CGE		TGE	
	e∉COS	e∈COS	e∉COS	e∈COS	e∉COS	e∈COS
P |e|=0	200	113	190	125	367	413
P |e|>0	100	17	80	32	151	18

TABLE 6. Example of claim-ontology signature (COS).

Test	HGE	CGE	TGE
Wilcoxon SRT	failed to reject	rejected	rejected
Kruskal-Wallis RST	rejected	rejected	rejected
Chi-Squared	rejected	rejected	rejected

The *Host-Guided embedding* use case described a scenario where the KG is sufficiently mature, and the stakeholders were mainly interested in curating new facts under the established KG ontology. As a result, the COS artifact was constructed based entirely on the basis of the host ontology graph. In the assessment of the effect of the artifact, the predicted edge types for the claim nodes exceeded the number of COS-prescribed edge types. While the Wilcoxon signed-rank test did not detect a statistically significant difference in the expected value, the Kruskal-Wallis rank-sum test showed that the variance between the two samples was statistically significant. This result shows that a COS based entirely on KG ontology can limit the number of edge types for some claim nodes but not all. Nevertheless, this result shows that the artifact can limit the growth of edges in the host.

Next, the tradeoff matrix created for the HGE use case showed that without the artifact, a total of 117 edges (the sum of edges in the column where the probability is greater than zero) could be added to the graph after the claim insertion. However, with the COS artifact, only 17 edges are defined by the host ontology and can be embedded. In addition, it was observed that the LP function assigned zero probability to the 113 edges that the COS flagged for embedding. These edges comprise relationships existing in the ontology graph but are not sufficiently represented in the fact graph. This observation raised the question of how to reconcile the information provided by ontology and topology.

The *Claim-Guided embedding* use case assumed that the KG was of sufficient maturity; however, the stakeholders were interested in expanding the KG ontology while simultaneously curating new facts. As a result, the COS artifact was constructed based on mapping the claim subgraph onto the graph ontology and adding new relationships defined by the claim. In the assessment of the effect of the artifact, the predicted edge types for the claim nodes exceeded the number of COS-prescribed edge types. Both, the Wilcoxon signed-rank test and the Kruskal-Wallis rank-sum test showed that the difference in the mean and variance in the two samples were statistically significant. Furthermore, the tradeoff matrix created for the CGE use case revealed that the artifact can limit the total number of predicted edges. Without the artifact, 112 edges could be added to the graph after the claim insertion. However, only 32 edges were inserted after the COS was applied. Here again, it was observed that the LP function assigned zero probability to 125 edges that the COS identified for potential embedding, leading to the need to reconcile the information between ontology and topology.

The *Topic-Guided embedding* use case assumed that the KG is of sufficient maturity; however, the stakeholders were interested in learning the KG on a specific topic, including expanding the ontology and facts graphs. As a result, the COS artifact was constructed by mapping the claim onto the graph ontology and adding the new relationships within the chosen topic. In the assessment of the effect of the artifact, the predicted edge types for the claim nodes exceeded the number of COS-prescribed edge types. Both, the Wilcoxon signed-rank test and the Kruskal-Wallis rank-sum test showed that the differences in the mean and variance in the two samples were statistically significant. Furthermore, the tradeoff matrix created for the TGE use case revealed that the artifact can limit the total number of updated edges. Without the artifact, 169 edges could be added to the graph after claim insertion. However, only 18 COS-prescribed edges were inserted in the host. Once again, it was observed that the LP function assigned zero probability to 413 COS-flagged edges. This number is much higher than the numbers observed in the previous two use cases, indicating that using graph topology alone may lead to missing information provided by the ontology.

B. INTERPRETATION AND APPLICATIONS

Knowledge Graphs combine factual data according to an ontology, facilitating the derivation of new knowledge by identifying non-observed entities and relationships. These capabilities make KGs an essential component in AI and ML through two functions: (a) promoting learning and explainability and (b) encoding, representing, and discovering domain knowledge that would be prohibitive for learning from large data sets alone. The use cases designed for COS artifact evaluation, including HGE, CGE, and TGE, illustrate how the artifact supports the two functions. In all three use cases, the host's ontology, and the facts graphs were separated to promote learning and explainability by facilitating the creation and application of the COS artifact. In this manner, the artifact facilitates proper link encoding of the observable links and bounds the number of non-observable links suggested by the topology-driven LP algorithm executed to update the KG.

Each use case tradeoff matrix illustrates the COS artifact's impact, that is, to limit the number of edges introduced after inserting a new claim into the host KG. Furthermore, an interesting result was the significant number of edges with zero probability belonging to the COS set. For each use case, these relationships were defined by the KG ontology, however the LP algorithm did not identify possible connections. One way to explain this phenomenon is by considering the structural redundancy of the host KG or lack thereof because, by construction, the host KG is heterogeneous. Zhou (2021) reported that networks with greater structural redundancy are more predictable [19]. In other words, the uniqueness of the nodes and edges in the host KG deterred the LP algorithm from identifying relationships between nodes. However, the COS artifact identified these edges as potential embedding candidates. This finding is significant in demonstrating how topology-based LP algorithms can miss ontology-defined connections in heterogeneous KGs. It also reveals the role of the COS artifact in detecting such missed links.

Next, in the context of the impact of ontology and KG topology on knowledge discovery, the results of the COS artifact evaluation revealed that based on the KG state and stakeholders' objectives, ontology, and KG topology are in a tradeoff relationship that impacts the direction and rate of knowledge discovery. It is unclear to what extent this relationship has been understood by the research community beyond the fact that ontology serves as a schema for KGs and schemas facilitate encoding and representing knowledge. Because these are the prerequisites for knowledge discovery, in particular, how complete and correct information is characterized, one can claim that ontology impacts the quality of the knowledge discovery process and, by extension, the topology of the KG. However, this reasoning is not actionable, although logical. In other words, the relationship between ontology and topology must be understood and leveraged for purposeful and efficient knowledge discovery.

The tradeoff relationship between ontology and topology was revealed while exploring and comparing the findings of the three use cases each designed with a specific host KG state and stakeholders' objectives, determining the computation of the COS artifact, in turn, and guiding the edge embedding. In the HGE use case, the host ontology has precedence over the claim ontology, and the host ontology graph alone sets the upper bounds for each claim node. Consequently, the topology of the host KG evolves strictly enclosed under its ontology. In contrast, in the CGE use case, the ontology of the host is open to accepting new relationship information. The claim and host ontology graphs performed a handshake-like transaction where new relationships were introduced by the claim previously non-observable in the host ontology. This transaction increases the upper bound limits set by the COS artifact. As a result, the host KG evolves in both ontology and topology.

Further, the Topic-Guided embedding use case was set to explore ontology evolving in a specific direction, such as a topic of interest that has not been observed before. Similar to the CGE use case, the claim and host ontology graphs performed a handshake-like transaction, where new relationships were introduced by the claim previously not observed in the host ontology. This transaction increases the upper bound limits set by the COS artifact. As a result, the host KG evolves in both ontology and topology. However, the difference with this use case is that the ontology grows in a specific direction and at a much higher rate.

Figure 3 illustrates the footprint on a log2 scale of the rate of embedding the claim edges by use case, where the tradeoff between topology (LPA) and ontology (COS) can be observed. The radar chart shows the rate-embedding envelope for each use case. All use cases start with the maximum possible number of edges that can connect the nodes of the new claim to the nodes of the host. The subsequent execution of link prediction determines the edges that should be embedded based on the topology of the host. In turn, the COS artifact determines the number of relationships defined by the host's ontology. Finally, the intercept of the last two sets determines the edges that can be reliably embedded in the host.

The observed tradeoff relationship between ontology and topology can be described as the tradeoff between learning new concepts and learning new facts within the same concept. If the host KG preserves its ontology, then what is left to be added to the KG is new facts, and over time, the topology of this KG will become homogeneous and, as a result, more predictable. By contrast, if the ontology is allowed to evolve, whether in general or in a specific topic, the topology of the host will develop accordingly, but it will remain less predictable. Regardless of the scenario, the COS artifact can facilitate the management of KG embedding either by limiting the spurious growth of edges or by detecting edges that are missed by topology-based LP algorithms. To a broader degree, the claim ontology signature can facilitate a better understanding of the ontology-topology tradeoff.

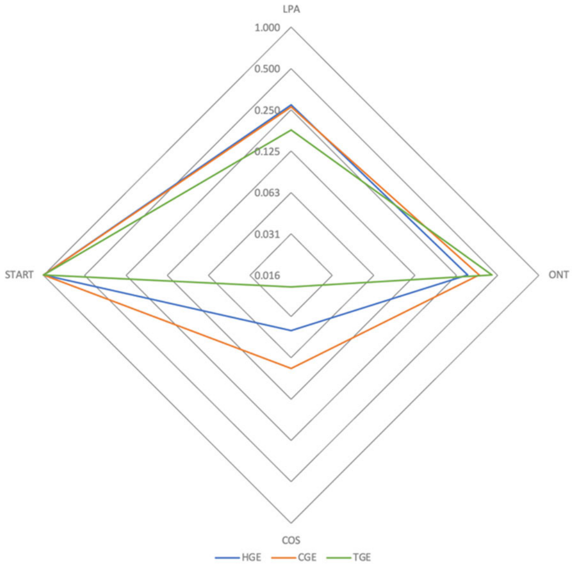

FIGURE 4. **Topology vs. ontology tradeoff envelop, illustrating the footprint in log2 scale of the rate of embedding the claim edges by use case.**

The application of the COS artifact as an instrument that facilitates the ontology-topology tradeoff can be explored in the pursuit of enforcing semantic knowledge. As demonstrated in this study, the COS artifact ensures that new information is (a) aligned with the host KG ontology and (b) each entity receives an ontology-validated upper bound of the number and type of potential links with other entities within the host. The COS artifact limits the number of unobserved links possibly predicted by the LP algorithms that are impacted by the graph topology. When transfer learning is performed using KGs representing the source and target domains, the challenges of transfer learning can be mitigated using the COS artifact, reinforcing positive transfer and limiting the negative transfer effect while improving interpretability. Enforcing semantic knowledge in generative AI systems by maintaining an ontology graph that supports the transformer and generator models can be a way to mitigate their limitations, help control this bias, limit hallucinations, and improve the accuracy of the generated content.

The impact of the COS artifact which aids the enforcement of semantic knowledge in transfer learning and generative AI systems can be evaluated through socioeconomic lenses, considering that AI-generated content has emerged as a promising force for innovation. A study by PriceWaterhouse-Coopers quoted by Du et al. reported that AI can potentially increase global GDP by 14% or nearly $15.7 trillion by 2030 [40]. While capabilities demonstrated by a variety of models such as ChatGPT developed by OpenAI and Stable Diffusion launched by Stability AI are rapidly becoming essential, the increasing costs of development and deployment due to large datasets and complex architecture required to achieve reliable performance still impede adoption across industries.

Professional users seek domain-grounded knowledge and transparency in these generative AI models.

C. LIMITATIONS

This section outlines the limitations of this research concerning the characteristics of the evaluation environment, including the type, size, and topology structure of the KG and the ability to explore the proposed COS artifact with available open-source link prediction algorithms. Limitations are weaknesses associated with the choice of research design, system or model constraints, or other factors outside of the researcher's control [41]. The choice of the evaluation environment impacts the artifact evaluation results primarily through the topology structure of the host KG and the choice of link prediction method. The network topology affects link predictability, revealing the effectiveness of the artifact. According to Zhou's review of link prediction methods, researchers have found that the network topology affects link predictability. In the same review, Zhou revealed that a network can be highly predictable if adding links does not significantly affect its topological structure [19].

The choice of the host KG was prompted primarily by the availability of domain-specific open-source KGs with established ontology graphs. Because of the rich entity and relationship ecosystem of domain knowledge, such KGs could have been a more suitable evaluation environment because their size and topology structure would impact the link prediction algorithm and, as a result, illustrate how the COS artifact performs within a rich and heterogeneous domain-driven environment and, consequently, strengthen the generalizability and transferability of the evaluation results. However, such KGs were not publicly available during this research. Although limited in size, both the ontology and fact graphs of the host KG have been designed to ensure their heterogeneous structures, that is, they contain nodes and edges of more than one type [42]. In this sense, the design supports the validity and reliability of artifact evaluation.

Next, the choice of link prediction algorithm was driven by the implementation availability within the igraph package. Zhou [19] stated that various automated link prediction algorithms have been proposed and implemented using network topology to estimate the likelihood of non-observed links. However, in the same review, he pointed out that very few algorithms have been compared using one or two metrics and on several small networks [19]. Based on these findings, the hierarchical random graph model to predict missing edges from a network implemented in the igraph package was chosen for artifact evaluation. Its impact on the validity and trustworthiness of the results is limited within the igraph package implementation, being an open-source product.

D. RECOMMENDATIONS FOR FUTURE RESEARCH

The utility value of Knowledge Graphs has been demonstrated throughout this paper, and the importance of researching ways to improve KG quality by limiting the spurious growth of its edges using the graph's ontology is paramount. Moreover, exploring ways to maintain and update KGs by enforcing semantic knowledge is in line with the recommendations of other researchers in the current knowledge-grounded scholarship. As a result, further exploration of methods to limit, reduce, or prevent the spurious growth of edges in KGs, especially utilizing the graphs' semantic knowledge, is recommended. Such efforts will increase the utility of KGs and their downstream applications, such as transfer learning and generative AI models.

The application of the research methodology described in the Methods section was limited to the characteristics of the evaluation environment, including the type, size, and topology structure of the KG and the ability to explore the COS artifact with available open-source link prediction algorithms. Research with alternative topologies, domain-specific KGs, and link prediction algorithms would be the next natural step that can facilitate the exploration of richer domain-specific topologies and help uncover potential challenges not encountered in this research.

Several directions are possible for future research regarding the integration and impact of the COS artifact on knowledge-grounded models. First, integrating the COS artifact into transfer learning and generative AI models requires further exploration owing to the complexity, wide range of applications, and challenges of these models. Researchers in transfer learning agree that transfer learning techniques can be further explored with new methods to solve complex knowledge transfer scenarios from relevant source domains, looking to avoid negative transfer and improve interpretability [29]. Therefore, it is recommended to research the integration of the COS artifact into transfer learning models and its impact on negative transfer and interpretability.

E. SUMMARY

The claim-ontology signature artifact is a data structure computed as a mapping of the claim ontology graph onto the host ontology graph. The artifact was evaluated in the context of three use cases: host-guided embedding, claim-guided embedding, and topic-guided embedding, each describing a specific KG state and stakeholder's objective and the corresponding COS artifact. In all three cases, the host ontology served as a blueprint for identifying explicit and implicit relationships between the entities defined within the claim. However, under each use case objective, either the host, claim, or topic subgraph guides the construction of the COS and determines the upper bound of the type and number of edges permitted to grow in the fact graph after the LP execution.

The evaluation results revealed that the COS artifact can limit the spurious growth of edges in the KG of the evaluation environment. Regarding specific use cases, the COS artifact can affect the connectivity of some claim nodes in the HGE and all nodes in the CGE and TGE use cases. One finding resulting from the tradeoff matrix calculations across all use cases revealed the existence of COS-identified zero-likelihood edges. In all use cases, these edges represent relationships described by the graph or claim ontology. This finding shows that the COS artifact effectively prevents knowledge loss when used with topology-based LP methods. A comparison of the three use cases revealed an element of a tradeoff between the host topology and ontology that plays a role in KG embedding.

VI. CONCLUSION

The purpose of this research was to establish a method to limit the spurious growth of edges in KGs by imposing an upper bound on the possible link embedding using the ontology graphs of the host KG and claim. This study was conducted to develop and evaluate a claim-ontology signature artifact, facilitating the validation of new links predicted for embedding by topology-driven link prediction algorithms. The main research question posed was: How can the spurious growth of KGs be limited by imposing an upper bound on the possible edge embedding using the claim and host ontology graph? The study presented the COS artifact's design and evaluation in the context of current knowledge graph scholarship using open-source software packages to construct a bibliography-based evaluation environment and practice-based use cases. The main finding from the research revealed that the spurious growth of KGs can be limited by imposing an upper bound on the possible edge embedding using the claim and host ontology graph. Further findings revealed that the COS artifact effectively prevents knowledge loss and could serve as an instrument to manage the ontology-topology tradeoff in KGs which plays a role in KG embedding.

The applicability and impact of this research were reviewed in the context of current KG research and its applications, including ML and AI. In particular, the findings of this research can be applied to areas of ML, such as transfer learning, which faces challenges, including negative transfer and lack of interpretability. This research links these concepts to the concept of ontology-topology tradeoff and the method of embedding new edges representing relationships in the target task ontology when training data is scarce. Specifically, in AI, knowledge-grounded dialog systems and generative AI models face challenges, including poor performance on topics unseen in the training data, inability to generalize to different knowledge source domains, bias, and accuracy. This study links these challenges to the idea of enforcing semantic knowledge and discusses the application of the COS artifact as an instrument for validating input or generating output by such AI systems. The limitations of this study relate to the characteristics of the evaluation environment, including the type, size, and topology structure of the KG and the ability to explore the COS artifact with available open-source link prediction algorithms. The choice of the host KG, prompted by the availability of domain-specific open-source KGs, impacts the performance of the HRG based link prediction algorithm and, as a result, demonstrates COS artifact performance.

With these findings and their potential applications, several focus areas for future research have been identified, including exploring alternatives to the COS artifact, exploring the artifact in alternative settings, and researching ways to integrate the artifact with existing knowledge-grounded ML and AI downstream applications. The first direction calls for further exploration of the COS artifact as a construct and potential data structure and storage solutions. The second direction aims to overcome the limitations of the current research framework and explore the COS artifact in an industry-domain ontology-rich environment. Finally, the third direction focuses on the practical applications of this research concerning integration of the COS artifact with knowledge-grounded transfer learning and generative AI models. Exploring how different industries with a defined ontology may integrate a COS artifact can be beneficial for evaluating the artifact's ability to enforce semantic knowledge.

REFERENCES

[1] B. Abu-Salih, "Domain-specific knowledge graphs: A survey," *J. Netw. Comput. Appl.*, vol. 185, Jul. 2021, Art. no. 103076.

[2] B. Xue and L. Zou, "Knowledge graph quality management: A comprehensive survey," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 5, pp. 4969–4988, May 2023.

[3] V. Chaudhri, C. Baru, N. Chittar, X. Dong, M. Genesereth, J. Hendler, A. Kalyanpur, D. Lenat, J. Sequeda, D. Vrandečić, and K. Wang, "Knowledge graphs: Introduction, history and, perspectives," *AI Mag.*, vol. 43, no. 1, pp. 17–29, Mar. 2022.

[4] M. T. Özsu, "A survey of RDF data management systems," *Frontiers Comput. Sci.*, vol. 10, no. 3, pp. 418–432, Jun. 2016.

[5] F. Lu, P. Cong, and X. Huang, "Utilizing textual information in knowledge graph embedding: A survey of methods and applications," *IEEE Access*, vol. 8, pp. 92072–92088, 2020, doi: 10.1109/ACCESS.2020.2995074.

[6] N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor, "Industry-scale knowledge graphs: Lessons and challenges," *Commun. ACM*, vol. 62, no. 8, pp. 36–43, Jul. 2019, doi: 10.1145/3331166.

[7] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic Web*, vol. 8, no. 3, pp. 489–508, Dec. 2016.

[8] M. Färber, F. Bartscherer, C. Menne, and A. Rettinger, "Linked data quality of DBpedia, freebase, OpenCyc, wikidata, and Yago," *Semantic Web*, vol. 9, no. 1, pp. 77–129, Nov. 2017.

[9] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer, "Quality assessment for linked data: A survey," *Semantic Web*, vol. 7, no. 1, pp. 63–93, Mar. 2015.

[10] D. Fensel, U. Simsek, and K. Angele, *Knowledge Graphs: Methodology, Tools and Selected Use Cases*, 1st ed. Cham, Switzerland: Springer, 2020.

[11] J. Tang, Y. Feng, and D. Zhao, "Learning to update knowledge graphs by reading news," in *Proc. 2019 Conf. Empirical Methods Natural Lang. Process., 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 2632–2641.

[12] E. Huaman, "Steps to knowledge graphs quality assessment," 2022, *arXiv:2208.07779*.

[13] M. Yahya, J. G. Breslin, and M. I. Ali, "Semantic web and knowledge graphs for industry 4.0," *Appl. Sci.*, vol. 11, no. 11, p. 5110, May 2021.

[14] Z. Chen, Y. Wang, B. Zhao, J. Cheng, X. Zhao, and Z. Duan, "Knowledge graph completion: A review," *IEEE Access*, vol. 8, pp. 192435–192456, 2020.

[15] X. Wang, L. Chen, T. Ban, M. Usman, Y. Guan, S. Liu, T. Wu, and H. Chen, "Knowledge graph quality control: A survey," *Fundam. Res.*, vol. 1, no. 5, pp. 607–626, 2021.

[16] P. He, G. Zhou, Y. Yao, Z. Wang, and H. Yang, "A type-augmented knowledge graph embedding framework for knowledge graph completion," *Sci. Rep.*, vol. 13, no. 1, p. 12364, Jul. 2023, doi: 10.1038/s41598-023-38857-5.

[17] I. Ferrari, G. Frisoni, P. Italiani, G. Moro, and C. Sartori, ''Comprehensive analysis of knowledge graph embedding techniques benchmarked on link prediction,'' *Electronics*, vol. 11, no. 23, p. 3866, Nov. 2022.

[18] V. Martínez, F. Berzal, and J.-C. Cubero, ''A survey of link prediction in complex networks,'' *ACM Comput. Surv.*, vol. 49, no. 4, pp. 1–33, Dec. 2017.

[19] T. Zhou, ''Progresses and challenges in link prediction,'' *Iscience*, vol. 24, no. 11, pp. 1–17, 2021.

[20] A. Rossi, D. Barbosa, D. Firmani, A. Matinata, and P. Merialdo, ''Knowledge graph embedding for link prediction: A comparative analysis,'' *ACM Trans. Knowl. Discovery Data*, vol. 15, no. 2, pp. 1–49, Apr. 2021, doi: 10.1145/3424672.

[21] J. L. Martinez-Rodriguez, A. Hogan, and I. Lopez-Arevalo, ''Information extraction meets the semantic web: A survey,'' *Semantic Web*, vol. 11, no. 2, pp. 255–335, Feb. 2020.

[22] A. Clauset, C. Moore, and M. E. J. Newman, ''Hierarchical structure and the prediction of missing links in networks,'' *Nature*, vol. 453, no. 7191, pp. 98–101, May 2008.

[23] E. Bryndin, ''Formation of technological cognitive reason with artificial intelligence in virtual space,'' *Britain Int. Exact Sci. (BIoEx) J.*, vol. 2, no. 2, pp. 450–461, May 2020.

[24] P. Ammanabrolu and M. O. Riedl, ''Transfer in deep reinforcement learning using knowledge graphs,'' 2019, *arXiv:1908.06556*.

[25] Y. Geng, J. Chen, E. Jimenez-Ruiz, and H. Chen, ''Human-centric transfer learning explanation via knowledge graph [extended abstract],'' 2019, *arXiv:1901.08547*.

[26] P. Ammanabrolu and M. O. Riedl, ''Playing text-adventure games with graph-based deep reinforcement learning,'' 2018, *arXiv:1812.01628*.

[27] F. Petroni, T. Rocktäschel, P. Lewis, A. Bakhtin, Y. Wu, A. H. Miller, and S. Riedel, ''Language models as knowledge bases?'' 2019, *arXiv:1909.01066*.

[28] S. Ambruster, ''Exploring strategies big data analysts need to improve classifiers for quality of personal data predictors,'' Ph.D. dissertation, Dept. Comput. Sci., Colorado Tech. Univ., Ann Arbor, MI, USA, 2019. [Online]. Available: https://www.proquest.com/dissertations-theses/exploring-strategies-big-data-analysts-need/docview/2313404900/se-2?accountid=144789

[29] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, ''A comprehensive survey on transfer learning,'' *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[30] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, ''Generative adversarial networks: Introduction and outlook,'' *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 4, pp. 588–598, 2017, doi: 10.1109/JAS.2017.7510583.

[31] R. Gozalo-Brizuela and E. C. Garrido-Merchan, ''ChatGPT is not all you need. A state of the art review of large generative AI models,'' 2023, *arXiv:2301.04655*.

[32] Z. Qiao, Z. Ning, Y. Du, and Y. Zhou, ''Context-enhanced entity and relation embedding for knowledge graph completion,'' 2020, *arXiv:2012.07011*.

[33] H. Xiao, M. Huang, L. Meng, and X. Zhu, ''SSP: Semantic space projection for knowledge graph embedding with text descriptions,'' in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1–7.

[34] Z. Zhang, L. Cao, X. Chen, W. Tang, Z. Xu, and Y. Meng, ''Representation learning of knowledge graphs with entity attributes,'' *IEEE Access*, vol. 8, pp. 7435–7441, 2020, doi: 10.1109/ACCESS.2020.2963990.

[35] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, ''A design science research methodology for information systems research,'' *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, Dec. 2007.

[36] S. T. March and G. F. Smith, ''Design and natural science research on information technology,'' *Decis. Support Syst.*, vol. 15, no. 4, pp. 251–266, Dec. 1995, doi: 10.1016/0167-9236(94)00041-2.

[37] D. Bisandu, ''Design science research methodology in computer science and information systems,'' *Int. J. Inf. Technol.*, vol. 5, no. 4, pp. 55–60, 2016. [Online]. Available: https://www.researchgate.net/publication/330041719_Design_science_research_methodology_in_Computer_Science_and_Information_Systems

[38] A. Elragal and M. Haddara, ''Design science research: Evaluation in the lens of big data analytics,'' *Systems*, vol. 7, no. 2, p. 27, May 2019, doi: 10.3390/systems7020027.

[39] A. Garcia-Duran and M. Niepert, ''KBLRN: End-to-end learning of knowledge base representations with latent, relational, and numerical features,'' 2017, *arXiv:1709.04676*.

[40] H. Du, Z. Li, D. Niyato, J. Kang, Z. Xiong, H. Huang, and S. Mao, ''Diffusion-based reinforcement learning for edge-enabled AI-generated content services,'' 2023, *arXiv:2303.13052*.

[41] D. Theofanidis and A. Fountouki, ''Limitations and delimitations in the research process,'' *Perioperative Nursing-Quarterly Sci., Online Off. J. GORNA*, vol. 7, no. 3, pp. 155–163, 2018.

[42] S. Gu, J. Johnson, F. E. Faisal, and T. Milenkovic, ''From homogeneous to heterogeneous network alignment via colored graphlets,'' *Sci. Rep.*, vol. 8, no. 1, p. 12524, Aug. 2018, doi: 10.1038/s41598-018-30831-w.

KINA TATCHUKOVA received the B.S. degree in applied and computational mathematics and sciences and the M.S. degree in applied mathematics from the University of Washington, Seattle, USA, and the Ph.D. degree in computer science from Colorado Technical University, Colorado Springs, CO, USA.

In her professional career, she has worked at various organizations and supporting multiple functions, including manufacturing, supply chain, and aftermarket services. She is an Applied Mathematician at an aerospace company in Seattle, WA, USA. Her current research interests include machine learning and artificial intelligence applied in domain-specific knowledge-based systems to improve knowledge curation, discovery, and management.

YANZHEN QU (Senior Member, IEEE) received the B.Eng. degree in electronic engineering from Anhui University, China, the M.Eng. degree in electrical engineering from the Chinese Science Academy, China, and the Ph.D. degree in computer science from Concordia University, Canada.

Throughout his industrial professional career, he has served in various executive-level management positions responsible for product research and development and IT operations in a few multinational corporations. He has led multinational engineering teams to successfully develop several of the world's first very large real-time commercial systems and technologies. He is currently the Dean and a Professor of computer science, engineering, and technology with Colorado Technical University, Colorado Springs, CO, USA. He is also a Dissertation Supervisor for many doctoral computer science students. He and his doctoral students have published several dozen scholarly articles, some of which have received the best paper award at several IEEE international conferences. His current research interests include data science, cybersecurity and privacy, machine learning, e-learning technologies, software engineering, cloud computing, and affective computing.

Dr. Qu has served as the General Chair, the Program Chair, and a Keynote Speaker at many IEEE, ACM, ASIS, and IFIP international conferences or workshops. He is also an editorial board member of several professional, peer-reviewed *Computer Science or Information Technology Journals*.