A Comprehensive Literature Review with Self-Reflection

Literature Review

October 5, 2025

Abstract

This literature review provides a comprehensive analysis of recent research in the field. The review synthesizes findings from 377 research papers, identifying key themes, methodological approaches, and future research directions.

Contents

1 Introduction

The effective representation and utilization of knowledge have long been central to the advancement of artificial intelligence. Knowledge Graphs (KGs) have emerged as a powerful paradigm for organizing vast amounts of structured information, representing real-world entities and their intricate relationships in a human-readable, symbolic format. However, despite their expressiveness, traditional symbolic KGs inherently face challenges such as data sparsity, computational inefficiency in large-scale reasoning, and difficulty in capturing nuanced semantic similarities or handling incompleteness. These limitations hinder their seamless integration with data-driven machine learning models.

To overcome these hurdles, Knowledge Graph Embedding (KGE) methods have revolutionized the field by transforming discrete, symbolic knowledge into continuous, low-dimensional vector representations. This paradigm shift enables entities and relations to be represented as points or vectors in a continuous space, where semantic relationships are captured through geometric or algebraic operations. Such embeddings provide a machine-understandable format, facilitating scalability, efficient computation, and the ability to infer latent relationships, thereby making KGs more actionable for a wide array of AI tasks, including link prediction, entity alignment, and question answering.

This introductory section lays the foundational context for understanding the significance and trajectory of KGE research. We begin by tracing the evolution of knowledge representation, introducing the fundamental concepts and inherent challenges of knowledge graphs. Subsequently, we delve into the core motivations driving the development of KGE methods, explaining how they address the limitations of symbolic representations and enable new capabilities for AI systems. Finally, this section delineates the comprehensive scope and organizational structure of this literature review, providing a roadmap for the detailed discussions that follow on foundational models, advanced architectures, practical considerations, and diverse applications of KGE.

1.1 Background: Knowledge Graphs

Knowledge Graphs (KGs) have emerged as a pivotal technology for organizing and representing real-world knowledge in a structured, machine-readable format. Fundamentally, a knowledge graph is a directed graph composed of entities (nodes) and relations (edges) that connect them (104; 48). Each piece of information within a KG is typically represented as a triple (head entity, relation, tail entity), often denoted as (h, r, t). For instance, the triple (Barack Obama, bornIn, Hawaii) explicitly states that "Barack Obama was born in Hawaii." This structured representation allows for unambiguous storage and retrieval of facts, forming a rich network of interconnected information (18).

The conceptual roots of knowledge graphs can be traced back to early efforts in Artificial Intelligence to represent knowledge symbolically. Semantic networks, developed in the 1960s and 1970s, were among the first attempts to model knowledge as a graph of concepts and relationships. These early networks, while foundational, often suffered from ambiguity, a lack of formal semantics, and limited scalability, making complex reasoning difficult. The subsequent rise of expert systems in the 1980s, which relied on explicit rules and symbolic logic, further highlighted the challenges of managing large, intricate knowledge bases.

The advent of the Semantic Web in the early 2000s marked a significant step towards formalizing knowledge representation on a global scale. Technologies like Resource Description Framework (RDF) and Web Ontology Language (OWL) provided standardized frameworks for defining ontologies and expressing knowledge with formal semantics, aiming to make web content machine-understandable. This era saw the emergence of early large-scale KGs like DBpedia, which systematically extracted structured information from Wikipedia infoboxes, effectively bridging the gap between unstructured text and structured knowledge (48). DBpedia's success demonstrated the potential of automatically constructing KGs from existing data sources, laying the groundwork for more ambitious projects.

Modern knowledge graphs, such as Freebase and Wikidata, represent the culmination of these historical developments, leveraging massive datasets and advanced computational

techniques. Freebase, a collaborative knowledge base initially developed by Metaweb and later acquired by Google, played a crucial role in organizing a vast array of general-purpose facts, significantly contributing to Google's own proprietary Knowledge Graph that powers its search engine (104). Wikidata, an open and collaborative knowledge base maintained by the Wikimedia Foundation, has become a central hub for structured data across various Wikimedia projects and beyond. Its multilingual nature and community-driven approach underscore its role in democratizing access to structured world knowledge and enabling a wide range of intelligent systems, from virtual assistants to sophisticated data analytics platforms. These modern KGs are characterized by their immense scale, dynamic nature, and their ability to integrate information from diverse sources, making them indispensable resources for contemporary AI applications.

Despite their power in organizing information, symbolic knowledge graphs inherently face several significant challenges that limit their utility, particularly in the context of modern data-driven AI.

Computational Inefficiency and Scalability: Reasoning directly with symbolic representations, especially on large-scale KGs containing billions of facts, is computationally intensive. Tasks like inferring new facts through logical rules or answering complex queries often involve combinatorial search problems, leading to prohibitive computational costs and poor scalability (104; 48). For instance, traditional rule-based inference engines struggle with the sheer volume of facts in KGs like Wikidata, where deriving multi-hop relations can quickly become intractable. This inefficiency makes real-time applications challenging and limits the depth of reasoning that can be performed.

Sparsity and Incompleteness: Real-world knowledge graphs are notoriously incomplete. Even the largest KGs, such as Freebase or DBpedia, contain only a fraction of all possible facts about entities and relations. Symbolic methods struggle to generalize from observed facts to infer missing ones, often requiring explicit rules or complete data to make deductions. This "cold-start" problem, where new entities or relations lack sufficient explicit connections, is a major bottleneck (48). For example, if a KG lacks the explicit triple (Albert Einstein, studiedAt, ETH Zurich), a purely symbolic system might

fail to infer this, even if it has related facts. This limitation contrasts sharply with human ability to infer and generalize from partial information.

Difficulty in Capturing Nuanced Semantics and Similarity: Symbolic representations are discrete and rigid. They treat each entity and relation as a distinct, atomic symbol, making it difficult to capture subtle semantic similarities or relationships between them. For instance, "car" and "automobile" are semantically very close, but a symbolic KG would treat them as entirely separate entities unless explicitly linked by an equivalence relation. Similarly, understanding that "fatherOf" is the inverse of "childOf" requires explicit logical rules, rather than being inherently captured by the representation itself. This rigidity hinders tasks that rely on semantic proximity, such as recommendation systems or natural language understanding, where nuanced meanings are paramount.

Incompatibility with Machine Learning: Perhaps the most critical limitation in the era of deep learning is the inherent incompatibility of symbolic representations with numerical machine learning models. Modern AI algorithms, particularly neural networks, operate on continuous vector spaces. Directly feeding discrete symbols into these models is inefficient and often requires extensive feature engineering, which is labor-intensive and prone to error. This disconnect prevents KGs from being seamlessly integrated into powerful, end-to-end machine learning pipelines.

These collective challenges of symbolic KGs—computational inefficiency, pervasive incompleteness, difficulty in capturing nuanced semantics, and incompatibility with modern machine learning paradigms—collectively motivate the need for more advanced representation techniques. This is precisely where Knowledge Graph Embedding (KGE) emerges as a transformative solution. KGE aims to address these limitations by transforming entities and relations from their sparse, discrete symbolic forms into dense, continuous, low-dimensional vector representations, often referred to as embeddings, in a continuous vector space (18; 104). By embedding knowledge into a continuous space, KGE models enable several critical advantages: capturing semantic similarity, enhancing computational efficiency, handling incompleteness through generalization, and facilitating seamless integration with machine learning models. This paradigm shift from explicit symbolic reason-

ing to implicit, distributed representations in vector spaces has unlocked the full potential of knowledge graphs, making them more accessible, actionable, and powerful for a diverse range of intelligent systems.

1.2   Motivation for Knowledge Graph Embedding

Building upon the discussion in the previous subsection regarding the inherent limitations of symbolic Knowledge Graphs (KGs), the core motivation for Knowledge Graph Embedding (KGE) emerges as a transformative paradigm shift in knowledge representation. Traditional symbolic KGs, while powerful for explicit fact storage, face significant hurdles that impede their utility in modern AI applications. These limitations primarily stem from their sparse, discrete nature, which struggles with nuanced semantics, computational efficiency, and handling incompleteness (48; 18). KGE directly addresses these challenges by projecting entities and relations into continuous, low-dimensional vector spaces, thereby converting complex symbolic problems into efficient vector operations and making KGs more accessible and actionable for AI (48).

One of the most critical limitations of symbolic representations is their **inability to capture nuanced semantic similarities**. Each entity and relation is treated as an atomic, distinct symbol, meaning that semantic proximity (e.g., "car" and "automobile" being synonyms) is not inherently encoded unless explicitly defined by an equivalence relation. This rigidity hinders tasks that rely on understanding subtle semantic relationships, such as recommendation systems, natural language understanding, or information retrieval, where a query might use different but semantically similar terms. KGE overcomes this by representing entities and relations as dense vectors, where semantic similarity is naturally captured by vector proximity in the embedding space. For instance, entities with similar meanings or roles will have vectors that are close to each other, allowing for generalization and inference based on continuous semantic gradients rather than discrete matches (18).

Furthermore, **computational inefficiency in large-scale reasoning** is a major bottleneck for symbolic KGs. Reasoning over billions of facts using traditional rule-based

or logical inference engines often involves combinatorial search, leading to prohibitive computational costs and poor scalability. This makes real-time applications, such as dynamic question answering or complex multi-hop inference, extremely challenging. KGE models, by contrast, transform these symbolic operations into efficient vector arithmetic. For example, a relation can be modeled as a translation operation in the embedding space, where (h, r, t) is plausible if the vector of h plus the vector of r is close to the vector of t (as in TransE). This converts a search over discrete symbols into a distance calculation in a continuous space, which is highly optimized for modern hardware and machine learning frameworks, significantly boosting computational efficiency and scalability for large KGs (48).

The pervasive **difficulty in handling incompleteness** is another fundamental problem that KGE aims to solve. Real-world KGs are inherently incomplete, containing only a fraction of all possible facts. Symbolic methods struggle to generalize from observed facts to infer missing ones, often requiring explicit rules or complete data to make deductions. This "cold-start" problem for new entities or relations is a significant barrier. KGE, however, leverages patterns learned from existing facts in the embedding space to predict missing links. By observing that certain entity pairs are connected by specific relation vectors, the model can infer the likelihood of unobserved triples. For example, early translational models like TransH (60) and TransD (78) were specifically designed to better handle complex relation types (e.g., one-to-many, many-to-one) that TransE struggled with, thereby improving the ability to infer missing facts in diverse scenarios. This generalization capability is crucial for knowledge graph completion, a primary application of KGE (108).

The paradigm shift to continuous vector representations also facilitates **seamless integration with modern machine learning pipelines**. Modern AI algorithms, particularly deep neural networks, are designed to operate on continuous numerical inputs. Symbolic representations are inherently incompatible, requiring extensive and often lossy feature engineering to be used with these models. KGE provides a natural bridge, transforming discrete symbolic knowledge into a format directly consumable by advanced ma-

chine learning models. This integration allows KGs to serve as rich, structured knowledge sources for tasks that benefit from both symbolic structure and statistical learning, such as natural language processing, computer vision, and recommendation systems.

This transformation into dense vector spaces provides a powerful mechanism for a multitude of downstream AI tasks:

- **Link Prediction**: KGEs are fundamentally designed for link prediction, where the goal is to infer missing relations between existing entities or to identify new entities for a given relation. Models like RotatE (77) exemplify this, defining relations as rotations in complex vector spaces to effectively model and infer complex relation patterns like symmetry, antisymmetry, inversion, and composition, significantly outperforming earlier translational models.

- **Entity Alignment**: This task involves identifying equivalent entities across different, often heterogeneous, KGs. By embedding entities from multiple KGs into a shared vector space, entity alignment becomes a problem of finding nearest neighbors in that space. Approaches like bootstrapping entity alignment (77) and multi-view KGE for entity alignment (43) leverage embedding similarities to iteratively discover and refine alignments, demonstrating superior performance over symbolic matching methods.

- **Question Answering (QA)**: KGEs enhance QA systems by enabling the conversion of natural language queries and KG facts into a common embedding space, facilitating efficient matching and retrieval of answers. The Knowledge Embedding based Question Answering (KEQA) framework (37) jointly recovers entity, predicate, and tail entity representations in the KG embedding spaces, allowing for direct matching with KG facts. More sophisticated systems, such as Marie and BERT for chemistry QA (16), integrate hybrid KGEs with advanced NLP models to handle complex queries and deep ontologies.

- **Recommendation Systems**: KGEs can model user-item interactions and item properties within a KG, providing rich contextual information for personalized rec-

ommendations. Recurrent KGE (RKGE) (77) uses a recurrent network to learn semantic representations of paths between entities, characterizing user preferences and even providing meaningful explanations for recommendations. Similarly, contextualized KGE (CKGE) (4) integrates neighbor semantics and high-order connections for explainable talent training course recommendations.

In essence, the motivation for KGE is to overcome the inherent limitations of sparse, symbolic knowledge representation by providing a scalable, semantically rich, and computationally efficient alternative. Early work, such as TransH (60), began to address the limitations of simpler models by introducing hyperplanes to better capture diverse relation properties, a step towards more expressive embeddings. This evolution from discrete symbols to continuous vectors represents a fundamental paradigm shift, making KGs not just repositories of facts, but dynamic, actionable components within modern AI systems. While early KGE models sometimes faced a trade-off between expressiveness and computational efficiency, the field has continuously evolved, with models like RotatE (77) demonstrating how sophisticated geometric operations can capture complex logical patterns while maintaining scalability. The collective efforts in KGE research have thus transformed KGs from static, hard-to-manage data structures into versatile, machine-understandable knowledge sources, crucial for the advancement of artificial intelligence.

1.3   Scope and Structure of the Review

Building upon the preceding discussion regarding the fundamental motivations for Knowledge Graph Embedding (KGE), this comprehensive literature review systematically delineates the intellectual landscape of the field, tracing its evolution from foundational theoretical underpinnings to cutting-edge advancements and diverse real-world applications. The scope of this review is intentionally broad, encompassing the primary paradigms and significant innovations that have shaped KGE research. We begin by examining the core conceptual models that first enabled the representation of entities and relations in continuous vector spaces, subsequently progressing to advanced architectural designs, critical practical considerations, and the wide array of applications that leverage KGE for

enhanced artificial intelligence capabilities. This pedagogical progression ensures a coherent narrative, illustrating how the field has continuously evolved to address the inherent complexities and limitations of knowledge representation.

The review's comprehensive scope is structured to guide the reader through a logical progression of KGE research. Initially, we delve into the **foundational models** that established the geometric and algebraic paradigms for embedding knowledge graphs. Early translational models, such as TransH (60) and TransD (78), are critically analyzed for their core innovations in handling complex relational patterns like one-to-many and many-to-one, and their trade-offs between expressiveness and computational efficiency. These models, while efficient, often struggled with intricate relational semantics, a limitation that spurred the development of more sophisticated geometric approaches. This initial phase of research, as highlighted by surveys like (48) and (18), laid the groundwork by converting symbolic knowledge into machine-understandable vector operations, addressing the sparsity and computational inefficiency inherent in traditional KGs.

The narrative then progresses to **advanced architectures**, particularly those leveraging deep learning, which represent a significant paradigm shift in KGE. This includes the integration of Convolutional Neural Networks (CNNs), Graph Neural Networks (GNNs), and Transformer models. These architectures, unlike their geometric predecessors, excel at automatically learning complex, non-linear feature interactions and capturing intricate structural patterns directly from the graph topology. For instance, while RotatE (77) introduced rotations in complex space to model symmetry and inversion more effectively than translational models, deep learning models offer even greater capacity for capturing hierarchical and contextual relationships. This evolution reflects the field's response to the need for more expressive and context-aware representations, moving beyond simple triplet-based interactions to leverage broader graph structures.

Beyond core model development, the review critically examines **practical considerations** essential for the real-world deployment of KGE. This includes strategies for improving computational efficiency, managing memory footprint, and ensuring scalability for increasingly massive knowledge graphs. The field has seen a continuous tension

between model expressiveness and practical feasibility; highly expressive models often incur significant computational costs, limiting their applicability to large-scale KGs. This section also addresses methods for enhancing model robustness against noisy data, optimizing training processes, and the crucial aspects of rigorous evaluation, benchmarking, and reproducibility. As noted by (108), inconsistent reporting practices and evaluation biases can obscure true model performance, underscoring the importance of standardized methodologies. For example, methods like HyTE (2), while innovative in incorporating temporal dynamics, introduce additional complexity that must be managed for scalability.

Finally, the review explores the **diverse applications** of KGE, demonstrating their transformative impact across various AI domains. This includes fundamental tasks like link prediction and knowledge graph completion, where KGE models infer missing facts by leveraging learned patterns. The utility extends to more complex applications such as entity alignment across heterogeneous KGs, exemplified by bootstrapping approaches (77) and multi-view frameworks (43) that leverage embedding similarities. Furthermore, KGEs have significantly enhanced question answering systems, with frameworks like KEQA (37) demonstrating how jointly learned entity and predicate representations can improve answer retrieval. In recommender systems, recurrent KGEs (RKGE) (77) have been shown to effectively characterize user preferences and provide explainable recommendations. These applications collectively illustrate the practical value of KGE in making knowledge graphs actionable and intelligent components within modern AI systems.

The organizational structure of this review is designed to provide a clear roadmap through this complex landscape:

1. **Section 1: Introduction** establishes the foundational context, motivations, and the overarching scope and structure of the review.

2. **Section 2: Foundational KGE Models and Geometric Paradigms** delves into the theoretical underpinnings, examining early translational models (e.g., TransH (60), TransD (78)) and more advanced geometric approaches like RotatE (77). This section highlights the initial breakthroughs in representing symbolic knowledge in continuous spaces and their inherent capabilities and limitations.

14

3. **Section 3: Deep Learning Architectures for Knowledge Graph Embedding** explores the integration of advanced neural network architectures (CNNs, GNNs, Transformers), which represent a shift towards learning more expressive and context-aware representations, often overcoming the limitations of purely geometric models in capturing complex, non-linear patterns.

4. **Section 4: Enriching KGE: Auxiliary Information, Rules, and Multi-modality** discusses methods that augment KGE with external knowledge, such as entity types, attributes, logical rules, and multi-modal data. This addresses the challenge of data sparsity and enhances semantic understanding, moving beyond purely structural information.

5. **Section 5: Dynamic, Inductive, and Distributed KGE** focuses on adapting KGE models to the evolving, incomplete, and decentralized nature of real-world knowledge graphs, covering temporal KGE (e.g., HyTE (2)), inductive learning for unseen entities, and privacy-preserving federated approaches. This section addresses the critical need for models that can operate in dynamic, real-world environments.

6. **Section 6: Practical Considerations: Efficiency, Robustness, and Evaluation** addresses the crucial aspects of deploying KGE models, including scalability, robustness against noise, and the importance of rigorous evaluation and reproducibility in research, as emphasized by comparative analyses like (108).

7. **Section 7: Applications and Real-World Impact of KGE** showcases the practical utility of KGE across diverse domains, including link prediction, entity alignment (77; 43), question answering (37), and recommender systems (77). This section bridges the gap between theoretical advancements and tangible societal benefits.

8. **Section 8: Conclusion and Future Directions** synthesizes the key findings, identifies persistent open challenges, and outlines emerging trends and ethical considerations that will shape the future trajectory of KGE research, drawing insights from comprehensive surveys (48; 18).

This structured approach ensures that readers gain a holistic understanding of KGE, from its theoretical underpinnings to its real-world impact and future potential, providing a comprehensive roadmap for navigating this rapidly evolving field.

2 Foundational KGE Models and Geometric Paradigms

This section marks a pivotal point in our review, transitioning from the foundational motivations for Knowledge Graph Embedding (KGE) to the initial, groundbreaking methodologies that transformed symbolic knowledge into continuous vector representations. Building upon the understanding that traditional Knowledge Graphs (KGs) struggle with sparsity, computational inefficiency, and capturing nuanced semantics, this section delves into the bedrock of KGE research: the foundational geometric and algebraic paradigms. These early and influential models laid the groundwork for representing entities and relations in continuous vector spaces, establishing the core principles for how relational patterns could be encoded and inferred.

The evolution began with simple yet powerful translational models, such as TransE and its crucial extensions like TransH (60) and TransD (78). These models conceptualized relations as direct translations between head and tail entities in an embedding space, offering significant advancements in efficiency and the ability to infer missing links. However, their limitations in capturing complex relational patterns, such as symmetry, antisymmetry, and composition, quickly became apparent. This challenge spurred the development of more sophisticated geometric approaches.

Consequently, the field progressed to models that leveraged rotations and complex-valued spaces, exemplified by RotatE (77). These innovations allowed for the elegant modeling of intricate logical patterns through algebraic transformations, significantly enhancing expressiveness beyond simple translations. Further explorations extended to multi-dimensional geometric approaches, including embeddings on Lie groups, different metric choices, and advanced transformations like Householder parameterization. This continuous refinement of mathematical foundations, exploring diverse embedding spaces and transformation mechanisms, was driven by the imperative to capture richer semantic

interactions and improve the theoretical soundness and robustness of KGE models. Ultimately, this section traces the critical journey from basic vector operations to complex geometric transformations, forming the theoretical and practical basis for all subsequent advancements in KGE.

2.1 Core Translational Models and Extensions

The foundational era of Knowledge Graph Embedding (KGE) research was significantly shaped by a family of models that conceptualized relations as translational operations within a continuous vector space. These pioneering translational models, notably TransE, TransH, TransR, and TransD, established a fundamental paradigm for representing symbolic knowledge in a machine-understandable format. They offered enhanced efficiency and expressiveness over purely symbolic methods, marking a crucial step in transforming knowledge graph analysis and laying the groundwork for subsequent geometric and deep learning advancements (116).

The seminal work, **TransE** (Translating Embeddings) (**?**), introduced the intuitive idea that if a triple (h, r, t) exists in a knowledge graph, then the embedding of the head entity ($\mathbf{h}$) plus the embedding of the relation ($\mathbf{r}$) should approximately equal the embedding of the tail entity ($\mathbf{t}$), i.e., $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$. This simple vector addition in Euclidean space, typically minimized using a scoring function like $f(h, r, t) = ||\mathbf{h} + \mathbf{r} - \mathbf{t}||\_L1/L2$, offered remarkable efficiency and a straightforward mechanism for link prediction. However, TransE suffered from a critical theoretical limitation: its inherent inability to effectively model complex relation patterns such as one-to-many, many-to-one, and many-to-many relations. This limitation arises because TransE represents each entity and relation as a single point or vector in the same embedding space. For instance, if a relation (`country, hasCapital, city`) is one-to-many (e.g., (`France, hasCapital, Paris`) and (`France, hasCapital, Lyon`) if considering historical capitals), TransE would force the embeddings of `Paris` and `Lyon` to be very close to each other relative to `France` and `hasCapital`, which is semantically incorrect and leads to a "blurring" of entity representations. This issue, as highlighted by (60), motivated the development of more sophisticated translational

models.

Building directly on TransE's insights, (60) introduced **TransH** (Translating on Hyperplanes) to address the challenge of modeling one-to-many and many-to-one relations more effectively. The core innovation of TransH lies in representing each relation not as a single vector in the entity embedding space, but as a hyperplane ($\mathbf{w}\_r$) along with a translation vector ($\mathbf{r}$) on that hyperplane. For a given triple (h, r, t), TransH first projects the head and tail entity embeddings onto the relation-specific hyperplane: $\mathbf{h}\_\perp = \mathbf{h} - \mathbf{w}\_r^\top \mathbf{h}\mathbf{w}\_r$ and $\mathbf{t}\_\perp = \mathbf{t} - \mathbf{w}\_r^\top \mathbf{t}\mathbf{w}\_r$. The translational assumption then applies to these projected embeddings: $||\mathbf{h}\_\perp + \mathbf{r} - \mathbf{t}\_\perp||\_{L1/L2}$. Unlike TransE, which uses a single embedding for each entity regardless of the relation, TransH allows a single entity to have different representations (projections) when involved in different relations. This mechanism enables the model to handle multiple tail entities for a single head (one-to-many) or multiple head entities for a single tail (many-to-one) without forcing their original embeddings to be identical. TransH successfully claims to solve the problem of preserving diverse mapping properties of relations with almost the same model complexity as TransE, achieving significant improvements in predictive accuracy on benchmark datasets like WordNet and Freebase (60). Its practical limitation is a slight increase in parameters compared to TransE due to the additional hyperplane normal vector for each relation, but it maintains competitive scalability.

Following TransH, models like **TransR** (?) and CTransR further explored the idea of relation-specific transformations by introducing relation-specific projection matrices. These models generally map entities from the entity embedding space into a relation-specific space *before* applying the translation. For instance, TransR projects $\mathbf{h}$ and $\mathbf{t}$ into a relation-specific space using a matrix $\mathbf{M}\_r$, then applies the translation: $||\mathbf{h}\mathbf{M}\_r + \mathbf{r} - \mathbf{t}\mathbf{M}\_r||\_{L1/L2}$. This approach aimed to capture more fine-grained relational semantics than TransH's hyperplane projection. However, a significant practical limitation of TransR/CTransR was the high parameter count. Each relation required its own dense projection matrix $\mathbf{M}\_r$, leading to a parameter complexity of $O(N\_r \cdot d\_e \cdot d\_r)$ where N_r is the number of relations, d_e is entity embedding

dimension, and d_r is relation embedding dimension. This quadratic growth in parameters with respect to embedding dimensions hindered scalability for large knowledge graphs, making them less practical for real-world applications with many relations.

This parameter explosion limitation was directly addressed by (78) with the introduction of **TransD** (Knowledge Graph Embedding via Dynamic Mapping Matrix). TransD claims to solve the problem of high parameter count in TransR/CTransR while considering the diversity of both relations and entities. Its core innovation is the use of dynamic mapping matrices. Instead of a static, dense projection matrix for each relation, TransD assigns two vectors to each entity $(\mathbf{e}, \mathbf{e}\_p)$ and relation $(\mathbf{r}, \mathbf{r}\_p)$. The first vector represents the entity's or relation's meaning, while the second is used to dynamically construct a mapping matrix for projection. Specifically, the projection of an entity $\mathbf{h}$ for relation $\mathbf{r}$ is given by $\mathbf{h}\_\perp r = \mathbf{h} + \mathbf{h}\_p \mathbf{r}\_p^\top \mathbf{h}$, and similarly for $\mathbf{t}$. The scoring function then follows the translational principle: $\|\mathbf{h}\_\perp r + \mathbf{r} - \mathbf{t}\_\perp r\|\_{L1/L2}$. In contrast to TransR's fixed, dense matrices, TransD's dynamic, entity-specific projection mechanism allows for more fine-grained transformations while significantly reducing the number of parameters to $O(N\_e \cdot d\_e + N\_r \cdot d\_r)$, as it avoids storing large, dense projection matrices. This makes TransD a more scalable solution for large-scale graphs, demonstrating superior performance on triplet classification and link prediction tasks (78). Its practical advantage lies in its parameter efficiency and absence of computationally intensive matrix-vector multiplication operations during projection, making it more applicable to large KGs. However, like other translational models, its theoretical limitations persist in fully capturing highly complex logical patterns such as composition or inversion, which later rotational models would specifically target.

Further extensions to the translational paradigm include models like **TransE-MTP** (TransE with Multi-Translation Principles) (**?**). This model directly addresses TransE's struggle with complex relations by proposing *multiple translation principles* tailored for different relation types (e.g., one-to-one, one-to-many, many-to-one, many-to-many). Unlike TransH and TransD, which modify the projection mechanism, TransE-MTP modifies the core translational rule itself based on the identified relation type. It combines these

MTPs with the original TransE framework, allowing for multiple optimization objectives during training on complex relations. This approach claims to provide superior prediction performance over both TransE and TransH, demonstrating its effectiveness at link prediction and triplet classification on Freebase and WordNet datasets (**?**). This illustrates a pattern in translational model evolution: the initial simplicity of TransE led to subsequent models either introducing relation-specific projections (TransH, TransR, TransD) or adapting the translational principle itself (TransE-MTP) to better handle the diversity of real-world relational patterns.

In summary, the evolution of core translational models showcases a continuous effort to balance efficiency with expressiveness. Early work like TransE established the foundational idea of relations as translations, prioritizing computational efficiency. Mid-period models such as TransH and TransD systematically addressed TransE's limitations regarding complex relation patterns (one-to-many, many-to-one) by introducing relation-specific projections, first via hyperplanes and then through dynamic, parameter-efficient mapping matrices. This progression from TransE to TransH and then to TransD (116) illustrates a clear causal relationship: the limitations of simpler models (e.g., TransE's inability to distinguish multiple tail entities for a single head) directly led to innovations in subsequent models (e.g., TransH's hyperplanes, TransD's dynamic matrices) to overcome these specific challenges. TransR, while innovative in its use of projection matrices, highlighted the trade-off between expressiveness and parameter count, a challenge that TransD successfully mitigated. The introduction of TransE-MTP further refined this by adapting the translational principle itself to relation types. This intellectual trajectory exemplifies the field's shift from basic vector operations to more nuanced geometric transformations, all while striving for models that are both effective and scalable for real-world knowledge graphs.

2.2   Rotational and Complex Space Embeddings

Building upon the foundational translational models discussed previously, which primarily focused on vector addition in real Euclidean spaces, a significant paradigm shift in

Knowledge Graph Embedding (KGE) research emerged with the introduction of rotational and complex space embeddings. These models address the inherent limitations of translational approaches, particularly their struggle to capture richer relational semantics such as symmetry, antisymmetry, inversion, and composition (77). Translational models like TransE (60) inherently assume that relations are simple displacements, making it difficult to distinguish between $h \xrightarrow{r} t$ and $t \xrightarrow{r} h$ for symmetric relations, or to model $h \xrightarrow{r\_1} i \xrightarrow{r\_2} t$ as equivalent to $h \xrightarrow{r\_3} t$ for compositional relations. Rotational models, by contrast, leverage the algebraic properties of complex numbers and higher-dimensional rotations to offer more nuanced and powerful transformations.

The seminal work in this area is **RotatE** (Knowledge Graph Embedding by Relational Rotation in Complex Space) (77). RotatE claims to solve the problem of effectively modeling and inferring various complex relation patterns, including symmetry ($r \iff r^{-1}$), antisymmetry ($r \implies \neg r^{-1}$), inversion ($r \iff r'^{-1}$), and composition ($r\_1 \circ r\_2 \implies r\_3$). Its core innovation lies in defining each entity as a vector in a complex vector space ($\mathbb{C}^k$) and each relation as an element-wise rotation from the head entity to the tail entity. Specifically, for a triple (h, r, t), RotatE models the relation as $\mathbf{h} \circ \mathbf{r} \approx \mathbf{t}$, where $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$ and $\circ$ denotes the Hadamard (element-wise) product. The relation vector $\mathbf{r}$ is constrained to have moduli of 1 for each dimension (i.e., $|r\_i| = 1$), ensuring it acts as a pure rotation. The scoring function is typically $f(h, r, t) = ||\mathbf{h} \circ \mathbf{r} - \mathbf{t}||\_L1/L2$. This elegant formulation inherently captures the desired patterns:

- **Symmetry**: If r is symmetric, then $\mathbf{r}$ should be close to a vector of all 1s (no rotation), or $\mathbf{r}^2 \approx \mathbf{1}$.

- **Antisymmetry**: If r is antisymmetric, then $\mathbf{r}$ should be far from a vector of all 1s.

- **Inversion**: If r' is the inverse of r, then $\mathbf{r}' \approx \mathbf{r}^{-1}$ (element-wise conjugate).

- **Composition**: If $r\_1 \circ r\_2 \implies r\_3$, then $\mathbf{r}\_1 \circ \mathbf{r}\_2 \approx \mathbf{r}\_3$.

RotatE succeeds under conditions where KGs exhibit a rich variety of these logical patterns. Its theoretical advantage stems from the algebraic properties of complex numbers,

which naturally encode rotations and their compositions. Empirically, RotatE significantly outperformed existing state-of-the-art models for link prediction on benchmark datasets like FB15k-237 and WN18RR, demonstrating its effectiveness (77). A practical limitation of RotatE, while more expressive than TransE, is its increased computational cost due to complex number arithmetic, and it might still struggle with very high-dimensional or extremely sparse relations where the "pure rotation" assumption might be too restrictive.

The success of RotatE spurred further research into higher-dimensional rotational embeddings. **Rotate3D** (64) extends the concept of relations as rotations from 2D complex space to 3D Euclidean space. This model maps entities to 3D vectors and defines relations as 3D rotations, leveraging the non-commutative property of 3D rotations. This non-commutativity is particularly beneficial for modeling non-commutative composition patterns, which are essential for multi-hop reasoning where the order of relations matters. Rotate3D claims to solve the problem of capturing non-commutative composition, which RotatE, despite its compositional capabilities, might not fully exploit due to the element-wise nature of its complex rotations. The core innovation is using rotation matrices in 3D space, where $\mathbf{R}\_r \in SO(3)$ is a 3D rotation matrix for relation r. The scoring function is $f(h, r, t) = ||\mathbf{R}\_r\mathbf{h} - \mathbf{t}||\_L1/L2$. Rotate3D demonstrated improved performance on link prediction and path query answering, especially for multi-hop reasoning tasks, compared to RotatE (64). Its theoretical limitation is that 3D rotations, while powerful, are still a specific type of transformation and may not capture all possible complex relational dynamics. Practically, managing 3D rotation matrices might introduce more parameters or computational complexity than element-wise complex rotations, though it offers a more direct geometric interpretation for certain patterns.

Moving beyond 3D Euclidean space, the use of quaternions offers an even richer algebraic structure for modeling rotations and addressing challenges like polysemy. **ConQuatE** (Contextualized Quaternion Embedding Towards Polysemy in Knowledge Graph for Link Prediction) (46) is a recent model that leverages quaternion rotations. Quaternions, which extend complex numbers to four dimensions (one real, three imaginary com-

ponents), can represent 3D rotations more compactly and efficiently than 3x3 matrices. ConQuatE claims to address the polysemy issue in KGs, where entities can exhibit different semantic characteristics depending on the relations they participate in. The core innovation is to incorporate contextual cues from various connected relations to enrich entity representations, which are then transformed using quaternion rotations. For a triple (h, r, t), entities and relations are embedded as quaternions, and the relation r acts as a quaternion rotation from h to t. The multiple imaginary components of quaternions allow for a more expressive representation of diverse relational contexts, enabling the model to differentiate between different "senses" of an entity based on the relation. ConQuatE's theoretical advantage is its ability to capture more intricate interactions and context-dependent semantics due to the higher-dimensional algebraic properties of quaternions. Empirically, it outperforms state-of-the-art models for link prediction on datasets like FB15k-237 and WN18RR, particularly in scenarios where polysemy is prevalent (46). A practical limitation is the increased complexity of quaternion arithmetic, which can be computationally more demanding than complex number operations.

Table 1: Comparative Overview of Rotational and Complex Space KGE Models

|l|X|X|X|X|

Model	Embedding Space	Core Operation	Key Patterns/Problems Addressed	Theoretical/Practical Considerations
RotatE (77)	Complex Vector Space ($\mathbb{C}^k$)	Element-wise Rotation (Hadamard product)	Symmetry, Antisymmetry, Inversion, Composition	Elegant mathematical formulation, good expressiveness, moderate computational cost.
Rotate3D (64)	3D Euclidean Space	3D Rotation Matrix	Non-commutative Composition, Multi-hop Reasoning	Direct geometric interpretation, higher parameter count for matrices, captures order in composition.
ConQuatE (46)	Quaternion Space	Quaternion Rotation	Polysemy, Diverse Relational Contexts	Richer algebraic structure, compact 3D rotation, higher computational complexity than complex numbers.
HousE (107)	Real Vector Space	Householder Transformations (Rotations & Projections)	General relation patterns, mapping properties, higher-dimensional rotations	Generalizes rotations, strong modeling capacity, increased complexity.
CompoundE (73)	Real Vector Space	Cascaded Translation, Rotation, Scaling	General relation patterns, diverse mapping properties	Unifies multiple operations, high expressiveness, potentially complex optimization.
SpherE (120)	Spherical Space	Rotational (entity as sphere)	Many-to-many relations, Set Retrieval	Novel entity representation, good for set-based queries, interpretability.

The intellectual trajectory in this domain demonstrates a clear progression from sim-

pler translational models to increasingly complex algebraic structures. While translational models like TransH (60) and TransD (78) introduced relation-specific projections to handle one-to-many/many-to-one relations, they fundamentally relied on vector addition, limiting their ability to model patterns that require more sophisticated transformations. RotatE (77) directly addressed this by introducing rotations in complex space, demonstrating that the algebraic properties of complex numbers are inherently better suited for patterns like inversion and composition. This exemplifies the field's shift from a purely geometric interpretation to leveraging richer algebraic structures.

Further innovations in this family include **HousE** (Knowledge Graph Embedding with Householder Parameterization) (107), which proposes a more powerful KGE framework based on Householder transformations. Householder transformations can represent both reflections (a type of rotation) and projections, allowing HousE to achieve superior capacity for modeling relation patterns and handling sophisticated relation mapping properties simultaneously. HousE generalizes existing rotation-based models by extending rotations to high-dimensional real spaces, achieving state-of-the-art performance on several benchmarks (107). Its core innovation is the use of Householder matrices, which are orthogonal and symmetric, providing a theoretically sound way to perform rotations and reflections.

Another approach, **CompoundE** (73) and its 3D extension **CompoundE3D** (104), takes a more generalized view by combining translation, rotation, and scaling operations. These models argue that a cascade of these fundamental geometric manipulations can capture a broader spectrum of relational semantics. CompoundE demonstrates that many existing scoring-function-based KGE models can be seen as special cases of its framework, highlighting a convergent research direction towards unifying different geometric operations. This approach offers high expressiveness but introduces a trade-off with increased model complexity and optimization challenges due to the compound nature of operations.

The concept of modeling transitivity, a specific type of compositional pattern, is addressed by **Rot-Pro** (Modeling Transitivity by Projection in Knowledge Graph Embedding) (71). Rot-Pro theoretically shows that transitive relations can be modeled with projections and combines this insight with relational rotation. It proves that this combination

can infer all common relation patterns, including transitivity, and achieves state-of-the-art results on datasets with transitive relations (71). This work builds on the rotational paradigm by adding a projection component to specifically target transitivity, a pattern that even pure rotational models might not perfectly capture.

More recently, **MQuinE** (106) identifies and resolves a theoretical deficiency, termed "Z-paradox," in some popular KGE models, which can degrade performance, especially on challenging test samples. MQuinE proposes a new KGE model that avoids this paradox while preserving strong expressiveness for various relation patterns, including symmetric/asymmetric, inverse, 1-N/N-1/N-N, and composition relations. This highlights a critical analysis of the theoretical underpinnings of existing models, pushing for more robust and theoretically sound designs. Similarly, **SpherE** (120) extends rotational embeddings by representing entities as spheres instead of vectors, specifically targeting many-to-many relations and enabling set retrieval. This novel entity representation, while inheriting the interpretability of rotational models, offers a unique way to handle set-based queries, a problem often overlooked by traditional KGEs.

The field also sees efforts to generalize orthogonal transformations. **GoldE** (Generalizing Knowledge Graph Embedding with Universal Orthogonal Parameterization) (120) introduces a framework based on a generalized form of Householder reflection. This universal orthogonal parameterization allows for dimensional extension and geometric unification, enabling the framework to simultaneously capture crucial logical patterns and inherent topological heterogeneity. GoldE represents a culmination of efforts to create a unified and highly expressive framework for orthogonal transformations, demonstrating state-of-the-art performance across benchmarks (120).

A critical tension across these models is the trade-off between increased expressiveness and computational complexity. While models like RotatE (77) offer a good balance, extensions to 3D rotations or quaternions, or the use of Householder transformations (107), often come with higher parameter counts or more involved arithmetic operations. For instance, the parameter efficiency of TransD (78) (from the previous subsection) is a practical advantage that more complex rotational models sometimes sacrifice. Fur-

thermore, the assumption that relations can be perfectly modeled by a single type of geometric transformation (e.g., pure rotation) might be an oversimplification for the vast diversity of real-world relations. The evaluation of these models primarily relies on link prediction metrics, which, while standard, may not fully capture the nuances of all the complex patterns they claim to model, especially for tasks like set retrieval (SpherE) or complex logical reasoning. The generalizability of these models to extremely large and sparse KGs, where the computational overhead of complex transformations can become prohibitive, remains an ongoing challenge. Despite these limitations, the shift to rotational and complex/higher-dimensional space embeddings marks a profound advancement, offering more nuanced and powerful transformations that move beyond the limitations of simpler translational approaches, thereby significantly enriching the semantic modeling capabilities of KGEs.

2.3   Other Geometric and Algebraic Innovations

Building upon the foundational translational models and the specific rotational and complex space paradigms discussed in preceding subsections, the field of Knowledge Graph Embedding (KGE) has continuously explored a broader spectrum of geometric and algebraic innovations. This subsection delves into models that challenge conventional Euclidean embedding spaces, introduce novel metric choices, or employ advanced, often compound, transformations beyond simple vector addition or pure rotation. The overarching goal is to refine the mathematical foundations of KGE, enabling more nuanced representations of entities and relations, particularly for capturing complex logical patterns and addressing inherent limitations of simpler geometric models. As highlighted by (18), the choice of representation space and its mathematical properties profoundly influences a model's capacity to capture diverse KG structures, often dictating its ability to model hierarchical, compositional, or context-dependent semantics.

A significant paradigm shift from standard Euclidean vector spaces is observed in models that embed entities on Lie groups or non-Euclidean manifolds, which offer inherent advantages for specific data structures. Unlike Euclidean spaces, which struggle with

certain topological properties like hierarchies or bounded representations, these alternative geometries can naturally accommodate them.

Non-Euclidean Geometries and Manifold Embeddings:

One pioneering work in this direction is **TorusE** (35), which embeds entities on a Lie group, specifically a torus (a compact manifold). TorusE claims to solve the regularization problem inherent in TransE-like models, where explicit regularization (e.g., constraining embeddings to a unit sphere) is often needed to prevent divergence during negative sampling. This regularization can inadvertently warp embeddings in Euclidean space, hindering their ability to accurately fulfill the translational principle ($h + r \approx t$). TorusE's core innovation is that the torus itself is a bounded space, naturally preventing embeddings from diverging while preserving the translational property. Empirically, TorusE demonstrated superior performance over TransE, DistMult, and ComplEx on link prediction tasks, and notably, was faster than the original TransE, showcasing that non-Euclidean spaces can offer both theoretical advantages and practical efficiency (35). However, while it elegantly solves the regularization issue, TorusE, being translation-based, still inherits the expressiveness limitations of TransE for complex relation patterns like symmetry or composition, which rotational models (e.g., RotatE (77) from Section 2.2) are better equipped to handle. Its core assumption of relations as simple translations on a torus, while regularization-friendly, might be too restrictive for the rich semantics found in many KGs, failing to capture the diverse transformations relations can represent.

Further exploring non-Euclidean geometries, hyperbolic spaces have gained traction due to their natural ability to model hierarchical structures prevalent in KGs. Unlike Euclidean space, which expands uniformly, hyperbolic space expands exponentially, allowing for more efficient embedding of tree-like or hierarchical data with fewer dimensions. Early hyperbolic KGE models often adopted a "hybrid" approach, mapping data between hyperbolic and tangent Euclidean spaces for transformations, which introduces computational overhead and numerical instability. **Fully Hyperbolic Rotation (FHRE)** (33) addresses this by defining the KGE model entirely within hyperbolic space, specifically the Lorentz model. FHRE's core innovation is to treat each relation as a Lorentz rotation,

transforming head entity embeddings to tail entity embeddings directly within hyperbolic space, eliminating iterative logarithmic and exponential mappings during training. This direct operation, leveraging the Lorentz rotation theorem, leads to a more stable and efficient model for hierarchical data. FHRE achieved state-of-the-art performance on challenging datasets like CoDEx-s and CoDEx-m, demonstrating its effectiveness and generalization ability, particularly for KGs with pronounced hierarchical patterns (33). A key limitation of FHRE, despite its efficiency, is its reliance on the Lorentz model, which might not be universally optimal for all types of hierarchical structures or logical patterns, and its "pure rotation" assumption might be too restrictive for relations that involve more complex transformations beyond simple hierarchical shifts.

In contrast to FHRE's fully hyperbolic design, **Unified Geometry Knowledge Graph Embedding (UniGE)** (?) takes a different tack by seamlessly integrating KGE in *both* Euclidean and hyperbolic geometric spaces. UniGE proposes an embedding alignment method and fusion strategy using optimal transport techniques and the Wasserstein barycenter method. This hybrid approach aims to capture the diverse geometric data structures (chains and hierarchies) within KGs that might exceed the capacity of a single embedding space. UniGE's theoretical analysis suggests a more robust error bound, and it empirically outperforms state-of-the-art methods on benchmark datasets, indicating that a judicious combination of geometries can be more expressive than relying solely on one (?). The trade-off for UniGE's enhanced expressiveness is the increased complexity of managing and aligning embeddings across different geometric spaces compared to FHRE's simpler, fully hyperbolic operations. While FHRE prioritizes computational stability and directness within a single hyperbolic geometry, UniGE sacrifices some of that simplicity for broader applicability across diverse KG structures.

Another hyperbolic approach, proposed by (118), leverages an extended Poincaré Ball and a polar coordinate system within hyperbolic space to capture both hierarchical structures and logical patterns. This model addresses the boundary conditions of the Poincaré Ball by expanding the modulus length, demonstrating that even within hyperbolic geometry, there are diverse ways to construct embedding spaces, each with its own advantages

for specific KG characteristics. Similarly, **Hierarchical Hyperbolic Neural Graph Embedding (H2E)** (**?**) also employs a hyperbolic polar embedding space with a dual-embedding (modulus and phase parts) to explicitly model inter-level and intra-level hierarchies. H2E further integrates an attentional neural context aggregation to enhance its ability to preserve hierarchical relations. Both (118) and H2E highlight the potential of polar coordinates in hyperbolic space for hierarchical modeling, but their optimization complexity, especially in managing boundary conditions or dual embeddings, can be a practical challenge, potentially leading to slower convergence or sensitivity to hyperparameters.

Further advancing the use of non-Euclidean geometries, **HolmE** (99) introduces a general form of Riemannian KGE, specifically leveraging hyperbolic space, with the novel property of being "closed under composition." This means that the composition of any two relation embeddings remains within the same embedding space, aligning with the theoretical nature of real-world relations and enabling robust modeling of composition patterns regardless of their representation in training data. Unlike prior KGEs that often implicitly assume relations are non-compositional if their patterns are not well-represented, HolmE explicitly designs its relation embedding space to ensure this closure. HolmE extends Möbius addition to product spaces for computational efficiency and theoretically proves that prominent KGE models like TransE (**?**) and RotatE (77) (discussed in Section 2.2) are special cases of its framework, offering a unifying perspective. Empirically, HolmE demonstrates notable advantages in modeling composition patterns, especially for long-tail patterns, and effectively extrapolates to unseen relations (99). However, its focus on composition patterns, while crucial, might mean less emphasis on other complex patterns like symmetry or inversion, which RotatE, for instance, explicitly targets. Its theoretical elegance comes with the inherent complexities of Riemannian geometry, which can be challenging for practitioners.

Beyond the choice of embedding space, the underlying metric used to measure distances and define relationships also profoundly impacts model expressiveness. **CyclE** (114) addresses this by questioning the ubiquitous use of the Minkowski metric in KGE.

The paper argues that existing methods often overlook the implications of the embedding space's metric, leading to suboptimal performance. CyclE's core innovation is the introduction of a "Cycle metric," which is based on the oscillation property of periodic functions. It posits that a smaller function period in this metric leads to better expressive ability. By combining the Cycle metric with popular KGE models, CyclE empirically demonstrated that its proposed metric is more appropriate than the Minkowski metric for KGE, leading to improved expressiveness (114). This work highlights a fundamental, yet often unstated, assumption in KGE research: the choice of distance function. While CyclE offers a novel metric, its success is still contingent on the underlying KGE model's ability to leverage this metric effectively. Its practical limitation lies in the need for careful tuning of the function period, which can be a complex hyperparameter, and the theoretical justification for why a "cycle" metric is universally superior to Minkowski for all KG patterns is not fully explored, potentially limiting its generalizability.

Another innovative approach to geometric representation is exemplified by models that move beyond representing entities as single points. **ManifoldE** (76) proposes a manifold-based embedding principle that replaces the over-strict point-wise translation principle ($h + r \approx t$) with a manifold-wise principle. This model claims to solve the problem of *precise link prediction* by addressing the ill-posed algebraic system and over-rigid geometric forms of traditional KGEs. Instead of mapping true triples to a single point, ManifoldE maps them to a *manifold* (e.g., a high-dimensional sphere or hyperplane) defined by the head entity and relation. This provides a more flexible geometric representation, especially for complex relations like one-to-many or many-to-many. ManifoldE significantly improved HITS@1 scores on FB15K for N-N relations (e.g., 53.0

Similarly, **SpherE** (120) extends rotational embeddings by representing entities as spheres instead of vectors, specifically targeting many-to-many relations and enabling set retrieval. While ManifoldE uses manifolds to define the *target* of a relation, SpherE represents the *entities themselves* as spheres, aiming to capture semantic "spread" rather than a precise target. This novel entity representation, while inheriting the interpretability

of rotational models, offers a unique way to handle set-based queries, a problem often overlooked by traditional KGEs. In a different vein, **TransC** (45) introduces a novel geometric representation for entities by differentiating between concepts and instances. It encodes concepts as spheres and instances as vectors in the same semantic space. This allows for a more intuitive and effective modeling of `instanceOf` and `subClassOf` relations, demonstrating how varying geometric primitives for different entity types can enhance semantic capture. This approach requires explicit type information, which might not always be available in all KGs, limiting its generalizability compared to models like SpherE that learn representations without such external metadata.

Generalized and Compound Transformations:

Beyond these fundamental changes to space, metric, and entity representation, other models introduce specialized algebraic or linear transformations to refine relational modeling, often within conventional Euclidean spaces. The intellectual trajectory in this domain also demonstrates a clear progression towards unifying and generalizing transformations. For instance, **HousE** (Knowledge Graph Embedding with Householder Parameterization) (107) proposes a more powerful KGE framework based on Householder transformations. Householder transformations can represent both reflections (a type of rotation) and projections, allowing HousE to achieve superior capacity for modeling relation patterns and handling sophisticated relation mapping properties simultaneously. HousE generalizes existing rotation-based models by extending rotations to high-dimensional real spaces, achieving state-of-the-art performance on several benchmarks (107). Its core innovation is the use of Householder matrices, which are orthogonal and symmetric, providing a theoretically sound way to perform rotations and reflections. Building on this, **GoldE** (Generalizing Knowledge Graph Embedding with Universal Orthogonal Parameterization) (120) introduces a framework based on a generalized form of Householder reflection. This universal orthogonal parameterization allows for dimensional extension and geometric unification, enabling the framework to simultaneously capture crucial logical patterns and inherent topological heterogeneity. GoldE represents a culmination of efforts to create a unified and highly expressive framework for orthogonal transformations, demonstrating

state-of-the-art performance across benchmarks (120). GoldE's strength lies in its ability to generalize HousE, offering a more comprehensive and flexible approach to orthogonal transformations, but at the cost of increased mathematical abstraction.

Another approach, **CompoundE** (73) and its 3D extension **CompoundE3D** (104), takes a more generalized view by combining translation, rotation, and scaling operations. These models argue that a cascade of these fundamental geometric manipulations can capture a broader spectrum of relational semantics. CompoundE demonstrates that many existing scoring-function-based KGE models can be seen as special cases of its framework, highlighting a convergent research direction towards unifying different geometric operations. This approach offers high expressiveness by explicitly modeling relations as a sequence of multiple transformations, providing greater flexibility than models relying on a single type of transformation like RotatE (Section 2.2). However, this comes with a trade-off of increased model complexity and optimization challenges due to the compound nature of operations and the higher number of parameters to learn, potentially impacting scalability compared to more parameter-efficient models like TransD (78) (Section 2.1).

The concept of modeling transitivity, a specific type of compositional pattern, is addressed by **Rot-Pro** (Modeling Transitivity by Projection in Knowledge Graph Embedding) (71). Rot-Pro theoretically shows that transitive relations can be modeled with projections and combines this insight with relational rotation. It proves that this combination can infer all common relation patterns, including transitivity, and achieves state-of-the-art results on datasets with transitive relations (71). This work builds on the rotational paradigm by adding a projection component to specifically target transitivity, a pattern that even pure rotational models might not perfectly capture. This highlights a pattern of augmenting core geometric operations with additional transformations to target specific complex patterns, demonstrating that a combination of transformations can be more effective than a single one for certain logical properties.

Novel Metrics and Refined Algebraic Models (within Euclidean/Real space):

More recently, **MQuinE** (106) identifies and resolves a theoretical deficiency, termed "Z-paradox," in some popular KGE models, which can degrade performance, especially on

challenging test samples. MQuinE proposes a new KGE model that avoids this paradox while preserving strong expressiveness for various relation patterns, including symmetric/asymmetric, inverse, 1-N/N-1/N-N, and composition relations. This highlights a critical analysis of the theoretical underpinnings of existing models, pushing for more robust and theoretically sound designs. MQuinE's contribution is less about a novel geometric space or transformation, and more about refining the underlying mathematical formulation to avoid pitfalls, making it a general algebraic innovation that enhances reliability.

Finally, some models focus on refining transformations within the conventional Euclidean space, often prioritizing efficiency or specific relational nuances:

- **LineaRE** (110): This model simplifies KGE to a linear regression task, modeling relations as a linear function of two low-dimensional entity vectors with weight and bias vectors. Despite its simplicity, LineaRE claims to cover four connectivity patterns (symmetry, antisymmetry, inversion, composition) and four mapping properties (1-to-1, 1-to-N, N-to-1, N-to-N), demonstrating that powerful expressiveness can sometimes be achieved with straightforward linear algebra. Its practical advantage is scalability to large KGs due to its linear nature, making it a strong contender for efficiency-critical applications. However, its linearity might limit its ability to capture highly non-linear semantic relationships compared to models like CompoundE (73) or those leveraging deep learning (Section 3), which can model more intricate interactions.

- **TransMS** (26): An extension of translation-based models, TransMS introduces "multidirectional semantics" by translating and transmitting semantics from head/tail entities and relations using non-linear functions and linear bias vectors. This moves beyond simple vector addition by incorporating richer, non-linear interactions, improving performance on N-1 and 1-N relations. Unlike the purely linear transformations in LineaRE (110), TransMS leverages non-linearity to capture more complex relational dynamics while maintaining a relatively low computational footprint compared to matrix-based approaches. However, its non-linear functions are still relatively simple compared to the complex algebraic structures of rotational or hy-

perbolic models, potentially limiting its capacity for extremely nuanced semantics.

- **TranS** (91): A transition-based method that addresses complex scenarios where the same entity pair can have different relations. Its core innovation is replacing the single relation vector in traditional scoring patterns with a "synthetic relation representation," enabling it to handle diverse relations more effectively and efficiently. This contrasts with models that assign a single, fixed embedding to each relation, offering greater flexibility for polysemous relations. While it addresses polysemy, it does so by increasing the representational overhead for relations, which can impact scalability for KGs with many distinct relation instances, especially when compared to more parameter-efficient models like TransD (78) (Section 2.1).

- **lppTransE/R/D** (119): An early innovation that introduces "role-specific projections" to preserve the logical properties of relations (e.g., transitivity, symmetricity) within translation-based embeddings. By projecting head and tail entities using distinct operators, these models aim to overcome the limitations of undifferentiated entity roles in earlier TransE variants. This represents an early attempt to inject more semantic nuance into translational models, increasing parameter count compared to base TransE but offering more expressive power for specific logical patterns. This approach, while effective for its time, is less sophisticated than later generalized transformations like HousE (107), which offer a broader range of geometric operations through more advanced matrix parameterizations.

This diverse set of models highlights several patterns and tensions in KGE research. There is a continuous effort to move beyond simple vector operations towards more mathematically sophisticated transformations and embedding spaces. For instance, TorusE (35), FHRE (33), UniGE (**?**), (118), H2E (**?**), and HolmE (99) all challenge the fundamental assumptions about the embedding space, demonstrating that non-Euclidean geometries or hybrid spaces can offer significant advantages for specific data structures like hierarchies, unlike purely Euclidean models that struggle with such representations. This pattern suggests a growing recognition that the intrinsic geometry of KGs is often non-

34

Table 2: Comparative Overview of Diverse Geometric and Algebraic KGE Innovations

Model	Embedding Space/Metric	Core Operation/Mechanism	Key Patterns/Problems Addressed & Theoretical/Practical Considerations
TorusE (35)	Lie Group (Torus)	Translation on manifold	TransE regularization problem, embedding divergence & Solves regularization elegantly, maintains translation, but may lack expressiveness for complex patterns beyond translation.
FHRE (33)	Lorentz Hyperbolic Space	Fully Hyperbolic Rotation	Hierarchical structures, instability of hybrid hyperbolic models & Direct hyperbolic operation, stable, efficient for hierarchies, specific to Lorentz model, "pure rotation" assumption can be restrictive.
Pan et al. 2021 (118)	Extended Poincaré Ball (Hyperbolic)	Polar coordinates, boundary handling	Hierarchical structures, logical patterns, Poincaré boundary issues & Novel hyperbolic space design, effective for hierarchies, optimization complexity due to boundary conditions.
H2E (?)	Hyperbolic Polar Space	Dual-embedding (modulus/phase), attentional context	Inter/intra-level hierarchies, rich relational context & Enhanced hierarchical modeling, complex dual-embedding, attention adds computational cost.
UniGE (?)	Hybrid (Euclidean + Hyperbolic)	Embedding alignment (Optimal Transport)	Diverse geometric data (chains, hierarchies), single-space limitations & Unifies geometries, robust error bound, complex fusion strategy, higher computational cost due to alignment.
HolmE (99)	Riemannian (Hyperbolic)	Closed under composition (Möbius addition)	Composition patterns (especially long-tail), unifying TransE/RotatE & Novel theoretical property, robust for composition, but Riemannian geometry adds complexity.
ManifoldE (76)	Manifold (Sphere/Hyperplane)	Manifold-based principle	Precise link prediction, ill-posed algebraic systems, over-strict point-wise models & Flexible geometric representation, improves HITS@1, efficient, but manifold choice impacts performance and complexity.
SpherE (120)	Spherical Space	Rotational (entity as sphere)	Many-to-many relations, Set Retrieval & Novel entity representation, good for set-based queries, interpretability.
TransC (45)	Real Vector Space (Spheres/Vectors)	Differentiating concepts (spheres) and instances (vectors)	`instanceOf`, `subClassOf` relations, ontological distinctions & Innovative entity representation, intuitive for hierarchies, requires explicit type information, limiting generalizability.
HousE (107)	Real Vector Space	Householder Transformations (Rotations & Projections)	General relation patterns, mapping properties, higher-dimensional rotations & Generalizes rotations, strong modeling capacity, increased complexity.
GoldE (120)	Real Vector Space	Universal Orthogonal Parameterization	Generalizes Householder, dimensional extension, geometric unification & Unifies orthogonal transformations, high expressiveness, abstract, potentially complex.
CompoundE (73)	Real Vector Space	Cascaded Translation, Rotation, Scaling	General relation patterns, diverse mapping properties & Unifies multiple operations, high expressiveness, potentially complex optimization and higher parameter count.
Rot-Pro (71)	Real Vector Space	Projection + Rotation	Transitivity, common relation patterns & Combines transformations for specific patterns, effective for transitivity, but adds complexity.
CyclE (114)	Cycle Metric (vs. Minkowski)	Metric choice (oscillation property)	Expressiveness limitations of Minkowski metric & Fundamental metric innovation, depends on base KGE, period tuning can be complex, not universally superior.
MQuinE (106)	Real Vector Space	Refined Scoring Function	Z-paradox, diverse relation patterns & Addresses theoretical deficiency, robust, general expressiveness, improves reliability.
LineaRE (110)	Real Vector Space	Linear Regression	Symmetry, inversion,

Euclidean, and a single Euclidean space may be an oversimplification. Similarly, CyclE (114) and ManifoldE (76) re-evaluate the underlying metric and the geometric primitive for entities, respectively, showing that even foundational choices can be refined for improved performance or theoretical soundness. This contrasts with models like TransMS (26) and LineaRE (110), which primarily refine the scoring function within a conventional Euclidean space, showing that significant gains can still be made through clever algebraic manipulations without resorting to exotic geometries.

A critical tension exists between model expressiveness and computational complexity. While models like UniGE, with its hybrid geometry and optimal transport, offer high expressiveness for diverse KG structures, they incur higher computational costs and parameter counts due to the complexity of managing and aligning embeddings across different spaces. This stands in stark contrast to the parameter efficiency of TransD (78) (discussed in Section 2.1) or the linear simplicity of LineaRE (110). The field implicitly assumes that increased mathematical complexity will yield better results, but this is often at the expense of scalability to extremely large KGs or interpretability. For example, while FHRE offers stability and efficiency for hierarchical data by operating fully in hyperbolic space, its reliance on the Lorentz model might limit its applicability to KGs where other geometric properties are more dominant, or where the "pure rotation" assumption is too restrictive. The evaluation of these models primarily relies on link prediction metrics, which, while standard, may not fully capture the nuances of all the complex patterns they claim to model, especially for tasks like precise link prediction (ManifoldE) or concept-instance differentiation (TransC). The generalizability of these models to extremely large and sparse KGs, where the computational overhead of complex transformations can become prohibitive, remains an ongoing challenge. Furthermore, many papers, while demonstrating empirical improvements, often lack a thorough critique of their own assumptions, such as the universal applicability of a single geometric primitive (e.g., spheres in SpherE) or the optimality of a chosen metric across all relation types (CyclE). This methodological oversight can lead to an overestimation of a model's robustness across diverse KG characteristics.

In summary, this subsection highlights the continuous and multifaceted effort to refine the mathematical foundations of KGE. From exploring non-Euclidean spaces and hybrid geometries to introducing novel metrics and specialized linear/algebraic transformations, researchers are pushing the boundaries of how entities and relations can be represented and interacted with algebraically. This trajectory demonstrates a clear evolution from simpler, often restrictive, geometric assumptions towards more flexible, powerful, and theoretically sound mathematical frameworks, albeit with persistent trade-offs in complexity, scalability, and the need for specialized solutions for specific relational patterns. The progression from basic translational models to sophisticated generalized transformations like CompoundE underscores a drive towards unified frameworks that can encapsulate a wider array of relational semantics.

3 Deep Learning Architectures for Knowledge Graph Embedding

While foundational geometric and algebraic models, discussed in the preceding section, laid crucial groundwork for representing entities and relations in continuous vector spaces, the field of Knowledge Graph Embedding (KGE) has witnessed a significant paradigm shift towards leveraging advanced deep learning architectures. This evolution directly addresses inherent limitations of purely geometric approaches, particularly their reliance on predefined transformation rules and their challenges in automatically extracting intricate, non-linear features from complex graph structures (?). Deep learning models offer unparalleled capacity to learn more expressive, context-aware, and hierarchical representations directly from knowledge graph data, pushing the boundaries of KGE performance beyond the constraints of fixed geometric transformations.

This section delves into how Convolutional Neural Networks (CNNs), Graph Neural Networks (GNNs), and Transformer models have been innovatively adapted to capture the rich structural patterns and complex interactions within knowledge graphs. We will explore how CNNs excel at extracting local features and modeling intricate interactions

between entity and relation embeddings, providing a powerful mechanism for discovering hidden patterns. Subsequently, we examine GNNs, which, through their message-passing and aggregation mechanisms, effectively capture structural information and neighborhood context, enabling the learning of richer, context-dependent embeddings that leverage the graph's topology. Finally, we discuss the emergence of Transformer architectures, which utilize self-attention to capture long-range dependencies and contextualized semantics, pushing the state-of-the-art in KGE by moving beyond localized triplet interactions (**? ?**). By enabling automatic feature learning and modeling of highly non-linear relationships, these deep learning paradigms have significantly advanced the state-of-the-art in KGE, paving the way for more robust, scalable, and powerful knowledge reasoning systems.

3.1 Convolutional Neural Networks (CNNs) for KGE

Convolutional Neural Networks (CNNs) have emerged as a powerful paradigm in Knowledge Graph Embedding (KGE), offering a distinct advantage over purely geometric models by automatically extracting intricate local features and modeling complex, non-linear interactions between entity and relation embeddings. This approach marks a significant shift from predefined mathematical transformations, enabling models to discover patterns directly from data. As discussed in previous subsections, while translational and rotational models (78; 35; 64; 107; 73; 104; 106; 46) provide elegant mathematical frameworks for specific relation types, they often struggle with the sheer diversity and complexity of real-world relational patterns, particularly N-to-1, 1-to-N, and N-to-N mappings. CNNs address this by treating concatenated entity and relation embeddings as a 2D input, over which convolutional filters can slide to learn hierarchical feature representations.

A core innovation of CNN-based KGE models lies in their ability to capture rich feature interactions through shared weights and local receptive fields. Unlike simple scoring functions that rely on dot products or Euclidean distances, CNNs can learn multiple interaction patterns simultaneously. For instance, the pioneering work of ConvE (not directly cited but foundational to many CNN KGEs) demonstrated how reshaping the concatenated head entity and relation embeddings into a 2D matrix and applying convolutional

filters could significantly enhance expressiveness. This mechanism allows the model to identify specific patterns that indicate the plausibility of a triplet (h, r, t), moving beyond global transformations to nuanced local feature detection.

One notable approach, **AcrE** (Knowledge Graph Embedding with Atrous Convolution and Residual Learning) (74), aims to address the computational complexity and potential information loss in deep neural network-based KGE models. AcrE's core innovation lies in its use of *atrous convolutions* and *residual learning*. Atrous convolutions enable the model to effectively enlarge its receptive field without increasing the number of parameters or losing resolution, allowing it to capture broader contextual information around the entity-relation pair. This is crucial for understanding relations that might depend on slightly more distant features within the concatenated embedding. Furthermore, the integration of residual learning helps in mitigating the vanishing/exploding gradient problem and ensures that original information is not forgotten as features are processed through deeper layers. AcrE claims to be simpler and more parameter-efficient than existing state-of-the-art methods (74), achieving superior results on diverse datasets. However, while atrous convolutions expand the receptive field, they still operate locally on the input, which might theoretically limit their ability to capture extremely long-range, non-contiguous dependencies compared to self-attention mechanisms found in Transformers (54; 109; 66). Practically, while parameter-efficient, the convolutional operations themselves can still be computationally intensive depending on the filter configurations and input dimensions.

Building on the idea of learning richer feature embeddings, the **Multi-Scale Dynamic Convolutional Network (M-DCN)** (72) was proposed to specifically tackle the challenge of handling complex relation types (1-to-N, N-to-1, N-to-N) that simpler models often struggle with. M-DCN's core innovation is its use of *multi-scale dynamic filters*. Instead of fixed filters, M-DCN generates filter weights that are dynamically related to each specific relation. This allows the model to adapt its feature extraction process based on the unique characteristics of the relation in question, making it highly flexible and expressive for diverse relational patterns. Additionally, M-DCN composes subject entity and relation embeddings in an alternating pattern in the input layer to extract

more feature interactions. This dynamic and multi-scale approach enables M-DCN to achieve state-of-the-art link prediction results on several benchmark datasets (72). The theoretical limitation of M-DCN, despite its power, lies in the increased complexity and potential opacity introduced by dynamic filter generation; understanding the precise rationale behind a filter's adaptation for a given relation can be challenging. Practically, the dynamic nature and multi-scale operations likely incur higher computational costs and parameter counts compared to more straightforward CNN architectures.

Further advancing the capabilities of CNNs, **ReInceptionE** (Relation-Aware Inception Network with Joint Local-Global Structural Information) (70) addresses the limitation of earlier CNN models, such as ConvE, which often lack explicit structural information in their embedding space. ReInceptionE's core innovation is a two-pronged approach: first, it leverages the *Inception network* architecture (popular in computer vision for its ability to capture features at multiple scales) to learn richer query embeddings, thereby increasing the number and diversity of interactions between head and relation embeddings. Second, and crucially, it integrates a *relation-aware attention mechanism* to enrich these query embeddings with *joint local neighborhood and global entity information*. This attention mechanism allows ReInceptionE to dynamically weigh the importance of neighboring entities and relations, effectively incorporating broader graph context into the CNN's local feature extraction process. This directly addresses the critique that CNNs, by their nature, are primarily local feature extractors. ReInceptionE demonstrates competitive performance on datasets like WN18RR and FB15k-237, which are known for their complex relational structures (70). However, the theoretical limitation arises from the heuristic combination of Inception modules and attention; while effective, the precise interaction and contribution of each component can be difficult to disentangle, potentially leading to a more complex "black-box" model. From a practical standpoint, the architectural depth and complexity of Inception networks combined with attention mechanisms generally lead to increased computational demands and a larger number of parameters compared to simpler KGE models.

More recent works continue to refine CNN-based KGE. **CNN-ECFA** (Convolutional

Neural Network-Based Entity-Specific Common Feature Aggregation) (59) focuses on an underexplored area: leveraging entity-specific common feature aggregation to enhance knowledge graph representation learning. Its innovation is a novel CNN-based strategy that aggregates common features specific to entities, which has proven valuable in other deep learning tasks like text classification. This approach aims to provide more discriminative and robust entity representations by focusing on shared, yet entity-specific, characteristics. CNN-ECFA outperforms state-of-the-art feature projection strategies on datasets like WN18RR, YAGO3-10, and NELL-995, demonstrating average improvements in MRR and Hits@1 (59). The theoretical underpinning of "common features" requires careful definition in the KGE context, and its effectiveness might depend on the inherent commonalities present in the dataset. Practically, the aggregation strategy, while beneficial, could introduce additional computational overhead during training.

Similarly, **SEConv** (A Semantic Enhanced Knowledge Graph Embedding Model With AIGC Designed for Healthcare Prediction) (53) addresses the neglect of complex structural features in traditional KGE models, particularly in the context of healthcare prediction. SEConv's core innovation combines a *less resource-consuming self-attention mechanism* with a *multilayer convolutional neural network*. The self-attention component generates more expressive embedding representations, while the multi-layer CNN is adopted to learn deeper structural features from triplets. This dual approach aims for both efficiency (crucial for resource-limited consumer electronics in healthcare) and enhanced expressiveness. SEConv demonstrates substantial improvements on medical datasets like UMLS and DBpedia50, validating its utility for healthcare prediction tasks (53). A theoretical limitation is ensuring the synergistic rather than redundant operation of self-attention and multi-layer CNNs. Practically, while claiming "less resource-consuming," deep neural networks inherently demand significant computational resources, and application-specific models like SEConv require rigorous validation for generalizability beyond their target domain.

The evolution of CNN-based KGE models demonstrates a clear intellectual trajectory. Early CNN applications in KGE primarily focused on leveraging their local feature extrac-

Table 3: Comparative Framework of CNN-based KGE Models

Model	Core Innovation	Problem Solved	Key Advantages	Limitations
AcrE (74)	Atrous convolutions + Residual learning	Complexity, training time, feature interaction	Increased receptive field, parameter efficiency, mitigates gradient issues	Still local, potential for long-range dependency struggle
M-DCN (72)	Multi-scale dynamic filters, alternating composition	Complex relations (1-to-N, N-to-N), limited expressiveness	Relation-adaptive feature extraction, rich interactions	Increased complexity, potential opacity of dynamic filters, higher computational cost
ReInceptionE (70)	Inception network + Relation-aware attention	Lack of structural info, limited interactions in ConvE-like models	Joint local-global structural info, multi-scale feature learning	High complexity (Inception + Attention), increased parameters/computation, interpretability challenges
CNN-ECFA (59)	Entity-specific common feature aggregation	Underexplored entity-specific feature learning	Enhanced entity representations, improved link prediction	Definition of "common features" can be vague, potential computational overhead
SEConv (53)	Multi-layer CNN + Resource-efficient self-attention	Neglect of complex structural features, healthcare prediction	Deeper structural features, expressive embeddings, efficiency for specific applications	"Less resource-consuming" is relative, application-specific generalizability

tion capabilities to move beyond the fixed operations of geometric models. AcrE (74) then introduced atrous convolutions to expand the receptive field while maintaining efficiency, marking a step towards capturing broader context. M-DCN (72) pushed this further by introducing *dynamic filters*, allowing the CNN to adapt its feature learning to specific relation types, thereby addressing the challenge of diverse relation patterns. ReInceptionE (70) represents a significant convergence, integrating attention mechanisms with CNNs to explicitly incorporate *global structural information* alongside local features, directly addressing the limitation of CNNs being purely local. This exemplifies the field's shift from purely local pattern recognition to a more holistic understanding of graph structure. More recent works like CNN-ECFA (59) and SEConv (53) refine these techniques, focusing on specific feature aggregation strategies or application-driven efficiency, indicating a maturation and specialization of CNN applications in KGE.

A critical tension across these models is the trade-off between model expressiveness and computational cost. While CNNs are powerful in learning intricate, non-linear feature interactions, the introduction of multi-scale filters, dynamic mechanisms, and attention layers (as seen in M-DCN (72) and ReInceptionE (70)) inevitably increases parameter

counts and computational demands. This can pose scalability challenges for extremely large knowledge graphs, a concern that is partially addressed by models like AcrE (74) focusing on parameter efficiency or SEConv (53) aiming for "less resource-consuming" designs. Furthermore, most CNN-based KGE models implicitly assume that concatenating entity and relation embeddings and applying convolutions is the optimal way to represent their interaction. However, the precise spatial arrangement of these input embeddings (e.g., (h, r) vs. (r, h)) and the optimal way to represent their interaction for convolutional filters are often unstated assumptions that can significantly impact performance. While these models achieve state-of-the-art results in link prediction, the interpretability of *why* a specific convolutional filter learns a particular pattern for a relation remains a challenge, often making them more "black-box" compared to the mathematically transparent geometric models discussed in Section 2. This lack of interpretability can be a practical limitation, especially in high-stakes applications like healthcare, as highlighted by SEConv (53).

In summary, CNNs have profoundly impacted KGE by enabling the automatic discovery of complex, non-linear patterns that are difficult to model with fixed geometric transformations. From early efforts to extract local features to advanced architectures incorporating dynamic filters and attention mechanisms, CNNs have consistently pushed the boundaries of link prediction performance. However, the field continues to balance the pursuit of higher expressiveness with the practical constraints of computational efficiency, scalability, and model interpretability, paving the way for further innovations in this promising area.

3.2 Graph Neural Networks (GNNs) and Attention Mechanisms

Building upon the local feature extraction capabilities of Convolutional Neural Networks (CNNs) discussed in the previous subsection (74; 70), Graph Neural Networks (GNNs) offer a more direct and inherently graph-aware paradigm for Knowledge Graph Embedding (KGE). While CNNs operate on flattened or reshaped embeddings to capture local interactions, GNNs are specifically designed to explicitly leverage the topological structure of

knowledge graphs through iterative *message passing* and *aggregation* mechanisms (**?**). This allows them to effectively capture structural information, neighborhood context, and relational paths, which are crucial for understanding complex relational patterns that extend beyond simple triplet interactions. The integration of attention mechanisms further refines this process by allowing models to dynamically weigh the importance of different neighbors or relational paths, leading to richer, context-dependent embeddings.

One of the foundational GNN architectures adapted for KGE is **Relational Graph Convolutional Networks (R-GCN)** (**?**). Unlike generic GCNs that typically operate on homogeneous graphs, R-GCN specifically addresses the multi-relational nature of knowledge graphs. Its core innovation lies in using *relation-specific weight matrices* for message passing, allowing it to learn distinct transformations for each relation type. For a given entity, R-GCN aggregates messages from its neighbors, where each message is transformed by a weight matrix corresponding to the relation connecting the neighbor to the central entity. This mechanism enables R-GCN to capture rich structural context and relation-specific patterns, directly addressing the challenge of modeling diverse relation types that simpler geometric models often struggle with. R-GCN demonstrated significant improvements in link prediction and entity classification tasks on benchmark datasets like FB15k-237 and WN18RR (**?**). However, a significant theoretical limitation of R-GCN is its high parameter count, which grows quadratically with the embedding dimension and linearly with the number of relation types. This can lead to scalability issues for KGs with many unique relations, as the model becomes prone to overfitting and computationally expensive. Furthermore, like many GNNs, R-GCN can suffer from the *over-smoothing problem* in deeper layers, where repeated aggregation can cause entity embeddings to become indistinguishable, limiting its capacity to capture long-range dependencies effectively (**?**).

Building on R-GCN's success while addressing its parameter efficiency, **Compositional Graph Neural Networks (CompGCN)** (**?**) introduced a more elegant way to handle relation-specific transformations. CompGCN's core innovation is to integrate *compositional operators* (such as DistMult, ComplEx, or RotatE) directly into the GNN

message-passing framework. Instead of learning a separate weight matrix for each relation as in R-GCN, CompGCN represents relations as operations that transform entity embeddings during message aggregation. This allows it to model diverse relational patterns (e.g., symmetry, antisymmetry, inversion, composition) with significantly fewer parameters, as the relation embeddings themselves participate in the compositional operation. CompGCN effectively solves the parameter explosion problem of R-GCN while simultaneously enhancing the model's expressiveness by leveraging the strengths of established KGE scoring functions. It achieves state-of-the-art results on various link prediction benchmarks (**?**). However, similar to R-GCN, CompGCN is still susceptible to the over-smoothing problem, especially in very deep architectures, as the aggregation process inherently blends information from neighbors. Practically, the choice of compositional operator can impact its expressiveness and the interpretability of the learned interactions, making it a more complex "black-box" model compared to simpler geometric approaches.

A critical challenge for GNNs, and KGE models in general, is *inductive learning*—the ability to embed new, unseen entities or relations without retraining the entire model. While R-GCN and CompGCN are primarily transductive, **Logic Attention Based Neighborhood Aggregation** (84) offered an early foray into inductive capabilities. Its core innovation is the *Logic Attention Network (LAN)*, designed to aggregate information from an entity's neighbors using both rules-based and network-based attention weights. This mechanism allows it to generate embeddings for new entities based on their existing neighborhood structure, a significant advantage over transductive models that require all entities to be present during training. Unlike R-GCN or CompGCN which focus on learning fixed transformations for known relations, LAN dynamically weighs the importance of neighbors, explicitly addressing the unordered and unequal natures of an entity's neighbors. However, its theoretical limitation lies in its reliance on existing neighbors; truly isolated new entities would still pose a challenge, as the model has no information to aggregate. Practically, the complexity of learning and applying dual attention weights can add computational overhead, especially for dense neighborhoods.

Further refining the use of attention for structural information, **Knowledge Graph**

Embedding via Graph Attenuated Attention Networks (GAATs) (41) aims to overcome the limitation of existing graph convolutional network-based models that often assign uniform weights to relation paths or incompletely mine triple features. GAATs' core innovation is an *attenuated attention mechanism* that assigns different weights to different relation paths and effectively acquires information from diverse neighborhoods. This allows entities and relations to be learned from any neighbors, capturing more nuanced contextual information. In contrast to the more general relation-specific transformations of R-GCN or the compositional operations of CompGCN, GAATs specifically focus on dynamically emphasizing more relevant parts of the neighborhood through attention, leading to significant improvements on benchmarks like WN18RR and FB15k-237 (41). A theoretical limitation, shared with other GNNs, is the potential for over-smoothing in deeper layers, where repeated aggregation can make entity embeddings indistinguishable. Practically, the attention mechanism adds computational complexity, and the "attenuated" aspect requires careful design to avoid simply ignoring less prominent but potentially relevant paths, which could lead to information loss.

Pushing the boundaries of representation learning, **DisenKGAT** (Knowledge Graph Embedding with Disentangled Graph Attention Network) (7) tackles the challenge of accurately capturing complex relations that are often oversimplified by single, static entity and relation representations. DisenKGAT's core innovation lies in its use of a *disentangled graph attention network*, leveraging both *micro-disentanglement* and *macro-disentanglement*. Micro-disentanglement is achieved through a novel relation-aware aggregation mechanism that learns diverse component representations for entities and relations. Macro-disentanglement, on the other hand, employs mutual information as a regularization term to enhance the independence of these learned components. This allows DisenKGAT to generate adaptive and diverse representations based on the given scenario, thereby improving its ability to model complex, polysemous relations. Unlike R-GCN and CompGCN which learn a single, albeit complex, representation for each entity/relation, DisenKGAT explicitly aims to break down these representations into independent semantic components. DisenKGAT claims to offer superior accuracy and explainability,

outperforming existing methods on public benchmark datasets (7). The theoretical limitation of disentanglement is ensuring that the learned components are truly independent and semantically meaningful, rather than just arbitrary splits. This often requires careful design of the disentanglement objective and can be sensitive to hyperparameters. Practically, the disentanglement process and mutual information regularization add significant complexity and hyperparameter tuning challenges, potentially increasing training time and requiring more sophisticated optimization compared to GAATs.

More recent advancements continue to refine GNN-based KGE, particularly in addressing scalability and inductive capabilities for dynamic KGs. **InGram** (INductive knowledge GRAph eMbedding) (14) directly addresses the limitation of many inductive KGE methods, including Logic Attention (84), which often assume new entities but not new relations. InGram's core innovation is its ability to generate embeddings for *both new relations and new entities* at inference time. It achieves this by defining a "relation graph" that captures affinities between relations and then uses an attention mechanism to aggregate neighboring embeddings based on both the original knowledge graph and this relation graph. This provides a more comprehensive inductive capability, crucial for real-world KGs where new relations frequently emerge. While Logic Attention provides inductive entity embeddings, InGram extends this to relations, offering a more complete solution for dynamic KGs. However, the complexity of constructing and leveraging the relation graph, especially for very large and sparse relation sets, can introduce practical challenges and computational overhead.

For large-scale KGs, computational efficiency remains a paramount concern. **CPa-WAC** (Constellation Partitioning-based Scalable Weighted Aggregation Composition) (94) proposes a lightweight GNN architecture that combines graph convolutional networks with a novel partitioning strategy. Its innovation lies in a three-stage approach: (1) *Constellation Partitioning (CPa)* uses modularity maximization (e.g., Louvain clustering) to partition the KG into topological clusters, minimizing cross-partition edges without relying on node/edge attributes. (2) *Weighted Aggregation Composition (WAC) Convolution* is an improved compositional GCN with attention layers for local embedding learning

within each partition. (3) A *Global Decoder (GD)* framework then effectively merges these cluster-specific embeddings for global inference. This approach directly tackles the high computational and memory costs of GNNs like R-GCN and CompGCN for large KGs. CPa-WAC demonstrates a significant reduction in training time (up to five times faster) while maintaining comparable prediction accuracy to training on the whole graph (94). This highlights a critical trade-off: achieving scalability through partitioning without sacrificing the global structural information that GNNs are designed to capture. The theoretical challenge lies in ensuring that the partitioning process does not destroy crucial long-range dependencies, and the Global Decoder effectively reconstructs the global context. Practically, the complexity of the partitioning and merging framework adds architectural overhead.

Table 4: Comparative Framework of GNN-based KGE Models with Attention

Model	Core Innovation	Problem Solved	Key Advantages	Limitations
R-GCN (?)	Relation-specific weight matrices for message passing	Multi-relational graphs, structural context	Explicit relation modeling, foundational GNN for KGE	High parameter count, prone to over-smoothing
CompGCN (?)	Compositional operators in GNN message passing	Parameter explosion (R-GCN), complex relation patterns	Parameter efficiency, models diverse relation patterns	Over-smoothing, interpretability of compositional ops
Logic Attention (84)	Logic Attention Network (LAN) for neighborhood aggregation	Inductive KGE for unseen entities	Inductive capability for entities, explicit attention	Relies on existing neighbors, dual attention complexity
GAATs (41)	Graph Attenuated Attention Mechanism	Incomplete feature mining, uniform path weighting	Context-sensitive path weighting, nuanced neighborhood info	Computational cost, potential over-smoothing, careful attenuation design
DisenKGAT (7)	Disentangled Graph Attention Network (micro/macro disentanglement)	Static/single representations, complex relations	Diverse, adaptive, independent component representations	High complexity, hyperparameter sensitivity, training time
InGram (14)	Inductive embedding for new entities *and* relations	Limited inductive scope (entities only)	Full inductive capability (entities + relations), robust for dynamic KGs	Relation graph complexity, potential sparsity issues
CPa-WAC (94)	Constellation Partitioning + Weighted Aggregation Comp. + Global Decoder	Scalability of GNNs for large KGs, long training times	Scalability (up to 5x faster), maintains accuracy, topology-aware partitioning	Architectural complexity, potential info loss across partitions

The evolution of GNNs in KGE demonstrates a clear intellectual trajectory from basic neighborhood aggregation to highly sophisticated, context-aware, and disentangled rep-

resentations, while increasingly addressing scalability and inductive capabilities. Early foundational work like R-GCN (**?**) and CompGCN (**?**) established the power of GNNs for capturing structural context in multi-relational graphs, moving beyond the fixed operations of geometric models (Section 2) and the local-only views of early CNNs (Subsection 3.1). These models primarily operate in a transductive setting. The field then progressed to tackle the critical challenge of inductive learning, with Logic Attention (84) demonstrating early success for new entities, and InGram (14) extending this to new relations, a more comprehensive solution for dynamic KGs. Simultaneously, the refinement of attention mechanisms, from the explicit weighting in Logic Attention to the nuanced "attenuated" approach of GAATs (41), and the disentangled representations of DisenKGAT (7), exemplifies a shift towards understanding the multifaceted semantics embedded within graph structures rather than just connectivity. This pattern of increasing expressiveness, however, often comes at the cost of computational complexity, leading to innovations like CPa-WAC (94) which explicitly addresses the scalability of GNNs for large KGs through partitioning.

A critical tension across these GNN-based models is the trade-off between increased expressiveness and computational complexity. While attention mechanisms and disentanglement (as seen in GAATs (41) and DisenKGAT (7)) significantly enhance the models' ability to capture intricate patterns, they invariably lead to higher parameter counts and longer training times compared to simpler models. For instance, DisenKGAT's mutual information regularization adds significant optimization challenges, potentially increasing training time by a substantial factor compared to a straightforward R-GCN. This can pose practical scalability challenges for extremely large knowledge graphs, a concern that models in Section 6.1 aim to address, and which CPa-WAC (94) directly confronts by reducing training time by up to five times through partitioning. Furthermore, GNNs, by their iterative message-passing nature, implicitly assume that local neighborhood information can be effectively propagated and aggregated to capture global graph structures. However, this assumption can break down in very deep GNNs due to the *over-smoothing problem* (**?**), where entity embeddings become indistinguishable after many layers, limiting their

capacity to capture long-range dependencies effectively. While attention mechanisms mitigate this by selectively focusing on relevant neighbors, they do not entirely eliminate the underlying issue. The interpretability of these complex attention mechanisms and disentangled components also remains a challenge; while DisenKGAT (7) claims explainability, the precise rationale behind the learned components can still be opaque to human understanding, a common limitation for deep learning models. The systematic search for optimal GNN message functions, as explored in (55), further highlights the complexity and non-triviality of designing effective GNN architectures, often requiring automated machine learning techniques like those for hyperparameter tuning (?).

In conclusion, GNNs, particularly when augmented with attention mechanisms, have profoundly advanced KGE by directly leveraging the graph's topology and relational paths. They have moved beyond simple triplet-based interactions to learn richer, context-dependent, and even disentangled representations. This paradigm shift has enabled significant improvements in tasks like link prediction and, crucially, in handling inductive settings for both entities and relations. However, the ongoing challenge lies in balancing the increasing architectural complexity and computational demands with the need for scalability, robustness, and transparent interpretability in real-world applications.

3.3 Transformer-based KGE Models

Following the advancements in Graph Neural Networks (GNNs) and attention mechanisms, which excel at aggregating local neighborhood information (7; 41), a new paradigm has emerged in Knowledge Graph Embedding (KGE) research: the adaptation of Transformer architectures. Originally designed for sequence modeling in natural language processing (?), Transformers, with their powerful self-attention mechanisms, offer a distinct advantage by inherently capturing long-range dependencies and global contextual information, which can be challenging for traditional GNNs that rely on iterative local message passing (?). This subsection explores how these architectures are innovatively applied to knowledge graphs, treating them either as sequences or by integrating graph structures directly into the Transformer framework, thereby enhancing expressiveness for complex

contextual and multi-structural features.

The initial adaptation of Transformers to KGE, exemplified by **CoKE: Contextualized Knowledge Graph Embedding** (54), marked a significant shift from static entity and relation embeddings. CoKE's core innovation lies in recognizing the intrinsic contextual nature of entities and relations within a knowledge graph. It claims to solve the problem of traditional KGE methods that assign a single, static embedding to each entity or relation, thereby ignoring their dynamic properties across different graph contexts. CoKE formulates both edges and paths within the KG as sequences of entities and relations. A Transformer encoder then processes these sequences to obtain dynamic, flexible, and fully contextualized representations. This mechanism allows CoKE to capture contextual meanings that are adaptive to the input sequence, offering an absolute improvement of 19.7

CoKE succeeds under conditions where contextual information, particularly along paths, is crucial for disambiguation and accurate prediction. Its theoretical limitation, however, stems from the inherent challenge of linearizing a graph structure into sequences. While effective for paths, this approach might implicitly lose some explicit graph-level structural information, such as cycles or broader neighborhood connectivity, that GNNs are specifically designed to capture. Practically, processing very long paths can lead to increased computational cost and memory usage, a common challenge for Transformer models. Compared to GNNs like DisenKGAT (7) which focus on disentangling local graph features, CoKE aims for broader contextualization by viewing KGs as sequences, offering a different perspective on capturing rich semantics.

A critical challenge in applying vanilla Transformers to KGs, particularly to triplets, is their inherent *order invariance*. The self-attention mechanism treats input tokens as an unordered set, making it unable to distinguish between a valid triplet (head, relation, tail) and its shuffled, semantically incorrect variants (e.g., tail, relation, head). **Knowformer: Position-Aware Relational Transformer for Knowledge Graph Embedding** (109) directly addresses this fundamental limitation. Its core innovation is a novel Transformer

architecture that incorporates *relational compositions* into entity representations, explicitly injecting semantics and capturing the role of an entity based on its position (subject or object) within a relation triple. This is achieved by defining relational compositions as operators on the relation and the object (or subject) and integrating them into the self-attention mechanism via a carefully designed residual block. Knowformer claims to solve the training inconsistency problem of order-invariant Transformers and formally proves its ability to distinguish entity roles and capture correct relational semantics. It achieves state-of-the-art performance on both link prediction and entity alignment across six benchmark datasets (109).

Knowformer succeeds under conditions where precise understanding of entity roles within a triplet is paramount. Its theoretical limitation is that while it effectively addresses the order invariance for triplets, it primarily operates at the triplet level, potentially not fully capturing the global graph structure or multi-hop dependencies as comprehensively as a dedicated graph-aware Transformer might. Practically, the design of relational compositions and their integration into self-attention adds complexity to the model architecture and training. Compared to CoKE (54), which implicitly handles order by framing paths as sequences, Knowformer directly confronts and resolves the order-invariance issue for individual triplets, making it more robust for fundamental KGE tasks.

Pushing the boundaries further, **TGformer: A Graph Transformer Framework for Knowledge Graph Embedding** (66) represents a significant step towards a more comprehensive integration of Transformer capabilities with graph structures. TGformer's core innovation is being the *first* framework to use a graph transformer to build knowledge embeddings that incorporate both *triplet-level* and *graph-level structural features* in both static and temporal knowledge graphs. It claims to solve the limitations of existing methods: triplet-based approaches that ignore the essential graph structure, and graph-based methods (like many GNNs discussed in the previous subsection) that overlook the contextual information of nodes. TGformer constructs a context-level subgraph for each predicted triplet and then employs a Knowledge Graph Transformer Network (KGTN) to fully explore multi-structural features. This allows the model to understand entities and

relations in different contexts, boosting performance in link prediction. Experimental results on several public datasets show that TGformer achieves state-of-the-art performance (66).

TGformer succeeds by leveraging the global receptive field of Transformers to integrate diverse structural information, making it particularly effective for KGs where both local triplet patterns and broader graph context are important. Its theoretical limitation lies in the increased complexity of integrating multiple levels of structural information and handling temporal dynamics, which can lead to higher computational demands and parameter counts. Practically, the construction of context-level subgraphs and the design of the KGTN require careful engineering and can be resource-intensive, especially for very large and dense KGs. TGformer represents a convergence, building on the contextualization ideas of CoKE (54) and the structural awareness of GNNs, while extending to temporal KGs, a domain often addressed by specialized models (see Section 5.1).

Table 5: Comparative Framework of Transformer-based KGE Models

Model	KG Representation	Core Innovation	Problem Addressed	Key Advantages	Limitations
CoKE (54)	Edges/Paths as sequences	Transformer encoder on linearized sequences	Static embeddings, ignoring context	Captures dynamic, contextualized meanings along paths	Potential loss of explicit graph structure, computational for long paths
Knowformer (109)	Triplet-level (h,r,t)	Position-Aware Relational Transformer with relational compositions	Transformer's order invariance for triplets	Distinguishes entity roles, robust for triplet semantics	Primarily triplet-focused, complex relational composition design
TGformer (66)	Triplet-level + Graph-level (static	Graph Transformer framework for multi-structural features	Triplet-based ignore graph, graph-based ignore node context	Integrates global/local, static/temporal features, state-of-the-art	High complexity, computational cost, resource-intensive

The intellectual trajectory of Transformer-based KGE models demonstrates a clear evolution from adapting sequence models to explicitly integrating graph structures. Early work like CoKE (54) assumed that KGs could be effectively represented as sequences (paths or edges), leveraging the Transformer's strength in contextualization. This approach, while innovative, implicitly makes the assumption that the linear ordering of elements in a path is sufficient to capture all relevant graph semantics, which may not hold for complex, non-linear graph structures. This tension between sequence-centric and

graph-centric views is a recurring theme. Knowformer (109) then addressed a fundamental limitation of applying vanilla Transformers directly to triplets, the order invariance, by introducing position-aware mechanisms. This marked a shift towards explicitly making Transformers "graph-aware" at a granular level. Finally, TGformer (66) represents a more comprehensive synthesis, moving beyond mere sequence adaptation or triplet-level fixes to a full graph transformer framework that explicitly models both local (triplet-level) and global (graph-level) structural features, and even extends to temporal dynamics. This progression exemplifies the field's shift from simply applying powerful neural architectures to carefully adapting and innovating them to respect the unique structural properties of knowledge graphs.

A critical pattern across these models is the trade-off between the simplicity of adaptation and the richness of graph structure capture. CoKE's sequence-based approach is relatively straightforward but might sacrifice explicit graph topology. Knowformer's focus on triplet-level order invariance is crucial but remains largely localized. TGformer, while offering the most comprehensive integration, introduces significant architectural and computational complexity. This reflects a broader tension in KGE research: how to leverage the immense power of general-purpose deep learning models while respecting the specific, often irregular, structure of knowledge graphs. Furthermore, while these Transformer-based models achieve state-of-the-art performance, their increased complexity often comes at the cost of interpretability, making it harder to understand *why* a particular prediction is made, a limitation shared with many deep learning approaches. This issue is particularly relevant for high-stakes applications discussed in Section 7.4. The reliance on large datasets for pre-training or fine-tuning is another practical limitation, as Transformers are notoriously data-hungry, which can be a challenge for smaller or domain-specific KGs.

In conclusion, Transformer-based KGE models have revolutionized the field by introducing powerful self-attention mechanisms to capture long-range dependencies and highly contextualized representations. From CoKE's sequence-based contextualization (54) to Knowformer's position-aware relational modeling (109) and TGformer's comprehensive

54

graph transformer framework (66), these models have significantly pushed the state-of-the-art in KGE performance. They demonstrate an innovative application of architectures originally designed for sequence modeling to graph structures, effectively capturing global and local semantic relationships. The ongoing challenge lies in balancing their immense expressive power with computational efficiency, scalability, and interpretability, particularly as KGs continue to grow in size and complexity.

4 Enriching KGE: Auxiliary Information, Rules, and Multi-modality

While the preceding section demonstrated the profound impact of deep learning architectures, such as CNNs, GNNs, and Transformers, in extracting intricate structural patterns and learning context-aware representations directly from knowledge graph topology, real-world knowledge graphs often present challenges that purely structural models cannot fully address. These include inherent data sparsity, the need for deeper semantic understanding beyond connectivity, and the demand for explicit logical reasoning capabilities. To overcome these limitations and move towards a more comprehensive and nuanced representation of knowledge, this section explores advanced Knowledge Graph Embedding (KGE) approaches that integrate diverse external knowledge sources and logical constraints, thereby enriching the learned embeddings.

This section delves into three primary avenues for KGE enrichment. Firstly, we examine methods that incorporate auxiliary information, such as entity types, attributes, and hierarchical structures, to provide richer semantic context and guide the embedding process, making models more robust to incompleteness (? ?). Secondly, we discuss the integration of explicit logical rules and constraints, which inject prior knowledge into the learning process, enhancing reasoning capabilities and improving the interpretability and consistency of the learned embeddings (?). Finally, we explore multi-modal and cross-domain KGE techniques that leverage complementary information from various sources, including textual descriptions and visual features, to alleviate data sparsity and enable a

more holistic understanding of entities and relations (? ?). By moving beyond solely structural information, these enriched KGE models aim to achieve superior performance, enhance semantic understanding, improve reasoning, and foster greater robustness and interpretability in complex, real-world scenarios.

4.1   Incorporating Auxiliary Information (Types, Attributes)

The effectiveness of Knowledge Graph Embedding (KGE) models, as discussed in previous sections, largely stems from their ability to capture structural patterns within the graph. However, relying solely on the triplet structure (head, relation, tail) can oversimplify the rich semantics inherent in real-world knowledge graphs, leading to limitations in expressiveness, robustness, and interpretability, especially when dealing with incomplete or noisy data. This subsection delves into advanced KGE approaches that enhance representations by integrating auxiliary semantic information, such as entity types, attributes, and hyper-relational facts, thereby grounding embeddings in richer context and providing more semantic guidance.

The integration of auxiliary information into KGE models marks a significant paradigm shift from purely structural learning towards semantically enriched representations. Early KGE models, like TransE and its extensions (as reviewed in (116)), primarily focused on geometric transformations in vector spaces. While effective for capturing basic relational patterns, they often struggled with the nuances of entity roles or the reliability of facts. The evolution towards incorporating auxiliary data directly addresses these limitations, aiming to produce more discriminative and robust embeddings.

Three major paradigms emerge in leveraging auxiliary information:

1. **Entity Type Integration**: Models that explicitly use entity type hierarchies or categorical labels to refine entity and relation embeddings.

2. **Entity Attribute Integration**: Approaches that incorporate descriptive attributes of entities to enrich their representations, often to enhance robustness against noise.

3. **Hyper-relational Fact Modeling**: Methods that extend beyond simple triplets

to capture more complex, n-ary relationships.

Entity Type Integration

The core problem addressed by entity type integration is the limited semantic guidance in traditional KGEs, which often treat all entities as abstract nodes without distinguishing their inherent categories or roles. This can lead to less discriminative embeddings and difficulties in handling complex relational patterns where entity types impose strong constraints.

TransET (81) is a pioneering framework that leverages entity types to learn more semantic features.

- **Problem Solved**: It addresses the lack of semantic guidance in traditional KGEs by incorporating well-constructed prior knowledge in the form of entity types, which helps in distinguishing entity roles and refining relation representations.

- **Core Innovation**: TransET utilizes a circle convolution based on the embeddings of entities and entity types to map head and tail entities to type-specific representations. This allows for a more nuanced understanding of how relations interact with entities of specific types. A translation-based score function then learns from these type-specific representations.

- **Conditions for Success**: TransET succeeds when reliable and sufficiently granular entity type information is available. Its performance is validated on real-world datasets for link prediction and triple classification, demonstrating superiority over state-of-the-art models in most cases.

- **Theoretical Limitations**: While effective, TransET's reliance on explicit type information means its performance can degrade if type data is sparse, noisy, or unavailable. The circle convolution, while innovative, might not capture all complex interactions between entity and type embeddings.

- **Practical Limitations**: The need for pre-defined and well-structured entity types can be a practical bottleneck, as many KGs might lack comprehensive type annota-

tions. The complexity of integrating type embeddings with entity embeddings also adds to the model's parameter count.

- **Comparison to Alternatives**: Compared to purely structural models like TransE (116) or RotatE (77), TransET provides a more semantically grounded representation. It offers a more direct way to inject prior knowledge than some deep learning models that might implicitly learn type-like features but without explicit guidance.

Building on this, **TaKE** (11) proposes a universal type-augmented KGE framework for knowledge graph completion.

- **Problem Solved**: TaKE aims to further improve KG completion by addressing the limitation of traditional KGE methods that primarily focus on structured triples, neglecting valuable entity type information. It seeks to provide more semantic guidance for embedding learning.

- **Core Innovation**: TaKE automatically captures type features without explicit type supervision, which is a significant advancement over methods requiring pre-defined types. It learns different type representations for each entity, allowing it to distinguish the diversity of types specific to distinct relations. Furthermore, it introduces a novel type-constrained negative sampling strategy to construct more effective negative samples during training. This is crucial as negative sampling quality (as reviewed in (117)) directly impacts KGE performance.

- **Conditions for Success**: TaKE demonstrates its merits on four datasets from Freebase, WordNet, and YAGO, achieving state-of-the-art performance when combined with models like SimplE. Its ability to automatically capture type features makes it more robust to KGs with implicit or less structured type information.

- **Theoretical Limitations**: While it mitigates the need for explicit type supervision, the quality of automatically captured type features might still depend on the richness of the existing graph structure. The effectiveness of type-constrained negative sampling relies on the assumption that type information provides meaningful constraints for valid triples.

- **Practical Limitations**: The framework's universality means its performance can still be influenced by the underlying KGE model it augments. The computational overhead of automatically capturing type features and implementing type-constrained negative sampling might be higher than simpler methods.

- **Comparison to Alternatives**: TaKE represents an evolution from TransET by reducing the reliance on explicit type supervision and introducing a more sophisticated negative sampling strategy. It offers a more flexible way to incorporate type information compared to models that hardcode type constraints.

A domain-specific application of type-constrained KGE is seen in **GeoEntity-type constrained KGE** (59).

- **Problem Solved**: This approach addresses the inaccurate prediction of natural-language spatial relations between geographic entities (geoentities) by prior studies that often overlook essential semantic attributes.

- **Core Innovation**: The SR-KGE framework integrates geoentity types as a constraint to capture spatial and semantic relations more accurately. It fuses graph structures and semantic attributes, considering the diversity of natural language expressions in the embedding process.

- **Conditions for Success**: It shows superior performance on both small-scale and large-scale knowledge graph datasets for spatial relation inference, outperforming general KGE models like TransE, RotatE, and HAKE. Its success is tied to the availability and relevance of geoentity type information for spatial reasoning.

- **Theoretical Limitations**: The domain-specific nature of this model means its innovations might not directly generalize to other types of relations (e.g., temporal, social) where geoentity types are irrelevant.

- **Practical Limitations**: Requires specialized geoentity type annotations, which might not be readily available in all KGs.

- **Comparison to Alternatives**: Unlike general KGEs, SR-KGE is tailored to a specific, complex relational task, demonstrating how auxiliary information can be crucial for domain-specific accuracy.

Entity Attribute Integration

Beyond categorical types, entities often possess rich descriptive attributes. Integrating these attributes can provide a deeper understanding of entities, particularly useful for handling noisy data.

AEKE: Attributed Error-aware Knowledge Embedding (112) is a notable framework that leverages entity attributes to guide error-aware embedding learning.

- **Problem Solved**: AEKE tackles the critical issue of erroneous triples inevitably injected during KG construction. Most KGE algorithms assume triples are correct, leading to significant performance degradation in downstream applications when errors are present.

- **Core Innovation**: AEKE leverages the semantics contained in entity attributes to guide the KGE model against the impact of erroneous triples. It designs two triple-level hypergraphs to model the topological structures of the KG and its attributes. A confidence score for each triple is jointly calculated based on self-contradiction within the triple, consistency between local and global structures, and homogeneity between structures and attributes. These confidence scores adaptively update the weighted aggregation in a multi-view graph learning framework and the margin loss in KGE, ensuring that potential errors contribute minimally to learning.

- **Conditions for Success**: AEKE demonstrates superior performance over state-of-the-art KGE and error detection algorithms on three real-world KGs. Its success hinges on the availability of meaningful entity attributes and the effectiveness of its confidence scoring mechanism in identifying erroneous triples.

- **Theoretical Limitations**: The reliance on entity attributes means that KGs with sparse or low-quality attributes might not fully benefit. The complexity of con-

structing and reasoning with triple-level hypergraphs and multi-view learning adds to the theoretical burden.

- **Practical Limitations**: The computational cost of building and processing two hypergraphs and calculating confidence scores for every triple can be substantial, especially for very large KGs. Data requirements for comprehensive attributes can also be high.

- **Comparison to Alternatives**: AEKE directly addresses data quality issues, a limitation often overlooked by KGEs that assume clean data. Unlike methods that only detect errors, AEKE integrates error awareness directly into the embedding learning process, making it more robust.

Hyper-relational Fact Modeling

Traditional KGE models are predominantly triplet-centric. However, many real-world facts are hyper-relational, involving more than two entities or additional key-value pairs that qualify the main triplet.

HINGE: Hyper-Relational Knowledge Graph Embedding ([68]) extends KGE beyond triplets to directly model hyper-relational facts.

- **Problem Solved**: HINGE addresses the oversimplification of complex hyper-relational facts by triplet-only KGE models. These models fail to capture additional key-value pairs associated with a base triplet, leading to suboptimal representations.

- **Core Innovation**: HINGE directly learns from hyper-relational facts, which include a base triplet (h, r, t) and associated key-value pairs (k, v). It captures both the primary structural information of the KG (triplets) and the correlation between each triplet and its associated key-value pairs. This moves beyond n-ary representations that discard the fundamental triplet structure.

- **Conditions for Success**: Extensive evaluation shows HINGE's superiority on various link prediction tasks, outperforming both triplet-only methods (by 0.81-41.45

61

- **Theoretical Limitations**: Modeling hyper-relational facts inherently increases the complexity of the data structure and the embedding space. Sparsity in key-value pairs for certain relations could limit its effectiveness.

- **Practical Limitations**: The need for KGs to be structured with hyper-relational facts is a prerequisite, which might not be the case for all datasets. The increased dimensionality and complexity of the model can lead to higher computational costs during training and inference.

- **Comparison to Alternatives**: HINGE offers a more comprehensive approach than methods that either ignore auxiliary information or transform hyper-relational facts into less informative n-ary representations. It provides a richer context for each fact, leading to more accurate predictions.

Comparative Framework and Evolution

The evolution of KGEs incorporating auxiliary information can be summarized as a progressive movement from relying solely on structural patterns to leveraging richer semantic context (Table

Table 6: Comparison of KGE Models Incorporating Auxiliary Information

Model Family	Key Papers	Auxiliary Info Type	Core Innovation	Problem Addressed	Key Advantage/Limitation
Entity Type	TransET [81], TaKE [11], GeoEntity-type [59]	Entity Types (categorical)	Type-specific representations, type-constrained negative sampling, automatic type feature capture	Limited semantic guidance, entity role ambiguity, domain-specific relations	Advantage: More discriminative, semantic embeddings. Limitation: Reliance on type quality/availability.
Entity Attribute	AKRE [12]	Entity Attributes (descriptive)	Multi-view hypergraph, confidence scores for error-aware learning	Performance degradation from erroneous triples	Advantage: Robustness to noisy data. Limitation: Computational cost, attribute quality dependence.
Hyper-relational	HINGE [81]	Key-value pairs (contextual qualifiers)	Direct learning from hyper-relational facts	Oversimplification by triplet-only models	Advantage: Captures richer fact semantics. Limitation: Increased data complexity, sparsity of qualifiers.

Connections Across Papers and Broader Themes

([81])'s **TransET** and ([11])'s **TaKE** both exemplify the field's shift towards leveraging well-structured prior knowledge to enhance KGEs. While TransET requires explicit entity types, TaKE builds on this by innovating with automatic type feature capture and a type-constrained negative sampling strategy, making it more adaptable to KGs with less explicit type annotations. This shows a clear progression in reducing the manual effort required for auxiliary information integration. The success of ([59])'s **GeoEntity-type constrained KGE** further reinforces this, demonstrating that even domain-specific type

62

information can significantly boost performance for specialized tasks, highlighting the general utility of type-based semantic guidance.

The work by (112) on **AEKE** addresses a different, but equally critical, challenge: the robustness of KGEs to noisy data. While TransET and TaKE focus on enriching semantic understanding for completion, AEKE leverages entity attributes to directly guide error-aware embedding learning. This highlights a broader tension in KGE research: whether to focus on improving expressiveness for ideal data or enhancing robustness for imperfect real-world data. AEKE's multi-view hypergraph and confidence scoring mechanism represent a sophisticated approach to mitigate the impact of errors, a problem that many KGEs implicitly assume away.

HINGE (68) offers a distinct yet complementary approach by extending the fundamental data structure itself. While TransET/TaKE enrich the *interpretation* of entities and relations within triplets, HINGE acknowledges that some facts are inherently *more complex than triplets*. This exemplifies the field's shift from a rigid triplet-centric paradigm to one that accommodates the intrinsic complexity of real-world knowledge. HINGE's ability to directly learn from hyper-relational facts provides a richer context for each fact, which can be seen as a form of auxiliary information that is integral to the fact itself, rather than external.

Patterns and Tensions

A recurring trade-off across these approaches is the balance between the richness of auxiliary information and its availability or complexity. While more semantic context generally leads to better performance, acquiring high-quality entity types, attributes, or hyper-relational facts can be challenging. Many KGs are incomplete not only in their structural triples but also in their auxiliary metadata. This leads to a tension between models that require explicit, well-structured auxiliary data (like TransET) and those that can infer or adapt to less complete information (like TaKE's automatic type feature capture).

Another tension lies in the computational cost. Integrating auxiliary information, especially through complex mechanisms like multi-view hypergraphs in AEKE or direct

hyper-relational modeling in HINGE, often increases model complexity and computational demands. This can contrast with the need for efficient and scalable KGEs, as discussed in Section 6.1.

The field implicitly assumes that auxiliary information, when available, is largely reliable. However, just as structural triples can be erroneous, so too can entity types or attributes. AEKE directly questions this assumption by building an error-aware framework, suggesting a future direction where the quality of auxiliary information itself is critically assessed and integrated into the learning process.

In conclusion, incorporating auxiliary information represents a crucial advancement in KGE, moving beyond purely structural patterns to leverage richer semantic context. By integrating entity types, attributes, and hyper-relational facts, models like TransET (81), TaKE (11), GeoEntity-type constrained KGE (59), AEKE (112), and HINGE (68) provide more semantic, discriminative, and robust representations. These approaches collectively demonstrate that external, well-structured knowledge is indispensable for enhancing KGE performance, particularly in dealing with incomplete or noisy knowledge graphs, by grounding embeddings in a more comprehensive understanding of the underlying knowledge.

4.2   Rule-based and Constraint-driven KGE

Building on the previous discussion of incorporating auxiliary semantic information like entity types and attributes, this subsection shifts focus to a more explicit form of prior knowledge: logical rules and structural constraints. While auxiliary information enriches the semantic content of individual entities and relations, rule-based and constraint-driven approaches aim to inject broader logical patterns and consistency requirements directly into the Knowledge Graph Embedding (KGE) learning process. This is crucial because purely data-driven KGE models, despite their success in capturing statistical patterns, often struggle with logical reasoning, ensuring semantic consistency, and providing interpretable explanations for their predictions (48). By making embeddings adhere to logical patterns, these methods enhance reasoning capabilities, improve interpretability, and lead

to more robust and semantically coherent representations, addressing the limitations of models that solely rely on observed factual triples.

The integration of logical rules and constraints into KGE models represents a significant advancement, moving beyond simple geometric transformations or deep learning architectures to explicitly guide the embedding space. This paradigm aims to bridge the gap between symbolic logic and distributed representations, leveraging the strengths of both. Three primary families of approaches can be identified:

1. **Semantic Smoothness and Simple Structural Constraints**: Early methods that impose general consistency or smoothness assumptions on the embedding space.

2. **Iterative Soft Rule Guidance**: Approaches that dynamically integrate uncertain logical rules into the training process, allowing for continuous refinement.

3. **Principled Rule Embedding and Regularization**: More sophisticated frameworks that jointly learn rule representations alongside entity and relation embeddings, or use rules as direct regularizers.

Semantic Smoothness and Simple Structural Constraints

The initial attempts to inject prior knowledge into KGE focused on enforcing general properties of the embedding space, rather than complex logical rules. These methods aimed to ensure that the learned representations respected basic semantic or structural consistencies.

Semantically Smooth Knowledge Graph Embedding (SSE) (97) addresses the problem of KGEs primarily relying on observed facts, which can lead to embedding spaces that lack intrinsic geometric structure or semantic coherence.

- **Core Innovation**: SSE proposes to enforce a "semantically smooth" embedding space where entities belonging to the same semantic category are encouraged to lie close to each other. It achieves this by formulating manifold learning algorithms, specifically Laplacian Eigenmaps and Locally Linear Embedding, as geometrically

based regularization terms. These terms constrain the embedding task, guiding the optimization towards a smoother, more semantically organized space.

- **Conditions for Success**: SSE demonstrates significant and consistent improvements in link prediction and triple classification on benchmark datasets. Its success is contingent on the availability of reliable semantic category information for entities, which serves as the basis for defining "smoothness."

- **Theoretical Limitations**: The effectiveness of SSE heavily relies on the quality and completeness of entity categorization. If categories are noisy or too coarse, the "smoothness" constraint might not provide meaningful guidance. Moreover, manifold learning, while powerful, can be computationally intensive for very large graphs.

- **Practical Limitations**: The requirement for additional semantic category information might not always be met in real-world KGs, necessitating pre-processing or external data integration. The regularization terms add complexity to the optimization problem.

- **Comparison to Alternatives**: Compared to purely structural models, SSE provides a more semantically grounded embedding. It differs from later rule-based methods by imposing a general geometric constraint rather than specific logical patterns, making it less expressive for complex reasoning but potentially more robust to rule sparsity.

Building on the idea of simple constraints, **Improving Knowledge Graph Embedding Using Simple Constraints** (90) investigates the potential of straightforward constraints to enhance KGEs, contrasting with the trend of designing increasingly complex scoring models.

- **Problem Solved**: This work addresses the need for more interpretable and structured embedding spaces, which are often lacking in complex KGE models. It aims to improve performance without significantly increasing model complexity.

66

- **Core Innovation**: The paper examines two types of simple constraints: non-negativity constraints on entity representations and approximate entailment constraints on relation representations. Non-negativity helps learn compact and interpretable entity vectors, while approximate entailment encodes logical regularities between relations (e.g., if A implies B, then their embeddings should reflect this). These constraints are imposed as prior beliefs on the structure of the embedding space.

- **Conditions for Success**: Evaluated on WordNet, Freebase, and DBpedia, the approach shows significant and consistent improvements over competitive baselines. Its success lies in the simplicity and intuitive nature of the constraints, which are generally applicable and do not require complex rule mining.

- **Theoretical Limitations**: While simple constraints improve interpretability and structure, they might not be expressive enough to capture highly complex logical patterns or multi-hop reasoning. The "approximate" nature of entailment means it's a soft guidance, not a strict logical enforcement.

- **Practical Limitations**: The benefits are tied to the underlying KGE model being constrained; these constraints are augmentations, not standalone KGEs. Their impact might be less pronounced than more sophisticated rule-integration methods for tasks requiring deep logical inference.

- **Comparison to Alternatives**: Unlike SSE (97) which uses manifold learning for smoothness, (90) employs direct mathematical constraints, offering a more lightweight approach. It provides a foundational step towards rule integration, demonstrating that even basic logical principles can significantly improve embedding quality, without the overhead of explicit rule mining required by later methods like RUGE (13).

Iterative Soft Rule Guidance

A more direct and powerful approach involves integrating logical rules, especially "soft rules" (rules with confidence levels), into the embedding process. This often involves

iterative mechanisms to refine both embeddings and rule applications.

Knowledge Graph Embedding with Iterative Guidance from Soft Rules (RUGE) (13) tackles the challenge of integrating logical rules, particularly soft rules, into KGE models. Previous attempts often made a one-time injection of hard rules, ignoring the interactive nature of embedding learning and logical inference, and the abundance of uncertain, automatically extracted rules.

- **Problem Solved**: RUGE addresses the limitations of one-time rule injection and the exclusive focus on hard, exception-less rules. It aims to effectively leverage automatically extracted soft rules to enhance KGE performance and transfer knowledge more deeply.

- **Core Innovation**: RUGE proposes a novel paradigm that iteratively guides embedding learning using soft rules. It learns simultaneously from labeled triples, unlabeled triples (whose labels are predicted iteratively), and soft rules with varying confidence levels. In each iteration, rules are queried to obtain soft labels for unlabeled triples, which are then integrated to update the embedding model. This iterative procedure allows knowledge from rules to be progressively transferred into the embeddings.

- **Conditions for Success**: Evaluated on link prediction tasks on Freebase and YAGO, RUGE achieves significant and consistent improvements over state-of-the-art baselines. Its success is particularly notable because it effectively leverages automatically extracted soft rules, even those with moderate confidence levels, demonstrating robustness to rule uncertainty.

- **Theoretical Limitations**: The iterative nature, while powerful, can be computationally more expensive than one-time rule injection. The quality of the soft rules themselves is paramount; noisy or contradictory rules could potentially mislead the embedding process. The framework's performance is also tied to the underlying KGE model it augments.

- **Practical Limitations**: Efficiently extracting high-quality soft rules from KGs remains a practical challenge. Balancing the influence of observed triples versus rule-derived soft labels requires careful hyperparameter tuning.

- **Comparison to Alternatives**: RUGE represents a significant evolution from earlier constraint-based methods like SSE (97) and (90) by directly integrating explicit logical rules, rather than general smoothness or structural properties. It moves beyond hard rules, which are scarce and difficult to validate, to embrace the more abundant but uncertain soft rules, making it more applicable to real-world KGs.

Further refining the integration of soft rules, **Knowledge Graph Embedding Preserving Soft Logical Regularity** (20) focuses on improving scalability and efficiency by imposing rule constraints directly on relation representations.

- **Problem Solved**: This paper addresses the challenge of efficiently and effectively integrating soft logical information, particularly the scalability issues that arise when rule learning complexity depends on the entity set size.

- **Core Innovation**: The method represents relations as bilinear forms and maps entity representations into a non-negative and bounded space. Crucially, it derives a rule-based regularization term that *merely enforces relation representations* to satisfy constraints introduced by soft rules. This design makes the complexity of rule learning independent of the entity set size, significantly improving scalability.

- **Conditions for Success**: The approach demonstrates effectiveness in link prediction on Freebase and DBpedia, outperforming many competitive baselines. Its scalability makes it particularly suitable for large KGs where entity-dependent rule learning would be prohibitive.

- **Theoretical Limitations**: While improving scalability, the approach might lose some granularity by focusing solely on relation representations, potentially overlooking entity-specific nuances that rules might imply. The non-negative and bounded space for entities might impose restrictions that are not always ideal for all KGE models.

- **Practical Limitations**: The method still relies on the availability of soft rules, and the process of extracting and validating these rules remains a prerequisite. The specific representation of relations as bilinear forms might not be universally compatible with all KGE architectures.

- **Comparison to Alternatives**: This work builds upon the principles of soft rule integration established by RUGE (13) but offers a more scalable solution by decoupling rule learning complexity from the number of entities. It provides a more direct and efficient way to regularize relation embeddings based on logical patterns, a critical improvement for large-scale applications.

Principled Rule Embedding and Regularization

The most recent advancements in this area go beyond simply using rules as regularization, instead learning explicit embeddings for rules themselves, allowing for more sophisticated logical inference and deeper integration.

RulE: Knowledge Graph Reasoning with Rule Embedding (44) introduces a novel and principled framework for enhancing KG reasoning by effectively leveraging logical rules.

- **Problem Solved**: RulE addresses the brittleness of traditional logic-based reasoning and the limited reasoning capabilities of purely KGE models. It aims to deeply intertwine logical rules with embedding learning in a unified space.

- **Core Innovation**: Unlike previous KGE methods, RulE learns *rule embeddings* from existing triplets and first-order rules by jointly representing entities, relations, and logical rules in a unified embedding space. This allows for a confidence score to be calculated for each rule, enabling soft logical inference. Concurrently, the learned rule embeddings inject prior logical information into the embedding space, enriching and regularizing entity/relation embeddings.

- **Conditions for Success**: Extensive experiments on multiple benchmarks reveal that RulE outperforms the majority of existing embedding-based and rule-based

approaches. Its success lies in its ability to perform soft logical inference and use rule embeddings to regularize and enrich entity/relation embeddings, leading to more robust and accurate reasoning.

- **Theoretical Limitations**: The joint learning of entity, relation, and rule embeddings can significantly increase model complexity and the number of parameters, potentially leading to higher computational costs. The quality of rule embeddings is highly dependent on the quality and completeness of the input rules.

- **Practical Limitations**: The framework requires a robust mechanism for rule extraction and representation. The computational overhead during training, especially for very large rule sets, could be substantial, although the paper emphasizes its conceptual simplicity.

- **Comparison to Alternatives**: RulE represents a sophisticated evolution from earlier rule-integration methods like RUGE ([13]) and ([20]). While RUGE iteratively guides embeddings with soft rules and ([20]) regularizes relation embeddings, RulE takes a more fundamental step by learning *embeddings for the rules themselves*. This allows for a more principled and unified approach to soft logical inference and regularization, making it a powerful framework for deep integration of logic and embeddings.

Comparative Framework and Evolution

The evolution of rule-based and constraint-driven KGEs demonstrates a clear progression from general consistency enforcement to explicit, iterative, and finally, deeply integrated logical reasoning (Table

Table 7: Comparison of Rule-based and Constraint-driven KGE Approaches

Model Family	Key Papers	Type of Logic/Constraint	Core Mechanism	Problem Addressed	Key Advantage/Limitation
General Consistency	SSE [87], [90]	Semantic smoothness, non-negativity, approximate entailment	Manifold learning regularization, direct mathematical constraints	Lack of intrinsic geometric structure, interpretability, basic consistency	**Advantage:** Simple, generalizable. **Limitation:** Limited expressiveness for complex logic.
Iterative Soft Rule Guidance	RUGE [13]	Soft logical rules (with confidence)	Iterative guidance, soft label generation	One-time rule injection, reliance on hard rules, knowledge transfer	**Advantage:** Leverages uncertain rules, deep knowledge transfer. **Limitation:** Computational cost, rule quality dependence.
Principled Rule Embedding	[20], RulE [44]	Soft logical rules, first-order rules	Rule-based regularization on relations, joint learning of rule embeddings	Scalability of rule learning, brittleness of logic, unified reasoning	**Advantage:** Scalable, unified reasoning, explicit rule representation. **Limitation:** Increased model complexity, rule extraction overhead.

Connections Across Papers and Broader Themes

The progression from (97)'s **SSE** to (13)'s **RUGE** highlights a fundamental shift in how "prior knowledge" is conceptualized and integrated. While SSE focuses on a global "semantic smoothness" constraint, RUGE moves to explicit, local logical rules, demonstrating the field's increasing ambition to inject more precise and actionable knowledge. This transition reflects a broader theme in AI: moving from implicitly learning patterns to explicitly encoding human-understandable knowledge.

(90)'s work on **simple constraints** can be seen as a precursor to more complex rule integration. By showing that even basic non-negativity and approximate entailment can improve KGEs, it laid the groundwork for the idea that *any* form of logical guidance, no matter how simple, is beneficial. This contrasts with the purely data-driven models discussed in foundational KGE sections (e.g., (76) TransA, (56) TransA, (76) ManifoldE, (76) TransG), which often lack such explicit semantic or logical structuring.

The work by (20) on **preserving soft logical regularity** directly addresses a practical limitation of iterative rule guidance like RUGE (13): scalability. By making rule learning complexity independent of entity set size, it allows the benefits of soft rule integration to extend to much larger knowledge graphs, which is critical for real-world applications. This also connects to the broader theme of efficiency and scalability discussed in Section 6.1, demonstrating how methodological innovations can address practical bottlenecks.

Finally, **RulE** (44) represents the culmination of this line of research by proposing a truly unified framework that learns embeddings for rules themselves. This goes beyond using rules as mere regularizers or guides; it treats rules as first-class citizens in the embedding space. This deep integration allows for more nuanced soft logical inference and directly contributes to the interpretability of KGEs, a persistent challenge in the field.

Patterns and Tensions

A recurring tension in this area is the trade-off between the strictness of logical adherence and the flexibility to capture exceptions or uncertainties. Hard rules are precise but brittle

and scarce; soft rules are more abundant and robust to noise but introduce ambiguity. Methods like RUGE (13) and RulE (44) explicitly embrace soft rules, acknowledging that real-world knowledge is often uncertain.

Another significant challenge is the **acquisition and quality of rules**. While these methods demonstrate the value of rules, they often assume the availability of a set of rules, whether manually curated or automatically extracted. The process of extracting high-quality, non-redundant, and consistent rules from large KGs is a research area in itself and remains a practical bottleneck. The field implicitly assumes that the provided rules are largely correct or that the models can robustly handle some level of noise within them.

The **scalability** of rule integration is another critical pattern. Early methods might struggle with large numbers of rules or entities. (20) directly addresses this by designing a mechanism where rule learning complexity is independent of the entity set size, showcasing a methodological trend towards optimizing for real-world graph scales.

In conclusion, rule-based and constraint-driven KGE approaches offer a powerful means to inject prior logical knowledge into embedding models, addressing key limitations of purely data-driven methods. From enforcing semantic smoothness (97) and simple structural constraints (90) to iteratively guiding embeddings with soft rules (13) and learning explicit rule embeddings (44), these methods collectively enhance reasoning capabilities, improve interpretability, and ensure semantic consistency. While challenges remain in rule acquisition and computational scalability, the trajectory of this research demonstrates a clear commitment to building more robust, semantically coherent, and logically sound knowledge representations.

4.3   Multi-modal and Cross-domain KGE

Building on the discussion of integrating logical rules and structural constraints, which enrich Knowledge Graph Embeddings (KGEs) with explicit consistency and reasoning capabilities, this subsection explores a complementary paradigm: leveraging multi-modal and cross-domain information. While rule-based methods inject logical structure, multi-

modal and cross-domain KGEs aim to enrich representations with diverse semantic content from external sources, such as textual descriptions, visual features, or data from heterogeneous domains. This approach is crucial for addressing data sparsity, enhancing semantic understanding, and enabling more comprehensive knowledge comprehension by leveraging complementary information that is often absent in the sparse, symbolic graph structure alone (48).

The integration of multi-modal and cross-domain data into KGE models represents a significant step towards grounding abstract knowledge in real-world observations and connecting disparate information silos. This paradigm moves beyond the limitations of relying solely on graph structure, which can be incomplete or lack fine-grained semantic detail, by drawing on the richness of natural language, visual cues, or domain-specific auxiliary data. Three main categories of approaches emerge:

1. **Text-Enhanced KGE**: Models that integrate textual descriptions of entities and relations to enrich their semantic representations.

2. **Cross-Domain KGE**: Approaches designed to learn embeddings that capture relationships and interactions across different knowledge domains, often for specific applications like recommendation.

3. **Domain-Specific Multi-modal Reasoning**: Models that combine various data modalities within a specialized domain to facilitate complex reasoning and knowledge discovery.

Text-Enhanced Knowledge Graph Embedding

The most common form of multi-modal integration in KGE involves leveraging textual descriptions associated with entities and relations. These descriptions provide rich semantic context that can resolve ambiguities and enrich embeddings, especially for sparse entities.

SSP: Semantic Space Projection for Knowledge Graph Embedding with Text Descriptions (95) is an early and influential work that addresses the "weak-

semantic" nature of purely geometric KGE models, which often struggle to capture nuanced meanings from symbolic triples alone.

- **Problem Solved**: SSP aims to overcome the semantic limitations of traditional KGEs by jointly learning from symbolic triples and textual descriptions, thereby providing more precise semantic embeddings. It addresses the challenge of integrating heterogeneous data sources (structured triples and unstructured text).

- **Core Innovation**: SSP proposes a semantic space projection model that builds an interaction between symbolic triples and textual descriptions. It uses textual descriptions to discover semantic relevance and project information into a shared semantic space. This allows the model to leverage the rich contextual information present in text to refine entity and relation representations.

- **Conditions for Success**: SSP demonstrates substantial improvements in knowledge graph completion and entity classification on benchmark datasets. Its success is contingent on the availability and quality of textual descriptions for entities and relations. The model performs well when text provides complementary semantic information that is not easily inferable from the graph structure.

- **Theoretical Limitations**: While effective, SSP's projection mechanism might be limited in capturing highly complex, multi-faceted semantic relationships compared to more advanced language models. The interaction between symbolic and textual information is modeled through a relatively simple projection, which might not fully exploit deep semantic patterns.

- **Practical Limitations**: The quality of the textual descriptions is paramount; noisy or irrelevant text can degrade performance. The model's reliance on pre-computed text embeddings (e.g., from word2vec) means its semantic understanding is bounded by the capabilities of those embeddings.

- **Comparison to Alternatives**: Compared to purely structural KGEs (e.g., (78) TransD, (35) TorusE, (114) CyclE), SSP provides a significant semantic boost by

75

grounding embeddings in natural language. It represents an early step in multi-modal KGE, preceding the widespread adoption of large pre-trained language models (PLMs).

More recently, the advent of powerful pre-trained language models has revolutionized text-enhanced KGE. **Joint Language Semantic and Structure Embedding for Knowledge Graph Completion** (22) exemplifies this by integrating deep language semantics with structural information.

- **Problem Solved**: This work addresses the need for more robust KGEs, particularly in low-resource settings, by effectively combining the rich semantics from natural language descriptions with the structural information of knowledge graphs. It aims to overcome data sparsity by leveraging external textual context.

- **Core Innovation**: The method fine-tunes pre-trained language models (PLMs) with a probabilistic structured loss. The forward pass of the PLM captures deep semantics from natural language descriptions of knowledge triplets, while the structured loss ensures the reconstruction of graph structure. This joint learning paradigm allows for a synergistic capture of both semantic and structural cues.

- **Conditions for Success**: The model achieves state-of-the-art performance on various knowledge graph benchmarks, showing significant improvements, especially in low-resource regimes. Its success relies on the powerful semantic understanding capabilities of PLMs and the ability to effectively fine-tune them for the KGE task.

- **Theoretical Limitations**: The computational cost and memory requirements of fine-tuning large PLMs can be substantial, limiting scalability for extremely large KGs or resource-constrained environments. The effectiveness is also tied to the quality of the PLM and the relevance of its pre-training data to the KG domain.

- **Practical Limitations**: Requires significant computational resources (GPUs) for training. The model's performance can be sensitive to hyperparameter tuning for PLM fine-tuning.

- **Comparison to Alternatives**: This approach represents a significant advancement over earlier text-enhanced methods like SSP (95). While SSP uses simpler text embeddings and projection, (22) leverages the deep, contextualized semantic understanding of PLMs, leading to superior performance, particularly in capturing nuanced meanings and handling data sparsity. This reflects the broader field's shift towards integrating advanced deep learning models for semantic enrichment.

Cross-Domain and Domain-Specific Multi-modal Reasoning

Beyond general text enhancement, KGEs are increasingly adapted to integrate diverse data types across domains or within specialized fields, enabling complex applications.

Cross-Domain Knowledge Graph Chiasmal Embedding for Multi-Domain Item-Item Recommendation (102) addresses critical challenges in recommender systems, namely cross-domain cold start and the provision of multi-domain recommendations, which traditional systems often struggle with due to data sparsity and domain heterogeneity.

- **Problem Solved**: This paper aims to efficiently model associations and interactions between items across diverse domains to enable multi-domain item-item (I2I) recommendations. It tackles the challenge of integrating information from different domains to enrich item representations.

- **Core Innovation**: A "cross-domain knowledge graph chiasmal embedding" approach is proposed, which efficiently interacts all items in multiple domains. To achieve both homo-domain (within-domain) and hetero-domain (cross-domain) embeddings, a novel "binding rule" is introduced. This allows the model to learn shared and domain-specific representations simultaneously, facilitating link prediction for I2I recommendations.

- **Conditions for Success**: The method achieves better link prediction and multi-domain recommendation results on two benchmark datasets. Its success depends on the existence of a knowledge graph that contains rich information about items and

their associations, both within and across domains. The "binding rule" is crucial for effectively managing the interaction between different domains.

- **Theoretical Limitations**: The effectiveness of the chiasmal embedding and binding rule is predicated on the assumption that meaningful cross-domain links or shared entities exist within the knowledge graph. If domains are too disparate with few bridging connections, the cross-domain interaction might be limited. The model's focus on I2I recommendation might not directly translate to other cross-domain tasks.

- **Practical Limitations**: Constructing a comprehensive cross-domain knowledge graph can be a significant data engineering challenge. The model's performance is sensitive to the quality and density of cross-domain links.

- **Comparison to Alternatives**: Unlike general KGEs that focus on single-domain link prediction, this work specifically targets the complex multi-domain recommendation problem. It offers a more integrated approach than simply training separate KGEs for each domain, by explicitly modeling cross-domain interactions through its chiasmal embedding and binding rule.

In specialized domains, multi-modal KGE can drive knowledge discovery. **Multi-modal reasoning based on knowledge graph embedding for specific diseases** (32) showcases this in the biomedical field.

- **Problem Solved**: This work addresses the challenge of discovering new and reliable knowledge from existing biomedical KGs, particularly for specific diseases, by leveraging multimodal reasoning. It aims to provide universal pre-trained knowledge for specific disease fields.

- **Core Innovation**: The paper develops a process for constructing Specific Disease Knowledge Graphs (SDKGs) and implements multimodal reasoning using reverse-hyperplane projection. This approach integrates structural, category, and description embeddings, allowing for a richer representation of biomedical entities and re-

78

lations. The combination of these modalities leads to the discovery of new, reliable drug-gene, gene-disease, and disease-drug pairs.

- **Conditions for Success**: Multimodal reasoning improves pre-existing models on all SDKGs using entity prediction tasks. The model's reliability in discovering new knowledge is verified through manual proofreading. Its success is highly dependent on the availability of structured biomedical data, entity categories, and textual descriptions.

- **Theoretical Limitations**: The reverse-hyperplane projection, while effective, might have limitations in capturing highly complex, non-linear interactions across diverse modalities compared to more advanced neural fusion techniques. The generalizability of the SDKG construction process to vastly different disease areas might require adaptation.

- **Practical Limitations**: The construction of high-quality SDKGs, including entity linking and relation linking, is a labor-intensive and domain-specific task. The manual proofreading for validation, while robust, is not scalable for large-scale knowledge discovery.

- **Comparison to Alternatives**: This model stands out by explicitly focusing on domain-specific multimodal reasoning, contrasting with general-purpose KGEs. It moves beyond purely structural or text-only approaches by integrating multiple distinct modalities (structure, category, description) to achieve more comprehensive knowledge discovery in a high-stakes domain.

Comparative Framework and Evolution

The evolution of multi-modal and cross-domain KGE demonstrates a clear trajectory from augmenting structural embeddings with simple textual features to leveraging sophisticated language models and integrating diverse data types for complex, application-specific reasoning (Table

79

Table 8: Comparison of Multi-modal and Cross-domain KGE Approaches

Model Family	Key Papers	Primary Modalities/Data	Core Mechanism	Problem Addressed	Key Advantage/Limitation
Text-Enhanced (Projection)	SSP [95]	Structural triples, text descriptions	Semantic space projection, interaction between sources	Weak-semantic KGEs, basic text integration	**Advantage:** Semantic enrichment. **Limitation:** Simpler text embeddings, limited deep semantic capture.
Text-Enhanced (PLM-based)	[22]	Structured triples, natural language descriptions	Fine-tuning PLMs with probabilistic structured loss	Data sparsity, deep semantic capture, low-resource settings	**Advantage:** State-of-the-art semantics, robust to sparsity. **Limitation:** High computational cost, PLM dependency.
Cross-Domain KGE	[102]	Multi-domain item KGs	Chiasmal embedding, binding rule for homo/hetero-domain interaction	Cross-domain cold start, multi-domain recommendation	**Advantage:** Handles domain heterogeneity. **Limitation:** Requires cross-domain links, specific to I2I rec.
Domain-Specific Multi-modal	[32]	Structure, category, description embeddings (biomedical)	Reverse-hyperplane projection, multimodal fusion	Knowledge discovery in specific diseases	**Advantage:** Deep domain understanding, new knowledge discovery. **Limitation:** Domain-specific, data engineering overhead.

Connections Across Papers and Broader Themes

The progression from **SSP** (95) to **Joint Language Semantic and Structure Embedding** (22) vividly illustrates the rapid advancement in leveraging textual information. While SSP uses a projection mechanism to integrate text embeddings, (22) capitalizes on the deep contextual understanding offered by pre-trained language models, fine-tuning them to simultaneously capture language semantics and structural consistency. This shift reflects a broader trend in AI towards integrating powerful, pre-trained foundation models into specialized tasks, moving from simpler feature engineering to complex, end-to-end learning.

These multi-modal and cross-domain approaches offer a distinct and complementary strategy to the rule-based and constraint-driven KGEs discussed in the previous subsection. While rule-based methods (e.g., (13) RUGE, (44) RulE) inject *logical consistency* and *reasoning capabilities* by enforcing explicit patterns, multi-modal methods like (95) SSP and (22) provide *semantic richness* and *contextual understanding* by drawing from diverse external data sources. For instance, a rule might state "if A is a parent of B, then B is a child of A," but a textual description might explain *what* A and B are, adding depth to their representations. This highlights that both logical and semantic enrichments are vital for comprehensive knowledge understanding.

Furthermore, the work on **Cross-Domain KGE Chiasmal Embedding** (102) and **Multimodal reasoning for specific diseases** (32) showcases the versatility and practical impact of KGE in complex real-world scenarios. (102) addresses the pragmatic problem of recommendation across disparate domains, which is a common challenge in e-commerce and content platforms. (32), on the other hand, applies multimodal KGE for scientific discovery in a high-stakes domain, demonstrating how integrating structural, categorical, and descriptive information can lead to verifiable new knowledge. Both papers exemplify how KGE moves beyond generic link prediction to solve specific, high-value problems by

adapting to the inherent heterogeneity and richness of real-world data.

Patterns and Tensions

A recurring pattern across these multi-modal and cross-domain approaches is their ability to significantly mitigate the **data sparsity problem**, a pervasive limitation of purely structural KGEs. By leveraging external information, even sparsely connected entities or relations can acquire rich representations. However, this comes with a critical tension: the **complexity of fusing heterogeneous information**. Effectively combining textual, structural, categorical, and potentially visual data requires sophisticated joint learning frameworks, which often increase model complexity and computational cost. For instance, fine-tuning PLMs in (22) is computationally intensive, contrasting with the relative efficiency of simpler translational models.

Another unstated assumption in many multi-modal KGEs is the **quality and alignment of the auxiliary data**. If textual descriptions are noisy, outdated, or misaligned with the symbolic facts, they can introduce errors rather than improvements. Similarly, for cross-domain KGEs, the existence of meaningful bridging entities or relations is crucial; without them, the "chiasmal embedding" of (102) might struggle to find relevant connections. This highlights a practical limitation: the performance of these models is often bounded by the quality of the *least reliable* modality or domain.

The field also grapples with the **interpretability** of fused multi-modal representations. While individual modalities might be interpretable (e.g., text descriptions), understanding *how* different modalities interact and contribute to a final embedding can be challenging, especially with deep learning fusion mechanisms. This is a persistent challenge that connects to broader themes of explainable AI, particularly relevant in high-stakes applications like biomedical knowledge discovery (32).

In conclusion, multi-modal and cross-domain KGEs represent a powerful evolution in knowledge representation, moving beyond the confines of symbolic graph structures to embrace the richness of diverse information sources. From projecting text into semantic spaces (95) to fine-tuning large language models for joint semantic and structural un-

derstanding (22), and from enabling cross-domain recommendations (102) to facilitating multimodal reasoning for specific diseases (32), these approaches consistently demonstrate improved expressiveness and robustness. While challenges remain in data quality, fusion complexity, and interpretability, the trajectory of this research points towards increasingly comprehensive and context-aware knowledge understanding, crucial for the next generation of intelligent systems.

5 Dynamic, Inductive, and Distributed KGE

The preceding section, "Enriching KGE: Auxiliary Information, Rules, and Multi-modality," demonstrated how integrating diverse external information and explicit logical constraints can significantly enhance the expressiveness and reasoning capabilities of Knowledge Graph Embeddings (KGEs). However, these advancements largely operated under the implicit assumption of a centralized, static, or quasi-static knowledge graph. In stark contrast, real-world knowledge graphs are inherently dynamic, constantly evolving with new facts, entities, and relations, often distributed across various data sources, and requiring continuous updates. This dynamic nature introduces critical challenges that necessitate a fundamental shift in KGE model design.

This section, "Dynamic, Inductive, and Distributed KGE," addresses these pivotal challenges, moving beyond static and centralized assumptions to explore how KGE models can be made adaptable, scalable, and secure in complex, operational environments. We delve into three interconnected areas. Firstly, we examine *Temporal Knowledge Graph Embedding (TKGE)*, which focuses on modeling the temporal dynamics of facts and entities, capturing their evolution and validity over time (? ?). Secondly, we explore *Inductive and Continual KGE*, tackling the crucial problem of learning embeddings for unseen entities and relations (inductive learning) and efficiently updating models with new information without full retraining (continual learning), thereby enabling KGEs to adapt to the ever-growing nature of knowledge bases (? ?). Finally, we investigate *Federated and Privacy-Preserving KGE*, addressing the complexities of collaborative learning across distributed and privacy-sensitive data sources, ensuring secure and efficient knowledge

sharing in decentralized settings (? ?). The overarching goal is to equip KGE models with the necessary mechanisms to thrive in the fluid, decentralized, and privacy-conscious landscape of modern knowledge management systems.

5.1 Temporal Knowledge Graph Embedding (TKGE)

The dynamic nature of real-world knowledge graphs (KGs), where facts evolve, appear, and disappear over time, necessitates embedding models capable of capturing these temporal dynamics. Moving beyond the static representations discussed in previous sections, Temporal Knowledge Graph Embedding (TKGE) models are specifically designed to represent entities and relations as they change through time, enabling reasoning over temporal sequences and understanding the evolution of facts. This subsection details the progression of TKGE, from early explicit temporal modeling to advanced geometric transformations and multi-curvature spaces, addressing the complexities of dynamic, spatiotemporal, and fuzzy knowledge.

The evolution of TKGE research can be broadly categorized into three main methodological families: (1) approaches that explicitly structure and model time, often through tensor representations or time series analysis; (2) methods leveraging geometric transformations like rotations to capture temporal evolution; and (3) advanced models employing multi-curvature spaces or Graph Neural Networks (GNNs) for more intricate temporal and spatiotemporal dynamics.

Explicit Temporal Modeling and Tensor/Time Series Approaches

Early efforts in TKGE focused on explicitly integrating time into the embedding space. (2) introduced **HyTE** (Hyperplane-based Temporally aware Knowledge Graph Embedding), a pioneering model that addresses the problem of static KGEs ignoring temporal validity. HyTE's core innovation is to associate each timestamp with a corresponding hyperplane in the entity-relation embedding space. This allows for temporal guidance during KG inference and the prediction of temporal scopes for facts with missing time annotations. HyTE succeeds under conditions where temporal information can be discretely

83

assigned to hyperplanes and offers a more expressive temporal representation than purely static models. However, its theoretical limitation lies in its potential inability to capture highly complex, non-linear temporal dependencies or continuous temporal evolution, as hyperplanes might oversimplify intricate temporal patterns. Practically, managing and learning distinct hyperplanes for a vast number of timestamps can introduce computational overhead, especially for fine-grained temporal data.

Building on the concept of explicit temporal integration, (79) proposed a **Tensor Decomposition-Based Temporal Knowledge Graph Embedding** model. This approach tackles the problem of sparse and time-evolving data by regarding the entire fact set as a fourth-order tensor (head, relation, tail, time). The core innovation is to apply tensor decomposition techniques (e.g., CP or TuckER decomposition extended to four dimensions) to learn dense, low-dimensional vectors for entities, relations, and time. This method claims to generalize well to other static tensor-based KGEs and demonstrates superior performance on temporal datasets. It succeeds in scenarios where temporal data can be naturally represented as a discrete dimension within a tensor. A theoretical limitation is that tensor decomposition often assumes a fixed dimensionality for time, which might not be ideal for continuously evolving or irregularly sampled temporal data. Practical limitations include the high computational cost and memory requirements associated with higher-order tensor operations, particularly for very large and dense temporal KGs, which can hinder scalability.

Addressing the challenge of temporal uncertainty and continuous evolution, (8) introduced **ATiSE** (Temporal Knowledge Graph Embedding Model based on Additive Time Series Decomposition). ATiSE's innovation lies in incorporating time using additive time series decomposition and mapping representations into multi-dimensional Gaussian distributions. The mean of each entity/relation embedding at a time step represents its expected position, while its covariance captures temporal uncertainty. This probabilistic approach claims to achieve state-of-the-art results on link prediction by accounting for the inherent fuzziness of temporal evolution. ATiSE excels when temporal uncertainty is a significant factor, providing a more robust representation than deterministic models.

However, its theoretical limitation is the assumption of additive time series decomposition, which might not hold for all complex, non-linear temporal dynamics. Practically, learning and managing Gaussian distributions for all entities and relations across time can be computationally intensive, and the interpretability of learned covariances might be challenging.

More recently, (109) presented **TeAST** (Temporal Knowledge Graph Embedding via Archimedean Spiral Timeline), which offers a novel way to model temporal relations. TeAST addresses the limitations of existing methods that fuse temporal information into entities, potentially leading to entity information evolution and limiting link prediction performance. Its core innovation is mapping relations onto an Archimedean spiral timeline, transforming the quadruple completion problem into a 3rd-order tensor completion problem. The spiral timeline ensures that simultaneously occurring relations are placed on the same timeline, and all relations evolve orderly over time with a spiral regularizer. TeAST claims to encode various relation patterns and provide interpretability. It succeeds in scenarios where a continuous, orderly temporal progression can be assumed. A theoretical limitation is that the Archimedean spiral might impose a specific, potentially restrictive, structure on temporal evolution that may not generalize to all types of temporal dynamics. Practically, the complexity of implementing and optimizing the spiral timeline and its regularizer adds to the model's intricacy.

Geometric Transformation-based Temporal Embeddings

A significant paradigm shift in TKGE involves leveraging geometric transformations, particularly rotations, to model temporal dynamics. (98) proposed **TeRo** (A Time-aware Knowledge Graph Embedding via Temporal Rotation), which defines the temporal evolution of entity embeddings as a rotation from an initial time to the current time in a complex vector space. TeRo's innovation lies in using dual complex embeddings for relations involving time intervals (beginning and end), allowing it to capture rich interaction between temporal and multi-relational characteristics. TeRo claims to overcome limitations of existing models and infer various relation patterns over time, outperforming

85

state-of-the-art models for temporal link prediction. It performs well on datasets with clear temporal sequences and where relational patterns like symmetry and inversion are prominent. However, the interpretability of complex rotations in high-dimensional spaces can be a theoretical limitation, and the computational cost of learning these transformations can be high.

Building on TeRo's insights, (105) introduced **ChronoR** (Rotation Based Temporal Knowledge Graph Embedding). ChronoR extends the concept by learning a k-dimensional rotation transformation, parameterized by both relation and time, to transform a head entity to fall near its tail entity. This approach, using high-dimensional rotation, aims to capture richer interactions between temporal and multi-relational characteristics. ChronoR demonstrates superior performance on benchmark datasets for temporal link prediction. While both TeRo and ChronoR focus on rotation, ChronoR's use of k-dimensional rotations offers potentially greater flexibility in modeling complex interactions. A common tension across these rotational models is balancing the expressiveness gained from complex rotations with the increased computational complexity and potential difficulty in interpreting the learned transformations.

More recently, the geometric transformation paradigm has been extended to address additional complexities like fuzziness and spatiotemporal dimensions. (111) presented **FSTRE** (Fuzzy Spatiotemporal RDF Knowledge Graph Embedding Using Uncertain Dynamic Vector Projection and Rotation). FSTRE addresses the problem of existing KGE models being insufficient for uncertain and dynamic knowledge. Its core innovation is to embed spatial and temporal information by projection and rotation within a complex vector space, while introducing fine-grained fuzziness through modal lengths of anisotropic vectors. This allows FSTRE to capture rich interactions between crisp/static and fuzzy spatiotemporal knowledge. FSTRE succeeds in scenarios where both spatial and temporal dynamics are present, along with inherent uncertainty. A theoretical limitation is the complexity of simultaneously modeling fuzziness, spatial, and temporal dimensions within a single geometric framework, which might lead to intricate optimization landscapes.

Further advancing this, (111) proposed a **Multihop Fuzzy Spatiotemporal RDF**

Knowledge Graph Query via Quaternion Embedding. This model tackles the significant challenge of multihop querying on *incomplete* fuzzy spatiotemporal KGs, where previous embedding-based approaches overlooked uncertainty and spatiotemporal sensitivity during reasoning. Its innovation lies in using quaternions to jointly embed spatiotemporal entities and represent relations as rotations from subject to object. Uncertainty is incorporated via the scoring function's bias factor, and the non-commutative compositional pattern of quaternions is exploited for more accurate multihop path reasoning. This approach offers a powerful algebraic structure for representing rotations and compositions, beneficial for complex path queries. However, the algebraic complexity of quaternions can lead to increased computational demands and potentially reduced interpretability compared to simpler vector spaces.

Multi-Curvature Space Embeddings and GNN-based Approaches

Recent advancements in TKGE have pushed towards even more sophisticated geometric and architectural solutions. (93) introduced **MADE** (Multicurvature Adaptive Embedding for Temporal Knowledge Graph Completion), which addresses the challenge of embedding TKGs with complex, interwoven geometric structures (hierarchical, ring, chain) that are poorly captured by single Euclidean spaces. MADE's core innovation is modeling TKGs in multi-curvature spaces (Euclidean, hyperbolic, hyperspherical) with an adaptive weighting mechanism. This mechanism assigns different weights to curvature spaces in a data-driven manner, strengthening ideal spaces and weakening inappropriate ones. It also includes a quadruplet distributor for information interaction and an innovative temporal regularization for timestamp smoothness. MADE claims to outperform existing state-of-the-art TKGC models by adaptively leveraging the strengths of different geometries. It succeeds when TKGs exhibit diverse underlying geometric patterns. A theoretical limitation is the increased complexity of optimizing embeddings across multiple, potentially disparate, geometric spaces. Practically, the computational overhead of managing different curvatures and the adaptive weighting mechanism can be substantial.

Concurrently, (93) proposed **IME** (Integrating Multi-curvature Shared and Specific

Embedding for Temporal Knowledge Graph Completion), which also models TKGs in multi-curvature spaces. IME addresses the problem that existing TKGC methods often neglect the heterogeneity of different curvature spaces. Its innovation lies in incorporating "space-shared" properties to learn commonalities across spaces and alleviate spatial gaps, and "space-specific" properties to capture characteristic features. IME also introduces an Adjustable Multi-curvature Pooling (AMP) approach and designs specific similarity, difference, and structure loss functions. IME, like MADE, achieves state-of-the-art results by leveraging the power of hybrid geometries. While both MADE and IME utilize multi-curvature spaces, MADE's adaptive weighting offers a more dynamic selection of geometries, whereas IME focuses on explicit shared/specific properties and tailored loss functions. A common practical limitation for both is the increased complexity in model design and training, making them potentially harder to implement and optimize than simpler models. The interpretability of embeddings in such hybrid spaces also remains a significant challenge.

Beyond geometric spaces, Graph Neural Networks (GNNs) have also been adapted for temporal KGE. [27] presented **TARGAT** (A Time-Aware Relational Graph Attention Model for Temporal Knowledge Graph Embedding). TARGAT addresses the limitation of previous GNN-based models that struggle to directly capture multi-fact interactions at different timestamps. Its core innovation is a relational generator that dynamically produces time-aware relational message transformation matrices, which jointly model relations and timestamp information. These matrices project neighborhood features into different time-aware spaces for aggregation, followed by a temporal transformer classifier for query quadruples. TARGAT claims to significantly outperform prior GNN-based models and achieve state-of-the-art results on benchmark datasets. It succeeds in capturing complex, localized interactions within the graph structure over time. A practical limitation is the scalability of GNNs to extremely large and dense TKGs, as message passing can be computationally intensive. Furthermore, while it captures interactions, the interpretability of the learned time-aware transformations might still be a challenge compared to more geometrically intuitive models.

Comparative Framework and Evolution

The intellectual trajectory of TKGE research demonstrates a clear evolution from explicitly modeling time as an additional dimension to integrating it deeply within complex geometric and algebraic structures. Early models like **HyTE** (2) and **Tensor Decomposition** (79) established the feasibility of temporal awareness, treating time as a distinct, often discrete, variable. This laid the groundwork for more sophisticated approaches. **ATiSE** (8) and **TeAST** (109) then introduced probabilistic and structured timeline representations, respectively, to better handle temporal uncertainty and continuous evolution, moving beyond simple discrete timestamps.

A significant shift occurred with the advent of **geometric transformation-based models** like **TeRo** (98) and **ChronoR** (105). These models evolved from the static rotational KGEs (e.g., RotatE (77) from Section 2.2) by parameterizing rotations with time, allowing entities and relations to evolve dynamically. This family excels at capturing complex relational patterns (symmetry, inversion, composition) as they change over time. However, these models often assume crisp temporal facts. The most recent advancements, such as **FSTRE** (111) and the **Quaternion Embedding** approach (111), further extend this by incorporating fuzziness and spatiotemporal dimensions, acknowledging the inherent uncertainty and geographical context of real-world knowledge. This exemplifies a broader theme in KGE research: the move from simplified, crisp data assumptions to models robust enough for noisy, multi-faceted real-world data.

The latest frontier, represented by **MADE** (93) and **IME** (93), addresses a fundamental geometric limitation. While earlier models focused on *how* entities change over time, these multi-curvature models tackle *where* they are best represented. They recognize that a single Euclidean space is insufficient for the diverse geometric structures (hierarchical, chain-like, ring-like) present in dynamic KGs. By adaptively combining hyperbolic, hyperspherical, and Euclidean spaces, they aim to capture these intricate structures more accurately, pushing the boundaries of representational capacity. This directly addresses the limitations of single-space embeddings, a recurring critique in KGE research as highlighted in (18).

TARGAT (27), on the other hand, represents a divergence in methodology, leveraging the power of GNNs. While many KGE models operate on graph data, TARGAT explicitly uses attention-based message passing to model multi-fact interactions across timestamps. This contrasts with the more mathematically constrained geometric approaches, offering a data-driven way to learn complex temporal dependencies.

A critical tension across TKGE models lies in the trade-off between *expressiveness* and *computational complexity*. Models like MADE and IME, while highly expressive due to their multi-curvature nature, introduce significant optimization challenges and computational overhead. Similarly, quaternion-based embeddings, while powerful for composition and spatiotemporal modeling, are algebraically more complex than simpler vector operations. This echoes the efficiency concerns raised in Subsection 6.1, where models like DualDE (61) and LightKG (81) aim to mitigate the parameter explosion in static KGEs; TKGE models face similar, if not greater, challenges due to the added temporal dimension.

Another unstated assumption in many TKGE models is the availability of precise timestamp information. While HyTE can predict missing time annotations, many models rely on well-defined discrete or continuous timestamps. This assumption breaks down in scenarios with fuzzy or uncertain temporal boundaries, which FSTRE (111) and the Quaternion Embedding (111) explicitly address. The field is steadily moving towards models that can handle such ambiguities, reflecting a broader trend towards robustness against data imperfections, as discussed in Subsection 6.2.

In conclusion, TKGE research has rapidly evolved from basic temporal awareness to highly sophisticated models that integrate complex geometries, algebraic structures, and deep learning architectures. The progression shows a clear drive to move beyond static, crisp, and Euclidean representations to encompass the full fluidity, uncertainty, and multi-faceted nature of real-world knowledge. The ongoing challenge remains to balance this increasing expressiveness with computational efficiency, interpretability, and robust handling of diverse temporal data characteristics.

5.2 Inductive and Continual KGE

The previous discussion on Temporal Knowledge Graph Embedding (TKGE) highlighted the necessity of adapting KGE models to evolving facts and time-dependent relationships. Building upon this, the challenge extends to handling the emergence of entirely new entities and efficiently updating models with novel information without full retraining. This subsection delves into *Inductive* and *Continual Knowledge Graph Embedding* (KGE) methods, which are crucial for adapting KGE models to the dynamic and ever-growing nature of real-world knowledge bases. Inductive KGE focuses on embedding unseen entities, while continual KGE aims to efficiently acquire new knowledge and retain previously learned information, mitigating catastrophic forgetting and ensuring scalability.

The research in this area can be broadly categorized into three evolving methodological families: (1) neighborhood aggregation-based approaches, which leverage local graph structures for inductive inference; (2) meta-learning strategies, designed to learn transferable knowledge for new entities or dynamic updates; and (3) parameter-efficient adaptation techniques, which enable incremental learning by selectively updating model parameters.

Neighborhood Aggregation for Inductive KGE

Early efforts to tackle the inductive KGE problem focused on leveraging the local neighborhood structure of new entities. Traditional KGE models, being transductive, require all entities to be present during training, making them impractical for real-world KGs where new entities constantly emerge. (84) addressed this by proposing **Logic Attention Based Neighborhood Aggregation for Inductive Knowledge Graph Embedding**. The core problem it claims to solve is enabling KGE models to embed unseen entities by learning a generalized representation function. Its core innovation is the *Logic Attention Network (LAN)*, which aggregates information from an entity's neighbors using both rules-based and network-based attention weights. This mechanism allows the model to dynamically compute an embedding for a new entity based on its existing connections, rather than requiring a pre-learned, static vector.

Table 9: Comparative Framework for Inductive and Continual KGE Approaches
|p0.18|p0.2|p0.2|p0.2|p0.15|

Dimension	Neighborhood Aggregation ([84])	Meta-Learning (Inductive) ([14])	Meta-Learning (Dynamic/Continual) ([92])	Parameter-Efficient Adaptation ([106])
Primary Problem	Inductive KGE (unseen entities)	Inductive KGE (general unseen entities, entity embeddings)	Dynamic KGE (emerging entities in evolving KGs)	Continual KGE (new knowledge, mitigate forgetting, efficiency)
Core Innovation	Logic Attention Network (LAN) for neighbor aggregation with rules- and network-based attention.	MorsE: Learns entity-independent meta-knowledge via meta-learning to *produce* entity embeddings.	MetaHG: Meta-learning with hybrid GNN (local + global) for emerging entities.	FastKGE: Incremental LoRA (IncLoRA) for efficient, isolated new knowledge acquisition and forgetting mitigation.
Conditions for Success	New entities have existing, informative neighbors.	Diverse source KGs for meta-training; generalizable meta-knowledge.	KGs with both local and global structural patterns; dynamic nature.	Large pre-trained KGE models; new knowledge can be localized.
Theoretical Limitations	Struggles with cold-start (isolated) entities or sparse neighborhoods. Limited by rule expressiveness.	Transferability depends on source-target KG similarity; struggles with highly novel entities.	Complexity of balancing local/global info; hybrid GNNs can be intricate.	Assumption that new knowledge can be "isolated" might not always hold; low-rank approximation limits expressiveness.
Practical Limitations	Performance sensitive to neighborhood quality/density; scalability for very large neighborhoods.	High computational cost of meta-learning; complex setup.	High computational cost of meta-learning + complex GNNs; tuning.	Requires careful design of layer isolation; small overhead from LoRA.
Comparison to Alternatives	Early inductive, simpler but less general than meta-learning.	More general inductive than aggregation, but computationally heavier.	Specialized meta-learning for dynamic KGs, integrates GNNs.	Addresses both efficiency and forgetting; adapts existing models rather than learning from scratch.

LAN succeeds under conditions where new entities have a sufficiently rich and informative neighborhood structure in the graph. The rules-based attention component can also leverage explicit logical patterns, enhancing interpretability. However, a significant theoretical limitation is its reliance on the presence of neighbors; truly cold-start entities (those with no or very few connections) would still pose a challenge, as there would be insufficient information to aggregate. Practically, the performance of LAN is highly sensitive to the quality and density of the neighborhood information. Furthermore, scaling neighborhood aggregation to extremely large graphs with vast neighborhoods can introduce computational overhead, as the aggregation process might become expensive. Compared to later meta-learning approaches, LAN offers a simpler, more direct inductive mechanism but is less generalizable to entities with entirely novel structural patterns.

Meta-Learning for Inductive and Dynamic KGE

Building on the need for more generalized inductive capabilities, meta-learning emerged as a powerful paradigm. Instead of learning entity-specific embeddings, meta-learning aims to learn *transferable meta-knowledge* that can be rapidly adapted to new tasks or entities.

(14) introduced **Meta-Knowledge Transfer for Inductive Knowledge Graph Embedding** (MorsE), directly addressing the limitation of existing inductive KGE methods that could only solve inductive relation prediction, not produce embeddings for unseen entities. The core innovation of MorsE is to learn entity-independent modules that encapsulate meta-knowledge, which can then be used to generate embeddings for new entities. This is achieved through a meta-learning framework, where the model learns how to learn, rather than learning specific entity representations. MorsE claims to significantly outperform baselines for both in-KG and out-of-KG tasks in inductive settings. It succeeds when there is sufficient diversity in the source KGs used for meta-training, allowing the model to learn truly generalizable meta-knowledge. A theoretical limitation is that the quality of meta-knowledge transfer is inherently dependent on the similarity between the meta-training KGs and the target KGs; highly novel entities or graph structures might still be

challenging. Practically, meta-learning frameworks are often computationally expensive and complex to set up and train, requiring careful design of meta-tasks and optimization strategies. MorsE represents a significant step beyond neighborhood aggregation by offering a more general mechanism for cold-start entity embedding, as it doesn't strictly rely on existing neighbors but on learned transferable knowledge.

Extending meta-learning to dynamic and evolving KGs, (92) proposed **MetaHG** for **Learning Dynamic Knowledge Graph Embedding in Evolving Service Ecosystems**. This work specifically tackles the challenge of efficiently updating incremental knowledge and representing emerging entities in dynamic service ecosystems, where conventional KGE methods often struggle with spatial deformation. MetaHG's core innovation is a meta-learning strategy that incorporates both local (via a GNN layer) and potential global (via a Hypergraph Neural Network, HGNN, layer) structural information from current snapshot KGs. It initializes entity embeddings using 'in' and 'out' relationship matrices and refines them through this hybrid GNN framework. This approach aims to enhance the representation quality of unseen entities by leveraging richer structural context. MetaHG succeeds particularly well in dynamic KGs that exhibit both localized and higher-order global structural patterns, such as those found in complex service ecosystems. A theoretical limitation is the increased complexity of balancing and optimizing the contributions from both local GNN and global HGNN layers, which might lead to intricate optimization landscapes. Practically, the combination of meta-learning with a hybrid GNN framework can result in high computational costs and requires extensive hyperparameter tuning. MetaHG specializes the meta-learning paradigm for dynamic updates and emerging entities, integrating advanced GNNs to capture richer structural context, a more nuanced problem than the general inductive entity embedding addressed by (14).

Parameter-Efficient Adaptation for Continual KGE

While meta-learning offers solutions for inductive capabilities and dynamic updates, the challenge of *continual learning* in KGE involves not only acquiring new knowledge effi-

ciently but also mitigating *catastrophic forgetting* of previously learned information. This is a critical tension in dynamic KGE, as models must balance plasticity (learning new facts) with stability (retaining old facts).

([106]) addressed this dual problem with **Fast and Continual Knowledge Graph Embedding via Incremental LoRA** (FastKGE). The problem it claims to solve is the inefficiency of fine-tuning KGE models for new knowledge acquisition and the persistent issue of catastrophic forgetting. Its core innovation is the *Incremental LoRA (IncLoRA)* mechanism, inspired by parameter-efficient fine-tuning techniques from large language models. FastKGE isolates and allocates new knowledge to specific layers based on the fine-grained influence between old and new KGs. It then devises an efficient IncLoRA mechanism, which embeds these specific layers into low-rank adapters, significantly reducing the number of trainable parameters during fine-tuning. Furthermore, IncLoRA introduces adaptive rank allocation, making the LoRA aware of entity importance. FastKGE demonstrates significant reductions in training time (34-49

Comparative Framework and Evolution

The intellectual trajectory in Inductive and Continual KGE demonstrates a clear evolution from basic inductive capabilities to more sophisticated and efficient dynamic learning paradigms, as summarized in Table

The field then progressed to more generalized inductive capabilities through meta-learning. ([14])'s MorsE represents a significant leap by learning transferable meta-knowledge to *produce* entity embeddings, thereby overcoming the reliance on direct neighborhood aggregation and enabling embedding for a broader range of unseen entities. This addresses a fundamental limitation of earlier methods by shifting from direct entity embedding to learning a meta-function. Building on this, ([92])'s MetaHG further refines meta-learning for *dynamic* KGs, specializing it for emerging entities in evolving service ecosystems and integrating hybrid GNNs to capture richer structural context. This highlights a trend towards tailoring general learning paradigms to specific challenges in dynamic environ-

ments.

The most recent advancements, such as (106)'s FastKGE, tackle the critical practical challenge of *efficiency* in continual learning, a concern that meta-learning approaches, despite their power, often face due to their computational complexity. FastKGE's IncLoRA mechanism directly addresses the dual problem of efficient new knowledge acquisition and catastrophic forgetting, a recurring tension in continual learning research. This approach leverages insights from parameter-efficient fine-tuning, demonstrating a convergence of techniques from different AI domains (e.g., large language models) to solve KGE challenges.

A critical tension across this subgroup is balancing the *generality* of inductive capabilities (e.g., embedding any unseen entity) with *computational efficiency* and the effective mitigation of *catastrophic forgetting*. While meta-learning approaches like MorsE and MetaHG offer strong inductive power, they can be computationally intensive. FastKGE, on the other hand, prioritizes efficiency and forgetting mitigation for continual updates, potentially at the cost of the full inductive generality offered by meta-learning for entirely novel entities. This echoes the broader efficiency concerns discussed in Subsection 6.1, where models like (61)'s DualDE and (81)'s LightKG aim to reduce parameter counts and inference times in static settings; FastKGE extends this focus to the dynamic update scenario.

Another unstated assumption in many inductive KGE models is the presence of sufficient training data for meta-learning or a rich enough neighborhood for aggregation. This assumption can break down in extremely sparse or rapidly changing real-world KGs. The evaluation of these models often requires specialized dynamic datasets, which can be less standardized than static link prediction benchmarks, making direct comparisons challenging and potentially limiting the generalizability of reported performance gains. The field is actively working towards more robust and universally applicable solutions for these complex, real-world scenarios.

5.3 Federated and Privacy-Preserving KGE

The previous subsection on Inductive and Continual KGE underscored the necessity for KGE models to adapt to dynamic and evolving knowledge. Extending this, the increasing demand for privacy-aware AI systems, coupled with the distributed nature of real-world knowledge, has propelled research into *Federated and Privacy-Preserving Knowledge Graph Embedding* (FKGE). This paradigm allows multiple clients to collaboratively train a shared KGE model without directly sharing their sensitive local knowledge graphs, thereby addressing critical privacy concerns. However, this distributed setting introduces unique challenges related to communication efficiency, personalization for diverse client data, and security vulnerabilities, which are not typically encountered in centralized KGE training. Furthermore, specific privacy-preserving mechanisms like differential privacy and secure multi-party computation become crucial to bolster the inherent privacy of federated learning.

Communication-Efficient Federated KGE

The distributed nature of Federated Learning inherently leads to significant communication overhead, especially when dealing with large models like KGEs that involve numerous entity and relation embeddings. This challenge is exacerbated by the need for multiple communication rounds between clients and a central server. While general Federated Learning (FL) research has explored techniques like increasing local training epochs to reduce communication rounds, these often overlook the substantial size of parameters transmitted in each round.

(_feds112) (FedS) directly addresses this by proposing **Communication-Efficient Federated Knowledge Graph Embedding with Entity-Wise Top-K Sparsification**. The problem it claims to solve is the high communication cost in FKGE, which hinders its scalability and deployment. Its core innovation lies in a bidirectional communication-efficient strategy. During the upload phase, clients identify and transmit only the Top-K entity embeddings that have undergone the most significant changes, rather than sending the entire model. Similarly, during download, the server performs

Table 10: Comparative Framework for Federated and Privacy-Preserving KGE Approaches

Dimension	Communication Efficiency ((_feds112))	Personalization ((_pfedeg112))	Security (Poisoning Attack) ((69))	Privacy-Preserving Defense (DP) ((?))	Cross-Domain Alignment ((?))
Primary Problem	High communication costs in FL due to large parameters and extensive communication rounds.	Semantic disparities and heterogeneity among clients; "one-size-fits-all" global model.	Vulnerability to malicious clients injecting poisoned data to manipulate predictions.	Quantifying and defending against privacy threats (e.g., inference attacks) in FKGE.	Fusing/embedding cross-domain KGs securely without sharing raw data, especially for entity/relation alignment.
Core Innovation	FedS: Bidirectional Entity-Wise Top-K Sparsification and Intermittent Synchronization Mechanism.	PFedEG: Client-wise relation graph to learn personalized supplementary knowledge from "neighboring" clients.	Systematizes poisoning attacks; KG component inference + dynamic poisoning scheme.	DP-Flames: Differentially private FKGE with private selection and adaptive budget allocation, exploiting entity-binding sparse gradients.	FedCKE: Encrypted entity/relation alignment using vertical FL, followed by parameter-secure aggregation in horizontal FL.
Conditions for Success	KGE models where changes in entity embeddings are sparse or can be effectively approximated by top-K changes.	Clients have discernible "affinity" or semantic overlap, allowing meaningful knowledge transfer.	Attacker can infer target relations and craft poisoned data; aggregation mechanism is susceptible.	Acceptable privacy-utility trade-off; sparse gradient property of KGE holds.	Existence of shared entities/relations across domains; secure encryption/alignment protocols.
Theoretical Limitations	Potential loss of precision for less significant embedding changes; risk of embedding inconsistency.	Defining "affinity" robustly; potential for personalized models to diverge too much from global consensus.	Assumes attacker has sufficient control over local training/updates; privacy mechanisms might hinder detection.	Noise addition inherently reduces utility; tight privacy accounting is complex; may not fully protect against all attack vectors.	Complexity of encrypted alignment; potential for information leakage during alignment or aggregation if not perfectly secure.
Practical Limitations	Requires careful tuning of K (top-K); overhead of sparsification/desparsification.	Increased complexity in managing client-wise graphs; computational cost of personalized aggregation.	No direct defense mechanism proposed; identifying malicious clients is hard in practice.	Tuning privacy budget (epsilon) is critical; computational overhead of DP mechanisms.	High computational cost for encryption/decryption; scalability for very large cross-domain KGs.
Comparison to Alternatives	Focuses on parameter size reduction per round, unlike general FL methods reducing rounds (e.g., multiple local epochs).	Moves beyond global model averaging, offering client-specific adaptations, unlike FedE or FedEC.	First systematic analysis of poisoning in FKGE, highlighting a novel attack vector compared to centralized KGE.	First holistic study of privacy threats and defense (DP) for FKGE, offering a better privacy-utility trade-off than naive DP.	Addresses cross-domain fusion with privacy, unlike general FKGE that assumes shared schema or simple aggregation.

personalized embedding aggregation for each client and then transmits only the Top-K aggregated embeddings back. This mechanism is complemented by an Intermittent Synchronization Mechanism designed to mitigate the negative effects of embedding inconsistency among shared entities, a common issue arising from the heterogeneity of federated KGs.

FedS succeeds under conditions where the most significant updates to entity embeddings can be captured by a sparse subset, allowing for substantial data reduction without critical information loss. Quantitative evidence from experiments on three datasets demonstrates that FedS significantly enhances communication efficiency with negligible, or even no, performance degradation. For instance, it can reduce communication by a factor of 50-100x while maintaining competitive link prediction accuracy. However, a theoretical limitation is the potential for precision loss for less significant, yet cumulatively important, embedding changes that fall outside the Top-K. This occurs because the Top-K selection implicitly assumes that smaller changes are less critical, an assumption that might not hold if many small changes collectively contribute to a significant shift in representation. Practically, determining the optimal value of K (the number of top entities to transmit) is crucial and dataset-dependent, requiring careful tuning. An overly aggressive K could lead to slower convergence or suboptimal model quality. Compared to general FL communication reduction methods that primarily focus on reducing the number of communication rounds (e.g., by increasing local computation), FedS specifically targets the *size* of the transmitted parameters within each round, offering a complementary approach. This aligns with broader efficiency concerns in KGE, similar to how (61)'s DualDE uses knowledge distillation or (30)'s Knowledge Graph Embedding Compression (LightKG) reduce parameter size in centralized settings, but FedS adapts this principle to the distributed FL environment.

Personalized Federated KGE and Cross-Domain Alignment

A fundamental tension in Federated Learning is between learning a robust global model and accommodating the inherent data heterogeneity across diverse clients. For FKGE, this

translates to semantic disparities, where different clients might have distinct knowledge patterns or even conflicting facts. Traditional federated averaging, which aims for a universal global model, often overlooks these client-specific semantic nuances, potentially leading to a "one-size-fits-all" model that performs suboptimally for individual clients.

(_pfedeg112) (PFedEG) addresses this by proposing **Personalized Federated Knowledge Graph Embedding with Client-Wise Relation Graph**. The problem it claims to solve is the compromise in embedding quality due to semantic disparities among clients and the limitations of a universal global model. Its core innovation is to move beyond a global consensus by learning personalized supplementary knowledge for each client. This is achieved through a novel "client-wise relation graph," which captures the semantic relevance or "affinity" between clients. PFedEG learns personalized supplementary knowledge for each client by amalgamating entity embeddings from its "neighboring" clients (based on their affinity) rather than simply averaging all client contributions. Each client then conducts personalized embedding learning using its local triples and this personalized supplementary knowledge.

PFedEG succeeds when there are meaningful semantic relationships or overlaps between client KGs that can be captured by the client-wise relation graph, allowing for beneficial knowledge transfer. Extensive experiments on four benchmark datasets demonstrate its superiority over state-of-the-art models, indicating that personalized aggregation significantly improves embedding quality. A theoretical limitation lies in the robust definition and dynamic adaptation of "affinity" within the client-wise relation graph; if affinity is poorly defined or static, it might not accurately reflect evolving semantic relationships, potentially leading to suboptimal personalization or even negative transfer. This is because the effectiveness of knowledge transfer is highly sensitive to the semantic alignment between source and target clients. Practically, managing and updating the client-wise relation graph adds complexity to the FKGE setup, and the computational cost of personalized aggregation can be higher than simple federated averaging. This approach directly contrasts with the implicit assumption of a homogeneous global model in many FL setups, offering a more nuanced way to handle non-IID data distributions. This

challenge is also recognized in general inductive KGE where meta-learning approaches like (14)'s MorsE aim to learn transferable knowledge across diverse tasks, but PFedEG specializes this for the federated context.

Extending the notion of personalization and heterogeneity to distinct knowledge domains, (?) proposes **Cross-domain Knowledge Graph Embedding in Federated Learning (FedCKE)**. This work tackles the challenge of securely interrelating, fusing, and embedding KGs from different domains without sharing raw data. Unlike PFedEG which focuses on semantic disparities within a broadly shared schema, FedCKE explicitly addresses cross-domain scenarios where entity/relation alignment is a primary hurdle. Its core innovation is an inter-domain encrypted entity/relation alignment method, leveraging encrypted sample alignment from vertical federated learning. This allows clients to identify shared entities/relations without revealing their full KG structures. Subsequently, a parameter-secure aggregation method from horizontal federated learning is used on the server to combine embeddings. FedCKE is particularly successful when there are common entities or relations that can serve as anchors for alignment across disparate domains. A theoretical limitation is the inherent complexity and potential computational overhead of cryptographic operations for alignment, which could limit scalability for very large KGs or a high number of clients. Practically, the reliance on encrypted alignment adds significant computational cost, making it more resource-intensive than simpler federated averaging methods. FedCKE thus offers a more robust solution for privacy-preserving *cross-domain* knowledge fusion, a problem distinct from but related to the semantic personalization addressed by PFedEG. Both methods highlight a pattern of adapting general FL principles (vertical FL for alignment, horizontal FL for aggregation) to the specific needs of KGE, demonstrating a convergence of techniques.

Security Vulnerabilities and Privacy-Preserving Defenses

While Federated Learning is designed to preserve data privacy by keeping raw data local, it does not inherently guarantee security against malicious actors. A critical, yet often unexplored, vulnerability in FKGE is the threat of poisoning attacks, where a malicious

client injects carefully crafted data to manipulate the global model's behavior or force specific false predictions.

(69) systematically explores this by presenting **Poisoning Attack on Federated Knowledge Graph Embedding**. The problem it claims to solve is the lack of understanding and systematization of poisoning attacks in FKGE, which can lead to biased decisions in downstream applications. Its core innovation is a novel framework for poisoning attacks that forces victim clients to predict specific false facts. Unlike centralized KGEs where attackers might directly inject poisoned triples, FKGE's distributed nature makes this challenging. Instead, the attacker in (69)'s framework first infers targeted relations in the victim's local KG via a "KG component inference attack." Then, to accurately mislead the victim's embeddings through the federated aggregation process, the attacker locally trains a shadow model using the poisoned data and employs an optimized dynamic poisoning scheme to adjust its model updates.

This attack framework demonstrates remarkable effectiveness, achieving success rates as high as 100

To counter such privacy threats and provide stronger guarantees, research has turned to mechanisms like Differential Privacy (DP). (?) conducts the first holistic study of privacy threats on FKGE from both attack and defense perspectives, proposing **DP-Flames**. The problem it claims to solve is quantifying and defending against inference attacks that reveal the existence of KG triples from victim clients. Its core innovation is a novel differentially private FKGE with private selection, which offers a better privacy-utility trade-off. DP-Flames exploits the entity-binding sparse gradient property of FKGE and incorporates state-of-the-art private selection techniques, coupled with an adaptive privacy budget allocation policy. This allows it to dynamically adjust defense magnitude during training. DP-Flames demonstrates its effectiveness by successfully mitigating inference attacks, reducing their success rate significantly (e.g., from 80-90

Patterns, Tensions, and Future Directions

The emerging field of Federated and Privacy-Preserving KGE is characterized by a critical tension between three key dimensions: *privacy*, *utility* (model performance and personalization), and *efficiency/security*. Solutions like FedS (_feds112) strive for communication efficiency, often by making trade-offs in the precision of transmitted information via Top-K sparsification, which could theoretically impact model utility if critical but small changes are discarded. PFedEG (_pfedeg112) pushes for personalization to enhance utility by tailoring knowledge transfer, but this adds complexity and potentially increases communication overhead for personalized aggregation. Similarly, FedCKE (?) enables cross-domain knowledge fusion with privacy, but at the cost of increased computational complexity due to encryption.

A recurring pattern is the adaptation of techniques from general FL or other AI domains to the specific challenges of KGE. For instance, sparsification for communication efficiency in FedS draws inspiration from general FL compression, while personalization strategies in PFedEG echo meta-learning's goal of transferability for diverse tasks, as seen in (92)'s MetaHG for dynamic KGE. The use of vertical FL for alignment in FedCKE further exemplifies this cross-domain technique transfer. The poisoning attack demonstrated by (69) reveals that the privacy guarantees of FL do not inherently translate to security, highlighting a fundamental vulnerability that could severely compromise model utility and trustworthiness. This systematically challenges the unstated assumption across much of the constructive FKGE research (i.e., not attack-focused) that clients are benign. In response, DP-Flames (?) introduces differential privacy as a robust defense mechanism, directly addressing the security gap exposed by poisoning attacks. This highlights a clear evolutionary trajectory: from initial focus on efficiency and basic privacy (FedS), to enhanced utility through personalization (PFedEG) and cross-domain integration (FedCKE), and finally to explicit security and quantifiable privacy guarantees (DP-Flames countering (69)'s attacks).

This field faces several open challenges. The generalizability of current FKGE solutions is an area requiring further scrutiny. While methods show promise on benchmark datasets,

their performance in highly heterogeneous, real-world distributed KGs with varying data quality and client capabilities remains to be fully evaluated. The trade-off between privacy guarantees (e.g., epsilon in DP) and model utility needs more sophisticated theoretical and empirical analysis, especially for complex KGE models. Furthermore, the field needs to move beyond simple link prediction metrics to assess the impact of FKGE on downstream applications, similar to the broader KGE field's shift towards application-specific evaluations discussed in Subsection 7.3.

In conclusion, Federated and Privacy-Preserving KGE represents a crucial frontier for leveraging decentralized knowledge while respecting privacy. The initial solutions address communication efficiency and personalization, but the identified security vulnerabilities and the need for stronger privacy guarantees underscore the need for a holistic approach that integrates robust defense mechanisms. Future research must focus on developing FKGE frameworks that can simultaneously optimize for privacy, utility, efficiency, and security, ensuring that collaborative KGE models are not only effective but also trustworthy and resilient in real-world, privacy-sensitive environments. This will likely involve more sophisticated privacy-preserving techniques beyond simple non-sharing of raw data, and adaptive personalization strategies that dynamically adjust to evolving client needs and data distributions.

6   Practical Considerations: Efficiency, Robustness, and Evaluation

While preceding sections have explored the theoretical foundations, advanced architectures, and adaptive capabilities for dynamic and distributed Knowledge Graph Embeddings (KGEs), their true utility hinges on practical viability in real-world scenarios. This section shifts focus from theoretical advancements to the critical operational challenges encountered when deploying and evaluating KGE models. It addresses the indispensable need for models that are not only expressive but also computationally efficient, resilient to data imperfections, and rigorously validated.

The first major theme, *Efficiency, Compression, and Scalability*, delves into strategies for optimizing KGE models to handle the immense scale of modern knowledge graphs. This includes techniques for reducing computational cost, minimizing memory footprint, and enhancing training and inference speed, which are paramount for large-scale deployments (**? ? ?**). Following this, *Robustness and Training Optimization* explores methods to fortify KGE models against the inherent noise and incompleteness of real-world data. This encompasses strategies for improving data quality, optimizing negative sampling, and making training processes more stable and effective, ensuring reliable performance even under challenging conditions (**? ? ?**). Finally, *Evaluation, Benchmarking, and Reproducibility* underscores the paramount importance of standardized metrics, fair comparisons, and transparent research practices. This subsection highlights efforts to establish unified frameworks, conduct large-scale comparative studies, and promote reproducibility, which are fundamental for fostering trustworthy scientific progress and ensuring the reliable application of KGE models in diverse domains (**? ?**). Collectively, these practical considerations are vital for bridging the gap between academic research and the successful implementation of KGE technologies in industry and real-world AI systems.

6.1 Efficiency, Compression, and Scalability

As Knowledge Graph Embedding (KGE) models have grown in expressiveness and complexity, as discussed in Sections 2 and 3, the practical challenges of their deployment have become increasingly prominent. The sheer scale of real-world knowledge graphs necessitates techniques that enhance efficiency, reduce memory footprint, and ensure scalability during both training and inference. This subsection examines key innovations addressing these practical bottlenecks, ranging from model-agnostic system optimizations to model-specific compression and parameter-efficient learning strategies.

A primary driver for this research area is the inherent trade-off between model capacity (and thus, performance) and computational resource requirements. Early KGE models, while efficient, often lacked expressiveness. Modern, highly expressive models, particularly those leveraging deep learning architectures, tend to be parameter-heavy and

computationally intensive. This tension has spurred the development of several distinct, yet complementary, methodological families:

1. **Knowledge Distillation and Embedding Compression:** These approaches aim to reduce the size of the embedding layer and accelerate inference by either training a smaller model to mimic a larger one or by encoding embeddings more compactly.

2. **Parameter-Efficient Learning:** This family focuses on designing KGE models that inherently require fewer parameters, especially for entities, thus mitigating the linear growth of parameters with graph size.

3. **Algorithmic and System-Level Optimizations:** These methods tackle the computational cost of training, either through novel, inherently efficient algorithms or by optimizing the underlying system infrastructure for parallel and distributed processing.

Knowledge Distillation and Embedding Compression

One prominent strategy to achieve efficiency is *knowledge distillation*, where a smaller, "student" KGE model is trained to replicate the behavior of a larger, more complex "teacher" model. (61) introduced DualDE, a dually distilling method designed to create low-dimensional student KGEs from high-dimensional pre-trained teachers. The core innovation of DualDE lies in its dual-influence consideration between teacher and student, employing a soft label evaluation mechanism to adaptively weigh soft and hard labels for different triples, alongside a two-stage distillation process. This approach claims significant parameter reduction (7-15x) and inference speedup (2-6x) while retaining high performance, making it suitable for resource-limited applications. A theoretical limitation of distillation is that the student model's performance is inherently bounded by the teacher's capabilities, and the distillation process itself can be complex to tune. Practically, the "minor loss in performance" often associated with compression, as noted in other works, can still be a critical factor in highly sensitive downstream tasks.

Complementary to distillation are direct *embedding compression* techniques. (30) pro-

posed an approach to compress the KGE layer by representing each entity as a vector of discrete codes, from which embeddings are composed. This method achieves massive compression (50-1000x) with only a minor loss in performance, demonstrating that the information content of embeddings can be significantly condensed. Building on this, (81) introduced LightKG, a lightweight framework for efficient inference and storage. LightKG's innovation lies in storing only a few codebooks and indices, drastically reducing storage requirements. It boosts inference efficiency through quick look-ups and incorporates a novel dynamic negative sampling method based on quantization. While (30) focuses on the compression mechanism, LightKG provides a more comprehensive framework, integrating residual modules for codebook diversity and a continuous function to approximate non-differentiable codeword selection. Both methods address the problem of large memory footprints, but LightKG specifically optimizes for both storage and inference speed, making it highly practical for real-time applications. However, the reliance on discrete codes or codebooks introduces a quantization error, which is the theoretical limitation that leads to the "minor loss in performance" compared to full-precision embeddings.

Parameter-Efficient Learning

Moving beyond post-hoc compression, *parameter-efficient learning* aims to design KGE models that are inherently less parameter-intensive. (3) proposed Entity-Agnostic Representation Learning (EARL) to tackle the inefficient parameter storage costs of conventional KGE methods, where the number of embedding parameters grows linearly with the number of entities. EARL's core innovation is to learn embeddings only for a small set of "reserved entities" and derive the embeddings for all other entities from their context using universal, entity-agnostic encoders. This approach results in a static and lower parameter count, making it particularly beneficial for massive and continuously growing KGs. EARL addresses a fundamental problem in KGE scalability by decoupling entity representations from direct parameter storage. Its success hinges on the assumption that sufficient distinguishable information can be extracted from an entity's context to form a high-quality embedding, which might be a practical limitation for very sparse entities

or those with limited contextual information. Compared to distillation or compression, EARL offers a more fundamental solution to parameter explosion by rethinking how embeddings are learned, rather than how they are stored or approximated.

Algorithmic and System-Level Optimizations

Beyond reducing model size, significant efforts have been directed towards accelerating the training process itself. (65) introduced a "Highly Efficient Knowledge Graph Embedding Learning" framework that significantly reduces training time and carbon footprint by orders of magnitude. Their core innovation is a closed-form solution using Orthogonal Procrustes Analysis, enabling full-batch learning and non-negative sampling. This represents a paradigm shift from iterative optimization methods, offering inherent efficiency. A key theoretical advantage is that their entity embeddings also store full relation information, suggesting improved interpretability. However, the applicability of such closed-form solutions might be limited to specific KGE model architectures, potentially not generalizing to the full complexity of GNN-based or Transformer-based models discussed in Section 3.

For Graph Neural Network (GNN)-based KGEs, which are increasingly popular for their expressiveness, scalability remains a challenge. (94) proposed CPa-WAC (Constellation Partitioning-based Scalable Weighted Aggregation Composition) to address the training time and memory cost of GNNs on large KGs. The innovation lies in using modularity maximization-based constellation partitioning to break down KGs into subgraphs, which are then processed separately. This method aims to reduce memory and training time while crucially retaining prediction accuracy by preserving local graph topology. CPa-WAC demonstrates that efficient partitioning strategies can enable GNNs to scale to larger KGs, achieving up to five times faster training while maintaining performance comparable to training on the whole graph. A practical limitation is the complexity of effectively aggregating information from partitioned subgraphs without losing critical global context, a common challenge in graph partitioning.

Finally, *system-level optimizations* provide a general infrastructure for efficient KGE

training. (99) introduced GE2, a "General and Efficient Knowledge Graph Embedding Learning System." GE2 tackles the long CPU times and high CPU-GPU communication overhead observed in existing graph embedding systems (e.g., PBG, DGL-KE, Marius). Its core innovations include a general execution model for various negative sampling algorithms, a user-friendly API, offloading CPU operations to GPU for higher parallelism, and the novel COVER algorithm for efficient data swap between CPU and multiple GPUs. GE2 is not a KGE model itself but a foundational system that can accelerate the training of various KGE models across different datasets, achieving speedups of over 2x and up to 7.5x. This work highlights that system-level engineering is as crucial as algorithmic innovation for practical KGE deployment. GE2's generality allows it to potentially accelerate the training of models like CPa-WAC (94) or even the efficient KGE learning from (65), representing a convergent research direction where optimized algorithms benefit from optimized infrastructure.

Synthesis and Tensions

The research in efficiency, compression, and scalability reveals a clear intellectual trajectory. Early KGE models often prioritized expressiveness, leading to large memory footprints and long training times. This necessitated the development of post-hoc solutions like knowledge distillation (61) and embedding compression (30; 81) to make models deployable. These methods inherently involve a trade-off: they achieve significant resource savings but often with a "minor loss in performance," as highlighted by (30).

More recent work has shifted towards designing KGE models that are *inherently* efficient, such as parameter-efficient learning with EARL (3), which fundamentally rethinks entity representation to avoid linear parameter growth. Simultaneously, the field has seen a push for radically more efficient training algorithms, exemplified by the closed-form solution of (65), which offers orders of magnitude speedup by leveraging mathematical properties. This contrasts with engineering solutions like graph partitioning (94) that optimize existing GNN architectures for scalability.

A critical tension across these approaches is the balance between efficiency and the

ability to capture complex relational patterns. While compression and parameter effi-
ciency reduce resource demands, they must ensure that the core expressiveness of the
KGE is not unduly compromised. Furthermore, the generalizability of these efficiency
techniques is an ongoing challenge. For example, while GE2 (99) offers a general system
for accelerating training, its impact is tied to the underlying KGE model and hardware.
Similarly, CPa-WAC (94) is tailored for GNN-based KGEs, and its partitioning strategy
might not be directly applicable to all KGE architectures. The unstated assumption in
many of these works is that the "minor" performance degradation from compression or
partitioning is acceptable for the gains in efficiency, an assumption that may not hold in
all high-stakes applications. These innovations are crucial for overcoming the practical
bottlenecks of KGE, making them deployable in resource-constrained environments and
for handling ever-growing knowledge bases, thereby bridging the gap between theoretical
advancements and real-world utility.

6.2 Robustness and Training Optimization

While the preceding sections have explored the evolution of Knowledge Graph Embedding
(KGE) models towards greater expressiveness (Section 2) and architectural sophistication
(Section 3), and the critical need for efficiency and scalability (Subsection 6.1), the practi-
cal utility of these models hinges significantly on their robustness against real-world data
imperfections and the optimization of their training processes. This subsection delves into
methodologies that ensure KGE models learn accurate and reliable representations even
from noisy, incomplete, or imbalanced data, leading to more trustworthy predictions and
better generalization. The focus here is on enhancing the resilience of KGE models and
refining the learning mechanisms that underpin their performance.

Probability Calibration for Reliable Predictions

A critical, yet often overlooked, aspect of KGE model reliability is the calibration of their
probability estimates. Many KGE models output raw scores that indicate plausibility,
but these scores are frequently uncalibrated, meaning a score of, for instance, 0.8 does

not reliably correspond to an 80

Problem Solved: (85) aims to make the probability estimates associated with predicted triples reliable, especially in scenarios where ground truth negatives, crucial for direct calibration, are scarce. **Core Innovation:** The paper proposes a novel method to calibrate KGE models post-training, even without readily available ground truth negatives. It leverages established techniques like Platt scaling and isotonic regression, adapting them for the KGE context. **Conditions for Success:** The method succeeds when the underlying KGE model provides a reasonably good ranking of triples, even if its raw scores are uncalibrated. It is a model-agnostic post-processing step, applicable to a wide range of KGE architectures. **Theoretical Limitations:** Calibration is a post-hoc adjustment; it does not enhance the intrinsic discriminative power of the KGE model itself. Its effectiveness is bounded by the quality of the raw scores provided by the base model. Furthermore, the absence of true negatives for calibration introduces its own set of challenges, requiring careful proxy selection. **Practical Limitations:** While effective, applying calibration methods like isotonic regression can be computationally intensive and may require a dedicated validation set for optimal performance, which might still be challenging to obtain with reliable negative samples. **Comparison to Alternatives:** Unlike methods that directly improve embedding quality or filter noisy inputs, calibration focuses solely on the interpretability and trustworthiness of the *output scores*. It complements other robustness techniques by ensuring that the model's confidence in its predictions is well-aligned with reality. This addresses a different facet of robustness compared to the input-focused methods discussed next. **Implications:** This work highlights that traditional KGE evaluation metrics (e.g., Hit@K, MRR) do not fully capture the reliability of predictions, advocating for a more comprehensive assessment of model trustworthiness.

Robustness Against Noisy Data and Imbalance

Real-world knowledge graphs are often constructed automatically, inevitably introducing noise and conflicts. Most KGE models, however, implicitly assume clean data, leading

to suboptimal or unreliable representations when trained on imperfect KGs. Concurrently, KGs frequently exhibit long-tail distributions, where a small fraction of entities and relations are highly frequent, while the vast majority are rare, leading to imbalanced training.

Noise Filtering via Reinforcement Learning: (36) addresses the challenge of noisy KGs by proposing a novel framework. **Problem Solved:** This method tackles the degradation of KGE performance due to noisy triples and knowledge conflicts, which are prevalent in automatically constructed KGs. **Core Innovation:** It introduces a multi-task reinforcement learning (RL) framework where an RL agent learns to dynamically select high-quality triples and filter out noisy ones during the training process. To enhance efficiency and leverage semantic similarities, the triple selection processes for semantically similar relations are trained collectively using multi-task learning. **Conditions for Success:** The approach is particularly effective in scenarios where KGs are known to contain a significant amount of noise. It is general enough to be extended to popular KGE models like TransE, DistMult, ConvE, and RotatE. **Theoretical Limitations:** RL-based training can be complex, requiring careful design of reward functions and state representations. The inherent uncertainty of identifying "true" noise in a large KG makes the optimal filtering strategy difficult to guarantee. **Practical Limitations:** Integrating an RL agent significantly increases the complexity and computational cost of the training pipeline compared to standard KGE models. Hyperparameter tuning for the RL component can also be challenging and time-consuming. **Comparison to Alternatives:** This approach represents a proactive, *in-training* mechanism for robustness, directly cleaning the input data stream. This contrasts with post-hoc calibration (85) or reactive negative sampling strategies, offering a more fundamental solution to input data quality issues. **Implications:** This work signals a shift towards more adaptive and intelligent training paradigms that actively manage data quality, rather than passively accepting noisy inputs.

Weighted Training for Imbalanced Data: (96) addresses the pervasive data imbalance issue in KGs, where entities and relations follow a long-tail distribution. **Problem**

Solved: Existing KGE methods often assign equal weights to all entities and relations during training, leading to unreliable representations for infrequent (long-tail) elements due to insufficient training exposure. **Core Innovation:** The proposed method, WeightE, employs a bilevel optimization scheme. The inner level focuses on learning reliable entity and relation embeddings, while the outer level dynamically assigns differential weights to entities and relations. It prioritizes infrequent entities and relations by endowing them with higher weights, ensuring they receive adequate training. **Conditions for Success:** WeightE is particularly beneficial for KGs with pronounced long-tail distributions. Its weighting technique is designed to be general and flexible, applicable to a number of existing KGE models. **Theoretical Limitations:** Bilevel optimization, while powerful, can be computationally more demanding than single-level optimization. The convergence properties and global optimality guarantees can be complex to establish, and the optimal weighting strategy might vary significantly across different KGs. **Practical Limitations:** The added complexity of bilevel optimization can increase training time and resource consumption. Careful tuning of the outer-level optimization parameters is crucial for effective performance. **Comparison to Alternatives:** WeightE tackles a different aspect of robustness: ensuring fair and effective learning across the entire spectrum of data frequency. This complements noise filtering by addressing data distribution issues and differs from calibration, which focuses on output reliability. It is a general training optimization technique that can be integrated with various KGE architectures. **Implications:** This highlights the importance of considering data distribution characteristics in KGE training, moving beyond uniform treatment to ensure high-quality representations for all knowledge elements, which is vital for comprehensive knowledge reasoning.

The Crucial Role of Negative Sampling

The training of KGE models typically relies on contrastive learning, which necessitates discriminating between positive (true) and negative (false) triples. Since KGs predominantly store only positive facts, the generation of effective negative samples is a critical and non-trivial aspect of the training process, profoundly impacting model accuracy and

robustness. The evolution of negative sampling (NS) methods reflects a continuous effort to overcome the challenges of generating high-quality, informative negative examples.

Table 11: Comparative Framework for Negative Sampling Strategies

Feature	Confidence-Aware NS (9)	NSCaching (100)	Non-Sampling (NS-KGE) (88)	Modality-Aware NS (_modality96)
Problem Solved	Training in noisy KGs, avoiding zero loss/false detection from uniform NS.	Efficiently finding "hard" negatives, avoiding complexity of GAN-based methods.	Instability and uncertainty inherent in all sampling procedures, considering all negatives.	Adapting NS for multi-modal KGEs, aligning structural and visual embeddings.
Core Innovation	Incorporating negative triple confidence to guide sampling in noisy environments.	Cache-based tracking and sampling of "hard" negative triplets.	Mathematical derivation to consider all negative instances without explicit sampling.	Aligning structural and visual embeddings during negative sample generation.
Conditions for Success	Noisy KGs where uniform sampling leads to issues; requires confidence estimation.	Scenarios where "hard" negatives are crucial for learning but are rare.	KGE models with square-loss or convertible loss functions; requires efficient mathematical optimization.	Multi-modal KGs where diverse data modalities need semantic alignment.
Theoretical Limitations	Relies on accurate confidence estimation, which can be challenging and noisy itself.	Still a heuristic sampling method; cache management introduces its own complexities.	Limited to specific loss functions; complexity reduction is often an approximation or specific to certain model forms.	Increased complexity due to modality fusion; optimal cross-modal alignment is difficult to define.
Practical Limitations	Adds complexity to the confidence estimation process and its integration into training.	Requires careful management of cache size and balancing exploration-exploitation.	Can have high initial computational cost and memory footprint without clever optimization.	Requires multi-modal data availability; complex fusion and sampling design.
Comparison to Alternatives	More robust than uniform NS in noisy settings; less complex than GAN-based methods.	Simpler and more efficient than GAN-based NS while achieving similar benefits of hard negative mining.	Offers more stable and accurate performance by removing sampling uncertainty, but with higher initial complexity.	Extends NS to a new domain; addresses multi-modal specific challenges, unlike purely structural NS.

Evolution of Negative Sampling Strategies: Early KGE models often relied on uniform random negative sampling, which, while simple, frequently generated "easy" negatives that provided little learning signal or "false negatives" that were actually true but unobserved. This led to issues like vanishing gradients or incorrect learning.

Heuristic-based Negative Sampling: Mid-period research introduced more sophisticated heuristic-based sampling. (9) proposed a *confidence-aware negative sampling method* to address the problem of noisy KGs. **Core Innovation:** It introduces the con-

cept of negative triple confidence, allowing the training process to better handle noisy environments where uniform sampling might lead to zero loss or false detection. By assigning confidence scores to negative samples, the model can learn more effectively from them. **Comparison:** This method improves upon uniform sampling by making the negative generation process more intelligent and robust to input noise, complementing the noise filtering of (36) by focusing on the negative side of the learning objective.

Building on the observation that "hard" negative triplets (those that are plausible but false) are crucial for effective learning, (100) introduced *NSCaching*. **Core Innovation:** NSCaching efficiently tracks and samples these hard negatives using a cache. This approach aims to distill the benefits of more complex Generative Adversarial Network (GAN)-based negative sampling methods into a simpler, more efficient framework, avoiding the additional parameters and training complexities associated with GANs. **Comparison:** NSCaching offers a practical and efficient alternative to GAN-based methods, achieving comparable performance gains by focusing on the most informative negative samples without the overhead of training a generative model.

Non-Sampling Approaches and Comprehensive Reviews: A more radical departure from traditional negative sampling is presented by (88) with *Efficient Non-Sampling Knowledge Graph Embedding (NS-KGE)*. **Core Innovation:** This method proposes to avoid negative sampling entirely by considering all negative instances in the KG for model learning. It leverages mathematical derivations to reduce the computational complexity of this "non-sampling" loss function, aiming for more stable and accurate performance by removing the inherent uncertainty and suboptimality of sampling. **Comparison:** NS-KGE fundamentally challenges the assumption that sampling is necessary, offering a theoretically more sound approach by considering the full negative space. This contrasts sharply with all sampling-based methods, including those of (9) and (100), which are inherently heuristic. While it promises greater stability, its applicability is currently limited to square-loss based KGE models or those whose loss functions can be converted to a square loss.

As KGE models integrate more diverse data, negative sampling strategies also need

adaptation. (_modality96) introduces *Modality-Aware Negative Sampling* for multi-modal KGE. **Core Innovation:** This method recognizes that when incorporating multi-modal information (e.g., text, images), negative sampling must align structural and visual embeddings for entities. It ensures that negative samples are meaningful across all modalities, leading to better-aligned and more robust multi-modal embeddings. **Comparison:** This extends the negative sampling problem to a new dimension, highlighting that as KGE models become more complex (as discussed in Subsection 4.3), their training components must also evolve to handle the increased complexity.

The critical importance of negative sampling is further underscored by comprehensive reviews. (47) and (31) provide systematic categorizations and analyses of existing negative sampling approaches. **Implications:** These surveys consolidate knowledge, categorize methods into distinct families (e.g., static, dynamic, custom cluster-based), and identify open research questions, guiding future advancements in this vital area. They highlight that the "true" negative distribution remains unknown, making the design of optimal NS methods an ongoing heuristic challenge and a fundamental theoretical limitation that no current method fully resolves.

Patterns and Tensions in Training Optimization: The research in robustness and training optimization reveals several key patterns and tensions. A recurring trade-off exists between the *simplicity/efficiency* of a training method and its *robustness/accuracy*. For instance, while uniform negative sampling is simple, it often yields suboptimal results in noisy or imbalanced settings, necessitating more complex solutions like confidence-aware sampling (9) or NSCaching (100). The emergence of non-sampling approaches (88) represents a paradigm shift, challenging the long-standing assumption that negative sampling is an unavoidable necessity, but this comes with its own computational complexities that require clever mathematical derivations.

Another tension lies between *proactive* and *reactive* robustness. Methods like RL-based noise filtering (36) proactively clean the input data during training, while probability calibration (85) reactively adjusts the model's output scores. Both contribute to overall trustworthiness but address different stages of the KGE pipeline. The field implicitly

assumes that the benefits of these complex optimizations outweigh their increased computational cost and implementation complexity, a trade-off that is not always explicitly quantified across diverse application scenarios.

Collectively, these advancements in robustness and training optimization are vital for ensuring that KGE models learn accurate representations even from noisy or incomplete data, leading to more reliable predictions and better generalization. They bridge the gap between theoretically powerful KGE architectures and their practical deployment in real-world, imperfect knowledge environments.

6.3   Evaluation, Benchmarking, and Reproducibility

While the preceding subsections have explored the evolution of Knowledge Graph Embedding (KGE) models towards greater expressiveness (Section 2), architectural sophistication (Section 3), and practical considerations like efficiency and robustness (Subsections 6.1 and 6.2), the scientific integrity and real-world trustworthiness of these advancements critically depend on rigorous evaluation, standardized benchmarking, and the reproducibility of reported results. This subsection highlights the field's introspection into these crucial aspects, discussing the development of unified frameworks, large-scale comparative studies that expose biases and failures, and analyses of hyperparameter effects, all aimed at moving KGE research towards higher standards of empirical validation and transparency.

The rapid proliferation of KGE models, each often with unique implementations, training procedures, and evaluation protocols, created a significant challenge for fair comparison and hindered cumulative scientific progress. This heterogeneity led to a "reproducibility crisis," where published results were often difficult or impossible to replicate, undermining the reliability of reported performance gains.

Unified Frameworks for Reproducible Research

To address the burgeoning reproducibility crisis, the community recognized the need for standardized tools and environments. This led to the development of unified libraries

designed to facilitate consistent experimentation.

LibKGE (82):

- **Context:** The lack of a standardized, configurable, and open-source environment for KGE research hindered reproducible experimentation and systematic comparative studies.

- **Core Innovation:** (82) introduced LibKGE, an open-source PyTorch-based library for training, hyperparameter optimization (HPO), and evaluation of KGE models. Its key mechanisms include decoupled components, comprehensive logging, and the ability to fully reproduce any experiment from a single configuration file. This design allows researchers to mix and match various training methods, model architectures, and evaluation strategies.

- **Conditions for Success:** LibKGE succeeds by providing a flexible and efficient platform for KGE experimentation, particularly beneficial for researchers familiar with PyTorch. It is designed to be as efficient as possible within the Python/Numpy/PyTorch ecosystem.

- **Theoretical Limitations:** While LibKGE offers a robust framework, it is fundamentally a tool. It does not inherently solve the underlying challenges of hyperparameter sensitivity or evaluation biases; rather, it provides the infrastructure to *study* and *mitigate* these issues. Its effectiveness is contingent on the user's diligence in configuration and analysis.

- **Practical Limitations:** Despite its configurability, extensive hyperparameter tuning within LibKGE can still be computationally intensive, as highlighted by (6). Users must also invest time in understanding its architecture to fully leverage its capabilities.

- **Comparison to Alternatives:** Compared to ad-hoc, custom implementations prevalent in early KGE research, LibKGE offers a significant leap in standardization and efficiency. It serves as a foundational platform for systematic research,

complementing large-scale empirical studies like (62) by providing the means to conduct them.

- **Implications:** LibKGE represents a crucial step towards fostering a more rigorous and transparent research environment in KGE, enabling researchers to build upon each other's work with greater confidence.

Large-Scale Benchmarking and Uncovering Evaluation Biases

Beyond providing tools, empirical studies were necessary to quantify the extent of the reproducibility problem and identify biases in common evaluation practices.

Bringing Light Into the Dark (62):

- **Context:** Motivated by the widespread difficulty in reproducing published KGE results, this work aimed to empirically assess the state of reproducibility.

- **Problem Solved:** The paper addressed the problem of heterogeneous implementations and inconsistent evaluation protocols that made fair and thorough comparisons of KGE models challenging.

- **Core Innovation:** (62) conducted a massive re-implementation and evaluation effort, re-implementing 21 KGE models within the PyKEEN software package (a unified framework similar to LibKGE). They performed a large-scale benchmarking study involving thousands of experiments and 24,804 GPU hours of computation.

- **Evidence:** The study explicitly detailed which results could be reproduced with reported hyperparameters, which required alternate hyperparameters, and which could not be reproduced at all. It provided concrete insights into best practices and optimal configurations, demonstrating that the combination of model architecture, training approach, loss function, and explicit modeling of inverse relations is crucial for performance. They showed that several architectures could achieve state-of-the-art results when carefully configured.

- **Limitations:** Re-implementations, despite best efforts, might not perfectly capture the nuances of original authors' intentions. The sheer scale of HPO means that finding a global optimum is still a heuristic search. The study is retrospective, identifying problems that have already impacted prior research.

- **Comparison to Alternatives:** Unlike individual model papers, this is a critical meta-study that systematically evaluates a broad spectrum of existing models. It provides empirical evidence for the necessity of unified frameworks like LibKGE and directly informs the community about the pitfalls of current research practices.

- **Implications:** This seminal work exposed widespread reproducibility failures, prompting a critical self-reflection within the KGE community and advocating for higher standards of reporting and experimentation. It underscored the paramount importance of hyperparameter tuning and detailed training configurations.

Knowledge Graph Embedding for Link Prediction (108):

- **Context:** Despite the rapid growth of KGE literature, insufficient attention had been paid to the effect of design choices and potential biases in standard evaluation methodologies.

- **Problem Solved:** This paper critically examined the common practice of reporting accuracy by aggregating over all test facts, which can be misleading due to the vastly uneven representation of entities, allowing models to achieve good scores by focusing only on frequently occurring entities.

- **Core Innovation:** (108) provided a comprehensive comparison of 18 state-of-the-art KGE methods, extending the dimensions of analysis beyond commonly reported metrics. Their detailed analysis over popular benchmarks systematically exposed biases in standard evaluation practices, particularly the over-representation of certain entities.

- **Evidence:** The study presented detailed experimental results demonstrating how standard metrics could be skewed, highlighting that performance might not reflect

true generalization across the entire KG.

- **Limitations:** While identifying critical biases, the paper primarily critiques existing practices rather than proposing a universally accepted, unbiased metric. Its findings are based on existing benchmarks, which themselves might have inherent structural biases.

- **Comparison to Alternatives:** Similar to (62) in its comparative nature, (108) distinguishes itself by focusing more on the *validity* and *fairness* of evaluation metrics themselves, rather than just the reproducibility of reported numbers. It directly questions the interpretability of aggregated performance scores.

- **Implications:** This work issued a crucial call for more nuanced and less biased evaluation methodologies, urging researchers to move beyond simple aggregated metrics to understand model performance across different parts of the knowledge graph, especially for long-tail entities and relations.

Assessing the Effects of Hyperparameters

The findings from large-scale benchmarking efforts underscored the critical role of hyperparameters. Understanding their impact became essential for efficient research and reliable model deployment.

Assessing the effects of hyperparameters on knowledge graph embedding quality (6):

- **Context:** Hyperparameter optimization is a computationally expensive and time-consuming process, as implicitly highlighted by the extensive HPO required in (62). Not all hyperparameters contribute equally to model quality.

- **Problem Solved:** This paper addressed the lack of understanding regarding the relative importance of hyperparameters on KGE quality, which leads to inefficient HPO. It also identified and mitigated data leakage issues in common benchmarks.

- **Core Innovation:** (6) employed Sobol sensitivity analysis to quantify the importance of different hyperparameters by performing thousands of embedding trials. This allowed them to determine which hyperparameters significantly impact embedding quality variance. Crucially, they identified several relations in the UMLS knowledge graph that cause data leakage via inverse relations and derived UMLS-43, a leakage-robust variant of that graph.

- **Evidence:** The study found substantial variability in hyperparameter sensitivities across different knowledge graphs, suggesting that optimal configurations are often dataset-dependent. Their analysis revealed that some hyperparameters could be eliminated from search without significantly impacting quality, leading to more efficient HPO. The identification and correction of data leakage in UMLS-43 provided a more reliable benchmark.

- **Limitations:** Sobol sensitivity analysis, while powerful, can still be computationally intensive for very large hyperparameter spaces. The findings on hyperparameter importance are specific to the KGE models and datasets studied and may not universally generalize.

- **Comparison to Alternatives:** This paper provides a *mechanistic and quantitative understanding* of *why* reproducibility is challenging, directly complementing the empirical observations of (62). It moves beyond simply identifying reproducibility failures to providing tools and insights for more efficient and robust hyperparameter tuning. It also addresses a fundamental issue of dataset quality, which (108) touched upon regarding evaluation biases.

- **Implications:** This work guides more efficient hyperparameter optimization by allowing researchers to focus on the most impactful parameters. More significantly, it highlights the critical need for meticulous dataset curation and validation to prevent data leakage, which can artificially inflate reported performance and undermine scientific claims.

Table 12: Comparative Framework for Key Contributions to KGE Evaluation and Reproducibility

Aspect	LibKGE (82)	Bringing Light Into the Dark (62)	KGE for Link Prediction (108)	Assessing Hyperparameters (6)
Primary Goal	Enable reproducible KGE research via a unified library.	Empirically assess reproducibility and benchmark models under unified conditions.	Comprehensive comparison of methods, exposing evaluation biases.	Quantify hyperparameter importance and identify dataset issues (leakage).
Core Mechanism	PyTorch library with configurable components, logging, single config file.	Re-implementation of 21 models in PyKEEN, large-scale HPO (24,804 GPU hours).	Extensive experimental comparison of 18 SOTA methods, detailed analysis of design choices.	Sobol sensitivity analysis on thousands of trials; identified and corrected data leakage.
Key Finding(s)	Provides a flexible framework for KGE experimentation.	Widespread reproducibility failures; HPs are crucial; careful configuration matters more than architecture alone.	Standard evaluation metrics biased by entity over-representation, leading to misleading performance.	HP sensitivity varies greatly by KG; identified data leakage in UMLS-43, proposing a robust variant.
Contribution Type	**Tool/Framework** for research.	**Empirical Study/Benchmarking** of existing models.	**Critical Analysis/Benchmarking** of evaluation practices.	**Methodological Analysis** of HPO and dataset quality.
Impact on Field	Fosters reproducibility, facilitates systematic studies.	Exposed reproducibility crisis, pushed for higher standards, provided best practices.	Called for more nuanced, less biased evaluation metrics and interpretation.	Improved HPO efficiency, highlighted dataset integrity issues for reliable benchmarking.

Patterns, Tensions, and Unstated Assumptions

The collective body of work in evaluation, benchmarking, and reproducibility reveals several critical patterns and tensions within KGE research. A significant pattern is the field's maturation from solely developing new models to critically examining *how* research is conducted and evaluated. This shift towards meta-research is a hallmark of a maturing scientific discipline.

A recurring tension exists between *achieving state-of-the-art performance* and ensuring *reproducibility and transparency.* As (62) empirically demonstrated, many published results were difficult to reproduce, often due to undisclosed or highly specific hyperparameter configurations. This complexity is exacerbated by the increasingly sophisticated deep learning architectures discussed in Section 3. The efficiency concerns raised in Subsection 6.1 are directly intertwined with this, as extensive hyperparameter tuning, if not managed efficiently, can be prohibitively costly (6).

Another tension lies between *generalizability* and *dataset-specific tuning.* The optimal hyperparameters and even the presence of evaluation biases are often highly dependent on the characteristics of the specific knowledge graph (6; 108). This implies that a model performing exceptionally well on one benchmark might not generalize effectively to another, challenging the notion of universally superior KGE models.

Furthermore, these papers systematically question several unstated assumptions prevalent in earlier KGE research. Many papers implicitly assumed that standard metrics like Mean Reciprocal Rank (MRR) and Hits@K on common benchmarks were sufficient indicators of real-world utility. However, (108) directly challenged this by showing how these metrics can be biased by entity over-representation. Similarly, the implicit assumption that published results are easily reproducible was empirically disproven by (62), highlighting the need for detailed reporting and open-source implementations. Even the integrity of benchmark datasets, often taken for granted, was questioned by (6), revealing data leakage that could artificially inflate performance. The choice of negative sampling strategies (as discussed in Subsection 6.2) also represents a critical hyperparameter and design choice that profoundly impacts reproducibility and model quality, further emphasizing

the interconnectedness of these practical considerations.

Conclusion

The emphasis on evaluation, benchmarking, and reproducibility marks a crucial phase in KGE research, moving the field towards higher standards of empirical validation and transparency. The development of unified frameworks like LibKGE (82), coupled with large-scale comparative studies (62; 108) and detailed analyses of hyperparameter effects (6), has exposed significant challenges in reproducibility and evaluation fairness. These efforts are indispensable for ensuring that KGE models are not only powerful but also reliable, comparable, and ultimately trustworthy for deployment in critical applications. The ongoing trajectory demands continued vigilance in experimental design, meticulous reporting, and the development of more robust, unbiased evaluation methodologies to foster reliable scientific progress.

7   Applications and Real-World Impact of KGE

Having thoroughly explored the foundational models, advanced deep learning architectures, methods for enriching embeddings with auxiliary information, and critical practical considerations concerning efficiency, robustness, and evaluation in previous sections, this part of the review shifts its focus from the "how" to the "what for." This section showcases the profound and diverse real-world impact of Knowledge Graph Embedding (KGE) across a myriad of artificial intelligence applications. It moves beyond theoretical advancements and model intricacies to demonstrate how KGE models are actively leveraged to solve complex, real-world problems, transforming symbolic knowledge into actionable insights for intelligent systems (? ? ?).

The utility of KGE is most evident in its ability to enhance core AI tasks that rely on structured knowledge. We will delve into how KGE underpins crucial applications such as link prediction and knowledge graph completion, where embeddings facilitate the inference of missing facts and the enrichment of existing knowledge bases (? ?). Furthermore, the section will explore entity alignment, a vital task for integrating disparate knowledge

sources by identifying equivalent entities across different graphs (? ?). Beyond these foundational tasks, KGE significantly boosts the performance of more interactive systems like question answering, enabling machines to comprehend and respond to natural language queries over vast knowledge repositories, and recommender systems, providing personalized and context-aware suggestions (?). Finally, we highlight domain-specific applications, illustrating the transformative potential of KGE in specialized fields such as biology, medicine, and intellectual property analysis, often emphasizing the growing demand for explainable and interpretable KGE solutions in high-stakes environments (? ?). Collectively, this section underscores the tangible benefits and broad applicability of embedding techniques, positioning KGE as a cornerstone technology in modern AI.

7.1   Link Prediction and Knowledge Graph Completion

The fundamental applications of Knowledge Graph Embedding (KGE) are link prediction (LP) and knowledge graph completion (KGC), which involve inferring missing facts within the graph structure. These tasks are paramount for enhancing the completeness and utility of knowledge graphs, making them more robust and informative for downstream AI systems (48; 104). LP typically focuses on predicting a missing head entity (h), relation (r), or tail entity (t) in an incomplete triple (h, r, t), while KGC broadly encompasses filling in these missing links to expand the knowledge base. The continuous effort in this domain aims to improve accuracy and handle complex relational patterns, forming the bedrock for many other applications.

Foundational Geometric Approaches

Early KGE models primarily leveraged geometric transformations to represent entities and relations, offering a computationally efficient way to model relational patterns for LP/KGC.

- **Translational Models**: Models like TransE (as a baseline for (60; 78)) represent relations as translations in the embedding space, where $h + r \approx t$. While simple and efficient, TransE struggles with complex relation patterns such as one-to-many,

many-to-one, and many-to-many. TransH (60) addressed this by projecting entities onto a relation-specific hyperplane before translation, allowing a single entity to have different representations for different relations. This significantly improved handling of complex mappings with comparable efficiency to TransE. Building on this, TransD (78) further refined the approach by using dynamic mapping matrices constructed from entity and relation vectors, enabling more fine-grained, entity-specific projections and reducing parameters compared to prior projection-based methods. TransA (5; 56) introduced an adaptive margin for the loss function, allowing for more flexible and context-aware embeddings, which was a crucial step towards locally adaptive learning.

- **Rotational and Compound Models**: A significant advancement came with models like RotatE (77), which defined relations as rotations in complex vector spaces. This innovation inherently allowed it to model and infer complex relation patterns like symmetry, antisymmetry, inversion, and composition more effectively than translational models. Rotate3D (64) extended this to three-dimensional space, leveraging the non-commutative property of 3D rotations for multi-hop reasoning. More broadly, CompoundE (73) and CompoundE3D (104) generalized these concepts by combining translation, rotation, and scaling operations, demonstrating that a richer set of cascaded geometric manipulations can lead to superior performance by capturing diverse underlying characteristics of a KG. HousE (107) introduced Householder transformations (rotations and projections) to achieve superior capacity for modeling relation patterns and mapping properties simultaneously, generalizing existing rotation-based models to high-dimensional spaces. MQuinE (106) identified and resolved a "Z-paradox" deficiency in some popular KGE models, ensuring stronger expressiveness for various relation patterns including symmetric/asymmetric, inverse, and compositional relations.

Method Family & Core Innovation & Problem Solved (LP/KGC) & Conditions for Success & Limitations
Translational (TransH [60], TransD [78]) & Project entities onto relation-specific hyperplanes/dynamic matrices. & Better handling of 1-to-N, N-to-1 relations. & Sufficient training data for relation-specific projections. & Still struggles with

The evolution from simple translational models to complex rotational and compound operations highlights a continuous effort to enhance the expressiveness of KGEs for LP/KGC. While TransH (60) and TransD (78) significantly improved upon basic translation by addressing mapping properties, they often fell short in capturing complex logical patterns like composition. This gap was notably addressed by RotatE (77), which demonstrated superior performance in modeling such patterns through complex space rotations. However, even RotatE might not fully exploit the theoretical underpinnings of composition, a limitation that HolmE (99) explicitly aims to overcome by ensuring its relation embedding space is closed under composition, offering stronger theoretical guarantees for handling under-represented patterns. This progression exemplifies the field's shift towards more mathematically robust and expressive models. A common practical limitation across these geometric models is the trade-off between increased expressiveness and computational cost, though many strive for competitive efficiency.

Deep Learning Architectures for LP/KGC

The advent of deep learning brought a paradigm shift, enabling KGE models to learn more intricate and context-aware representations for LP/KGC.

- **Convolutional Neural Networks (CNNs)**: CNNs have been adapted to extract local features and model interactions between entity and relation embeddings. AcrE (74) used atrous convolutions and residual learning to increase feature interactions and address gradient issues, achieving high parameter efficiency. RelnceptionE (70) further explored Inception networks and relation-aware attention to capture joint local-global structural information, enhancing interactions between head and relation embeddings. More recently, CNN-ECFA (59) proposed an entity-specific common feature aggregation strategy, demonstrating superior performance by leveraging CNNs for richer feature extraction. SEConv (53) (designed for healthcare prediction) also employs multi-layer CNNs for deeper structural features, highlighting their utility in specialized domains. These models excel at learning complex, non-linear feature interactions, often outperforming simpler geometric models in accuracy.

- **Graph Neural Networks (GNNs) and Attention Mechanisms**: GNNs, through message passing and aggregation, are naturally suited to capture structural information and neighborhood context. DisenKGAT (7) introduced a disentangled graph attention network to learn diverse component representations, enhancing independence and adaptability to various score functions. GAATs (41) incorporated an attenuated attention mechanism to assign different weights to relation paths and acquire information from neighborhoods, leading to more complete mining of triple features. For inductive KGE, Logic Attention Based Neighborhood Aggregation (84) used a Logic Attention Network (LAN) to aggregate information from an entity's neighbors, allowing for embeddings of unseen entities.

- **Transformer-based KGE Models**: Transformers leverage self-attention mechanisms to capture long-range dependencies and contextualized representations. CoKE (54) treated KGs as sequences of entities and relations, using a Transformer encoder to obtain dynamic, contextualized representations. Knowformer (109) addressed the order-invariance issue of standard Transformers by incorporating relational compositions and position-aware mechanisms, proving effective for both LP and entity alignment. TGformer (66) proposed a general graph Transformer framework that integrates triplet-level and graph-level structural features, boosting the model's ability to understand entities and relations in different contexts.

Method Family & Core Innovation & Problem Solved (LP/KGC) & Conditions for Success & Limitations
CNN-based (AcrE [74], ReInceptionE [70]) & Multi-scale filters, residual learning, inception networks for feature extraction. & Captures intricate non-linear feature interactions for complex relation types (1-to-N, N-to-1). & Requires suf-

Deep learning architectures address the limitations of foundational geometric models by automatically extracting complex features and capturing intricate structural and contextual information. For instance, while RotatE (77) effectively models compositional relations geometrically, CoKE (54) offers a more flexible, data-driven approach by using Transformers to learn contextualized representations, which can capture more nuanced semantic dependencies beyond simple geometric transformations. This exemplifies the field's shift towards leveraging powerful neural networks for enhanced expressiveness, albeit often at the cost of increased computational resources and reduced interpretability. The tension between model expressiveness and computational efficiency is particularly pronounced here, with models like AcrE (74) attempting to balance this through parameter-efficient designs.

Leveraging Auxiliary Information and Rules for LP/KGC

Beyond structural patterns, integrating auxiliary information and logical rules significantly enhances KGE for LP/KGC, especially in incomplete or noisy KGs.

- **Auxiliary Information**: Models like TransET (81) and TaKE (11) incorporate entity type features to learn more semantic representations, guiding the embedding process and improving KG completion. TaKE specifically provides a universal framework that can augment any traditional KGE model. HINGE (68) moved "beyond triplets" by directly learning from hyper-relational facts (triplets with associated key-value pairs), capturing richer data semantics that traditional triplet-based models overlook. GeoEntity-type constrained KGE (59) applied this principle to a specific domain, using geoentity types as constraints to predict natural-language spatial relations more accurately. These methods leverage well-structured prior knowledge to produce more semantic and robust embeddings.

- **Rule-based and Constraint-driven KGE**: This family injects explicit logical rules or constraints to ensure semantic consistency and improve reasoning. Semantically Smooth KGE (97) enforced a "semantically smooth" embedding space where entities of the same category lie close, using manifold learning as regularization.

RUGE ([13]) introduced an iterative guidance mechanism from soft rules, allowing KGE models to learn simultaneously from labeled triples, unlabeled triples, and automatically extracted soft rules. This was a crucial step beyond one-time rule injection. Improving Knowledge Graph Embedding Using Simple Constraints ([90]) showed that even basic constraints like non-negativity and approximate entailment can significantly improve interpretability and structure. More recently, Knowledge Graph Embedding Preserving Soft Logical Regularity ([20]) imposed soft rule constraints directly on relation representations, improving scalability. RulE ([44]) learned rule embeddings jointly with entity and relation embeddings, enabling soft logical inference and using rule embeddings to regularize and enrich the overall embedding space.

Comparative Framework: Auxiliary Information & Rule-based KGE for LP/KGC

Method Family & Core Innovation & Problem Solved (LP/KGC) & Conditions for Success & Limitations
Type/Attribute Augmented (TransET [81], TaKE [11]) & Integrates entity types/attributes into embedding process. & Enhances semantic understanding, especially for incomplete KGs; improves robustness to noise. & Requires availability and quality

These approaches address a critical assumption in many purely structural KGE models: that all necessary information is contained within the triples. TransET (81) and TaKE (11) demonstrate that leveraging entity types provides crucial semantic guidance, leading to more discriminative embeddings. This complements the structural learning of models like TransD (78) by adding an orthogonal dimension of knowledge. Similarly, rule-based methods like RUGE (13) and RulE (44) address the limitation of purely data-driven models by injecting logical consistency, which is vital for robust reasoning in LP/KGC. While TransH (60) improves mapping properties, it doesn't inherently enforce logical consistency; rule-based methods fill this gap. A practical limitation is the reliance on the availability and quality of auxiliary information or rules, which may not always be complete or accurate.

Addressing Dynamic and Complex Structures for LP/KGC

Real-world KGs are dynamic and often exhibit complex, hierarchical structures that Euclidean spaces struggle to represent efficiently.

- **Hyperbolic Embeddings**: Hyperbolic spaces, with their inherent negative curvature, are naturally suited for embedding hierarchical data with high fidelity and fewer dimensions. Hyperbolic Hierarchy-Aware KGE (118) explicitly targeted this by extending the Poincaré Ball model to capture hierarchical structures more effectively. Fully Hyperbolic Rotation (33) pushed for a more native hyperbolic formulation, defining the entire model directly in hyperbolic space using the Lorentz model, emphasizing hyperbolic rotation for relations. This offers superior representation for hierarchical KGs compared to Euclidean counterparts. Mixed Geometry Message and Trainable Convolutional Attention Network (115) proposed a versatile approach by integrating messages and scoring functions from hyperbolic, hypersphere, and Euclidean spaces, allowing for adaptive modeling of diverse local structures, acknowledging that not all KG structures are purely hierarchical.

- **Temporal KGE (TKGE)**: While a dedicated section (5.1) covers TKGE, it's crucial to note its role in LP/KGC for dynamic KGs. HyTE (2) explicitly incorporates

time by associating each timestamp with a hyperplane, enabling temporally aware inference and prediction of temporal scopes for facts. This is essential for predicting links that are valid only for specific time periods.

The tension here lies between the theoretical advantages of non-Euclidean geometries for certain graph structures (e.g., hierarchies) and the increased computational complexity of operations within these spaces. While Fully Hyperbolic Rotation (33) claims competitive results with fewer parameters, the overall training and inference can still be more involved than in Euclidean space.

Efficiency and Robustness in LP/KGC

Practical deployment of KGE for LP/KGC necessitates addressing efficiency and robustness concerns.

- **Negative Sampling and Training Optimization**: Training KGE models relies heavily on negative sampling to discriminate positive from negative triples. NSCaching (100) proposed an efficient method to track and sample "hard" negative triplets using a cache, distilling the benefits of complex GAN-based sampling into a simpler framework. Efficient Non-Sampling KGE (88) took a radical approach by considering all negative instances and leveraging mathematical derivations to reduce computational complexity, aiming for more stable and accurate performance by removing sampling uncertainty. Confidence-Aware Negative Sampling (9) addressed noise in KGs by introducing negative triple confidence. Modality-Aware Negative Sampling (96) extended this to multi-modal KGE, adapting sampling for diverse data modalities. These methods are crucial for optimizing the training process, which directly impacts the accuracy and efficiency of LP/KGC.

- **Compression and Parameter Efficiency**: For large-scale KGs, embedding size and parameter count are critical. DualDE (61) used knowledge distillation to build low-dimensional student KGEs from high-dimensional teachers, reducing parameters by 7-15x and increasing inference speed by 2-6x with minor performance loss. Knowledge Graph Embedding Compression (30) represented entities with discrete codes,

136

achieving 50-1000x compression. LightKG (81) introduced a framework that stores only codebooks and indices, drastically reducing storage and boosting inference. EARL (3) achieved parameter efficiency by learning embeddings for only a small set of "reserved entities" and deriving others from context using entity-agnostic encoders.

The tension between achieving high accuracy in LP/KGC and maintaining computational efficiency is a recurring theme. While models like RotatE (77) achieve state-of-the-art accuracy, their parameter count can be substantial. Methods like DualDE (61) and EARL (3) directly address this by reducing model size and computational cost, making KGE more deployable in resource-constrained environments. However, this often comes with a "minor loss in performance" (30), which might be unacceptable in high-stakes applications. The choice of negative sampling strategy, as reviewed by (47) and (31), is also critical; an inefficient or poor sampling method can severely degrade LP/KGC performance, even for highly expressive models.

In conclusion, LP/KGC remains the central benchmark for KGE models, driving innovations from foundational geometric transformations to advanced deep learning architectures and sophisticated strategies for incorporating auxiliary information and managing dynamic data. The field continuously navigates trade-offs between model expressiveness, computational efficiency, and robustness, with recent efforts increasingly focusing on making these models more adaptable and reliable for real-world applications.

7.2 Entity Alignment

Building upon the foundational capabilities of Knowledge Graph Embedding (KGE) for tasks like link prediction and knowledge graph completion, KGEs have emerged as a powerful paradigm for Entity Alignment (EA). Entity alignment is the critical task of identifying equivalent entities across different knowledge graphs (KGs), a process essential for integrating heterogeneous knowledge sources, enriching existing KGs, and enabling comprehensive cross-KG reasoning (101; 48). The core idea is to embed entities from multiple KGs into a shared, low-dimensional vector space, where semantically equivalent

entities are expected to be close to each other. This transforms the symbolic matching problem into a geometric proximity search, offering a data-driven approach to overcome the challenges of schema heterogeneity and data incompleteness inherent in real-world KGs.

The application of KGE to entity alignment has evolved significantly, addressing key challenges such as the scarcity of labeled alignment data, the integration of diverse entity features, and the leveraging of meta-information. Early approaches often adapted general KGE models (e.g., TransE, TransH (60), TransD (78)) by training them on a small set of pre-aligned entities and then using distance metrics (e.g., Euclidean or cosine similarity) in the embedding space to find new alignments. However, these basic adaptations often struggled with the inherent differences in structure and semantics between KGs, as well as the practical limitation of requiring substantial initial alignment seeds. This motivated the development of more specialized KGE-based EA methods, which can be broadly categorized into iterative/self-training, multi-feature integration, and schema-aware approaches.

Iterative and Semi-Supervised Alignment

A significant challenge in entity alignment is the scarcity of labeled training data, as manual annotation of equivalent entities across large KGs is prohibitively expensive. This limitation is particularly acute when dealing with KGs from different domains or languages.

- **Bootstrapping Entity Alignment**: To mitigate the reliance on extensive prior alignment, Bootstrapping Entity Alignment with Knowledge Graph Embedding (77) proposed an iterative approach.

  1. **Problem Solved**: Addresses the lack of sufficient prior alignment as labeled training data for embedding-based EA.

  2. **Core Innovation**: It iteratively labels likely entity alignments as training data, progressively expanding the set of known alignments. Furthermore, it incorporates an "alignment editing method" to reduce error accumulation, a critical concern in self-training loops.

138

3. **Mechanism**: The process typically starts with a small set of seed alignments. A KGE model is trained using these seeds, and then entity similarities in the learned embedding space are used to identify new, highly confident alignment candidates. These candidates are added to the training set, and the process repeats. The alignment editing method acts as a filter, removing potentially erroneous self-labeled alignments to maintain data quality.

4. **Conditions for Success**: This method succeeds when the initial seed alignments are of high quality and the KGE model can effectively generalize from them. The alignment editing mechanism is crucial for its robustness, preventing the propagation of errors.

5. **Theoretical Limitations**: The primary theoretical limitation is the potential for error accumulation. If the initial KGE model or the confidence threshold for new alignments is not robust, incorrect alignments can be introduced and reinforced, leading to a drift in the embedding space.

6. **Practical Limitations**: The performance is highly sensitive to the quality of the initial seed alignments and the effectiveness of the error reduction strategy. It can also be computationally intensive due to multiple retraining iterations of the KGE model.

- **Semi-Supervised Entity Alignment**: Building on the need to leverage unlabeled data, Semi-Supervised Entity Alignment via Knowledge Graph Embedding with Awareness of Degree Difference (SEA) (86) further refined the approach.

  1. **Problem Solved**: Addresses two key issues: the difficulty of acquiring labeled data (similar to bootstrapping) and the challenge posed by entity degree differences (high-frequency vs. low-frequency entities) which can affect KGE accuracy.

  2. **Core Innovation**: It employs a semi-supervised learning framework to leverage both labeled and abundant unlabeled entity information. Crucially, it improves the KGE with awareness of degree difference by performing adversarial

training, aiming to make embeddings more robust to varying entity frequencies.

3. **Mechanism**: Labeled entity pairs are used to initialize KGEs. For unlabeled entities, a self-training mechanism generates pseudo-labels based on embedding similarity. Adversarial training is introduced to encourage the KGE model to learn representations that are less sensitive to an entity's degree, thereby preventing high-degree entities from dominating the embedding space and improving alignment for low-degree entities.

4. **Conditions for Success**: Requires a robust semi-supervised learning framework and effective adversarial training to balance the influence of entity degrees. It performs well when there's a significant amount of unlabeled data that can be reliably pseudo-labeled.

5. **Theoretical Limitations**: Adversarial training can be notoriously difficult to stabilize and tune, potentially leading to training instabilities or suboptimal embeddings if not carefully managed. The effectiveness of mitigating degree bias might also vary across different KG structures.

6. **Practical Limitations**: Increased complexity due to the adversarial component. The quality of pseudo-labels, especially in early iterations, can still impact overall performance.

Multi-view and Ontology-guided Alignment

Beyond structural information, entities in KGs possess rich auxiliary features that, if leveraged effectively, can significantly enhance alignment accuracy.

- **Multi-view Knowledge Graph Embedding**: While early KGEs for EA primarily focused on relational structures, Multi-view Knowledge Graph Embedding for Entity Alignment (43) recognized the untapped potential of diverse entity features.

  1. **Problem Solved**: Addresses the limitation of previous methods that primarily focused on relational structure, sometimes incorporating attributes, but often overlooking or not equally treating a vast array of other entity features

(e.g., entity names, textual descriptions). This oversight impairs accuracy and robustness.

2. **Core Innovation**: Proposes a novel framework that unifies multiple views of entities (entity names, relations, and attributes) to learn more comprehensive embeddings for entity alignment.

3. **Mechanism**: The framework learns separate embeddings for each view (e.g., using character embeddings for names, structural KGE for relations, and attribute value embeddings). These view-specific embeddings are then combined using various strategies (e.g., concatenation, weighted sum, or attention mechanisms). Cross-KG inference methods are designed to further enhance alignment by propagating alignment signals between the KGs.

4. **Conditions for Success**: This approach thrives when multiple views provide complementary and non-redundant information. The effectiveness hinges on robust methods for learning embeddings from each view and intelligent strategies for combining them.

5. **Theoretical Limitations**: The optimal strategy for combining heterogeneous view embeddings is often heuristic and dataset-dependent, lacking a universal theoretical guarantee. The interplay between different views can be complex, and simple concatenation might not fully capture nuanced interactions.

6. **Practical Limitations**: Increased computational cost due to learning and combining multiple embedding spaces. Requires careful feature engineering and selection for each view. The quality of each view's data (e.g., completeness of attributes, clarity of names) directly impacts performance.

- **Ontology-guided Entity Alignment**: Recognizing that KGs often come with explicit schema or ontological information, OntoEA: Ontology-guided Entity Alignment via Joint Knowledge Graph Embedding (17) proposed to integrate this meta-knowledge.

  1. **Problem Solved**: Addresses the oversight of existing methods that ignore

ontological schemas, which contain critical meta-information such as classes, their hierarchies, and disjointness relationships. This meta-information can significantly constrain and guide the alignment process.

2. **Core Innovation**: Jointly embeds both KGs and their associated ontologies, explicitly utilizing class hierarchy (e.g., `subClassOf` relations) and class disjointness (e.g., `disjointWith` axioms) to avoid false mappings and guide the embedding learning.

3. **Mechanism**: `OntoEA` extends a base KGE model by incorporating additional loss terms that enforce ontological constraints. For instance, if entity e_1 is an instance of class C_1 and entity e_2 is an instance of class C_2, and C_1 is a subclass of C_2, then their embeddings should reflect this hierarchical relationship. Similarly, if C_1 and C_2 are disjoint, then instances of these classes should not be aligned. This joint embedding ensures that entity representations are consistent with both the factual triples and the underlying schema.

4. **Conditions for Success**: Requires the availability of explicit and reasonably complete ontologies for the KGs being aligned. It performs best when these ontologies are consistent and well-defined, providing strong semantic guidance.

5. **Theoretical Limitations**: The effectiveness is directly tied to the quality and completeness of the ontologies. Sparse or inconsistent ontologies can introduce noise or misleading constraints, potentially hindering alignment. The integration of hard logical constraints into continuous embedding spaces can also be challenging.

6. **Practical Limitations**: Many real-world KGs, especially those automatically extracted, may lack rich or explicit ontological schemas. The process of extracting or harmonizing ontologies can be complex and time-consuming.

Comparative Framework: KGE-based Entity Alignment Methods

Method Family & Core Innovation & Problem Solved & Conditions for Success & Limitations
Bootstrapping [77] & Iterative self-labeling with error reduction. & Scarcity of labeled training data. & High-quality initial seeds; effective error editing. & Risk of error accumulation; sensitive to initial seed quality.
Semi-Supervised

Synthesis and Critical Analysis

The evolution of KGE for entity alignment reflects a continuous effort to move beyond purely structural matching towards more semantically rich and robust approaches. The initial challenge of limited labeled data, which implicitly affects any supervised KGE model, was directly tackled by Bootstrapping Entity Alignment (77) and SEA (86). While Bootstrapping iteratively expands the training set, SEA further refines this by explicitly addressing the bias introduced by varying entity degrees through adversarial training. This highlights a crucial pattern in the field: the progression from basic self-training to more sophisticated mechanisms that account for inherent data characteristics. However, a common tension across these iterative methods is the trade-off between leveraging more unlabeled data and the risk of error propagation, which can lead to suboptimal or even incorrect alignments if not carefully managed by mechanisms like alignment editing (77).

Furthermore, the field has recognized that structural information alone is often insufficient for robust EA, especially when KGs are sparse or structurally dissimilar. This led to the development of methods that integrate auxiliary information. Multi-view Knowledge Graph Embedding (43) exemplifies this by unifying entity names, relations, and attributes. This approach directly addresses the limitation of earlier KGEs that might only consider relational triples (as discussed in Section 7.1), by showing that a richer feature set leads to more discriminative embeddings. Similarly, OntoEA (17) extends this by incorporating ontological information, which provides a powerful, explicit form of meta-knowledge. This demonstrates a broader trend towards leveraging all available information, moving from implicit structural patterns to explicit semantic constraints. While Multi-view focuses on diverse *data types*, OntoEA focuses on *schema-level knowledge*, offering complementary improvements. A critical assumption often made by these methods is the existence of high-quality auxiliary data or ontologies, which may not always hold true for all KGs, especially those automatically constructed or from highly specialized domains.

Methodologically, the quality of evaluation in EA is paramount. As highlighted by experimental reviews like (83), a systematic quantitative assessment of strengths and weaknesses across different KG characteristics is crucial. Many EA papers evaluate per-

formance using metrics like Hits@k and MRR on benchmark datasets such as DBP15K, which are valuable but may not fully capture real-world complexities like different entity granularities, varying levels of noise, or the presence of non-isomorphic structures. The generalizability of methods like SEA (86) in mitigating degree bias, for instance, needs to be thoroughly tested across KGs with diverse degree distributions. Similarly, the reliance of OntoEA (17) on well-defined ontologies means its performance might degrade significantly for KGs lacking such rich schema, an unstated assumption that limits its universal applicability.

In conclusion, KGEs provide a powerful, data-driven approach to integrate heterogeneous knowledge sources, which is crucial for building comprehensive knowledge bases and enabling cross-KG reasoning. The field has progressed from basic adaptation of KGE models to specialized frameworks that address data scarcity through iterative learning (77; 86), integrate diverse entity features (43), and leverage ontological information (17). These advancements collectively enhance the ability to find semantic correspondences between disparate knowledge structures. However, challenges persist in ensuring robustness to noise, handling highly sparse KGs, and achieving truly unsupervised alignment. Future research, as indicated by recent surveys (101; 50), will likely focus on integrating even more diverse information (e.g., temporal, multimodal), developing more robust self-training mechanisms, and improving the generalizability of these methods across a wider spectrum of real-world KGs.

7.3   Question Answering and Recommendation Systems

Leveraging the capability of Knowledge Graph Embeddings (KGEs) to represent complex relational data in a continuous, low-dimensional space, the field has significantly advanced applications in Question Answering (QA) and Recommender Systems. As discussed in previous sections, KGEs transform symbolic knowledge into machine-understandable formats, addressing issues of sparsity and enabling semantic reasoning. This subsection explores how KGEs bridge the gap between natural language and structured knowledge, fostering more intelligent, personalized, and interpretable user interactions in these critical

application contexts.

Knowledge Graph Embedding for Question Answering

Question Answering over Knowledge Graphs (QA-KG) aims to answer natural language questions by querying structured facts within a KG. This task is challenging due to the inherent ambiguity and variability of natural language, as well as the complexity of mapping linguistic expressions to precise KG entities and relations. KGEs offer a powerful mechanism to overcome these challenges by embedding both questions and KG elements into a shared semantic space.

- **KEQA: Bridging Natural Language and KG Embeddings**

  1. **Context**: Traditional QA-KG systems often rely on complex semantic parsing or template matching, which struggle with predicate variability and entity ambiguity in natural language questions. The success of KGE in tasks like link prediction ([108]) motivated its application to QA.

  2. **Problem Solved**: Knowledge Graph Embedding Based Question Answering (KEQA) ([37]) addresses the difficulty of identifying the correct head entity and predicate from a natural language question to retrieve a single-fact answer. It aims to make KG access more efficient and intuitive for end-users.

  3. **Core Innovation**: Instead of directly inferring the head entity and predicate, KEQA jointly recovers the question's head entity, predicate, and tail entity representations within the KG embedding spaces. It employs a carefully designed joint distance metric to find the closest fact in the KG as the answer.

  4. **Mechanism**: The natural language question is first processed to extract potential entities and relations. These are then mapped to their corresponding KGEs. The core idea is to find a (head, relation, tail) triple in the KG whose embeddings are maximally similar to the question's inferred (head, relation, tail) embedding, using a scoring function that measures this joint proximity.

5. **Conditions for Success**: KEQA performs well for "simple questions" that can be answered by a single KG triple. Its effectiveness relies on the quality of the underlying KGE model (e.g., TransE, TransH (60)) and robust entity linking to correctly identify entities mentioned in the question.

6. **Theoretical Limitations**: The primary theoretical limitation is its restriction to simple questions. It struggles with multi-hop reasoning, complex logical operations (e.g., aggregations, comparisons), or questions requiring external knowledge beyond a single fact. The joint distance metric, while innovative, may not fully capture the nuances of complex natural language semantics.

7. **Practical Limitations**: Scalability to very large KGs can be an issue if the search space for candidate triples is not efficiently pruned. It also requires a pre-trained KGE model, adding to the overall system complexity.

8. **Comparison**: Compared to rule-based QA systems, KEQA is more robust to linguistic variations due to the semantic capture of embeddings. However, for complex questions, it is less capable than systems that incorporate explicit logical reasoning or advanced NLP parsing.

- **Hybrid Systems: Marie and BERT for Domain-Specific QA**

  1. **Context**: While KEQA established the utility of KGE for simple QA, real-world applications, especially in specialized domains, demand handling complex queries, deep ontologies, and numerical filtering. This necessitates integrating KGE with advanced Natural Language Processing (NLP) models.

  2. **Problem Solved**: Marie and BERT (16) addresses fact-oriented information retrieval in complex, specialized domains like chemistry. It tackles challenges such as deep ontologies, implicit multi-hop relations, numerical filtering, and the need for dynamic calculations via semantic agents.

  3. **Core Innovation**: This system integrates hybrid KGEs (employing multiple embedding methods in parallel), a BERT-based entity-linking model for enhanced robustness, and specialized modules for implicit multi-hop relations

147

and numerical filtering. It also incorporates semantic agents for dynamic calculations.

4. **Mechanism**: The system queries multiple KGE spaces simultaneously, leveraging their diverse strengths. A BERT-based model significantly improves entity linking by capturing contextual semantics. An algorithm is implemented to derive implicit multi-hop relations, extending beyond direct connections. A score alignment model reranks answers from different KGEs, and a joint numerical embedding model handles numerical constraints.

5. **Conditions for Success**: Requires a rich, well-structured domain-specific KG and the availability of pre-trained large language models (LLMs) like BERT. Effective performance hinges on the careful integration and tuning of its heterogeneous components.

6. **Theoretical Limitations**: The complexity of integrating multiple KGEs and a BERT model can make the system difficult to interpret end-to-end, despite the individual components being understandable. Generalizability to vastly different domains might require significant re-training and re-engineering.

7. **Practical Limitations**: High computational cost due to running multiple KGE models and a BERT-based linker. Development and maintenance are complex given the multi-component architecture.

8. **Comparison**: Marie and BERT significantly surpasses KEQA in handling complex, domain-specific QA by combining the semantic power of KGE with the contextual understanding of BERT. It addresses multi-hop and numerical queries that KEQA cannot, but at a much higher computational and engineering overhead. This exemplifies the field's shift towards hybrid, multi-modal approaches (as discussed in Section 4.3) for complex tasks.

Knowledge Graph Embedding for Recommendation Systems

Recommender systems (RecSys) aim to predict user preferences and suggest relevant items. Traditional RecSys often suffer from data sparsity, cold start problems, and a

148

lack of explainability. KGEs enrich item and user representations by incorporating rich relational information from KGs, enabling more accurate, personalized, and interpretable recommendations.

- **RKGE: Recurrent KGE for Path-based Recommendation**

  1. **Context**: Many RecSys rely on collaborative filtering or content-based methods, which struggle with sparsity and often lack transparency. KGs contain rich relational information (e.g., "user buys item," "item is of category," "category has attribute") that can alleviate these issues. Existing KG-based methods often rely on hand-engineered meta-paths, which require domain expertise.

  2. **Problem Solved**: Recurrent knowledge graph embedding for effective recommendation (RKGE) (77) aims to improve recommendation by automatically learning semantic representations of paths between entities in a KG, thereby characterizing user preferences without manual feature engineering. It also provides meaningful explanations.

  3. **Core Innovation**: RKGE employs a novel recurrent neural network (RNN) architecture to model the semantics of paths linking entity pairs (e.g., user-item paths). These path representations are then seamlessly fused into the recommendation process. A pooling operator is used to discriminate the saliency of different paths, offering a form of explanation.

  4. **Mechanism**: User-item interactions and item attributes are represented as paths in a KG. The RNN processes these paths, learning a vector representation for each path. These path embeddings are then aggregated to form a comprehensive representation for user-item pairs, which is used for prediction. The pooling mechanism highlights important paths, providing a rudimentary explanation.

  5. **Conditions for Success**: Requires a sufficiently rich KG where user-item interactions can be meaningfully represented as paths. The RNN needs to be capable of learning effective representations from these paths.

6. **Theoretical Limitations**: The interpretability of RNN-learned path representations can be limited; while it identifies "saliency," the underlying semantic reason might not be immediately clear to a human. It may also struggle with very long or highly diverse paths.

7. **Practical Limitations**: Training RNNs on potentially numerous and long paths can be computationally intensive. Scalability to extremely large KGs with vast numbers of paths can be a challenge.

8. **Comparison**: RKGE offers a more automated approach than earlier meta-path-based methods, which require significant manual feature engineering. It provides a basic level of explainability, moving beyond black-box matrix factorization models.

- **CKGE: Contextualized and Explainable Recommendation**

  1. **Context**: Building on the success of path-based KGE for recommendation, there's a growing demand for more nuanced, motivation-aware, and transparent recommendations, especially in high-stakes domains like talent training. The limitations of RNNs in capturing global dependencies motivate the use of Transformer architectures.

  2. **Problem Solved**: Contextualized Knowledge Graph Embedding for Explainable Talent Training Course Recommendation (CKGE) (4) aims to provide precise and explainable recommendations for talent training courses, considering different learning motivations.

  3. **Core Innovation**: CKGE integrates "contextualized neighbor semantics" and "high-order connections" as "motivation-aware information" by constructing meta-graphs for talent-course pairs. It then employs a novel KG-based Transformer with relational attention and structural encoding, alongside "local path mask prediction" to reveal path importance for explainability.

  4. **Mechanism**: For each talent-course pair, a meta-graph is constructed, incorporating neighbors and meta-paths as contextual information. This meta-

graph is serialized and fed into a KG-based Transformer, which uses relational attention and structural encoding to model global dependencies. A local path mask prediction task explicitly identifies and quantifies the importance of different paths, leading to fine-grained, motivation-aware explanations.

5. **Conditions for Success**: Requires rich contextual information to build effective meta-graphs. The Transformer architecture needs to be robustly trained to capture complex relational patterns and long-range dependencies.

6. **Theoretical Limitations**: The construction of meta-graphs and the interpretation of "motivation-aware information" can be complex and potentially heuristic. While local path mask prediction offers explainability, ensuring that these explanations are truly intuitive and actionable for human users remains a challenge.

7. **Practical Limitations**: High computational cost associated with Transformer models and meta-graph construction. The system's specificity to talent training might limit its direct generalizability to other recommendation domains without significant adaptation.

8. **Comparison**: CKGE represents a significant advancement over RKGE (77) by incorporating contextualization and leveraging the powerful Transformer architecture. It moves beyond simple path saliency to provide more nuanced, motivation-aware recommendations and more precise, explicit path-based explanations. This aligns with the broader trend of integrating advanced deep learning models for enhanced expressiveness and interpretability.

Comparative Framework: KGE in Question Answering and Recommendation Systems

Application/Method & Core Innovation & Problem Solved & Key Strength & Primary Limitation	Method
QA: KEQA (37) & Joint embedding recovery for question elements. & Simple QA with predicate variability/entity ambiguity. & Robust to linguistic variations for simple questions. & Limited to simple, single-hop questions.	

Synthesis and Critical Analysis

The application of KGEs to Question Answering and Recommender Systems clearly illustrates the field's progression from foundational embedding techniques to sophisticated, hybrid, and application-specific solutions. A recurring pattern is the evolution from leveraging KGEs for basic semantic matching to integrating them within complex deep learning architectures to handle more nuanced problems.

For QA, the trajectory moves from KEQA's (37) foundational approach of mapping simple natural language questions to KG triples in an embedding space, to the highly specialized and integrated Marie and BERT system (16). While KEQA demonstrated the initial power of KGEs to overcome predicate variability, its theoretical limitation to single-hop questions highlighted the need for deeper NLP integration. Marie and BERT directly addresses this by combining multiple KGEs with a powerful BERT-based entity linker and dedicated modules for multi-hop and numerical queries. This exemplifies the broader trend of augmenting KGEs with advanced NLP models to capture the full semantic complexity of natural language, as also seen in multi-modal KGEs that integrate textual descriptions (22). The trade-off is evident: increased accuracy and capability for complex queries come at the cost of significantly higher computational complexity and domain specificity. Marie and BERT's reliance on a rich domain-specific KG and pre-trained LLMs is an unstated assumption that limits its direct generalizability to resource-poor or open-domain QA.

Similarly, in recommender systems, KGEs have evolved from enriching item representations to modeling complex user preferences and providing explicit explanations. RKGE (77) pioneered the automatic learning of path semantics using recurrent networks, moving beyond manual feature engineering. This built upon the understanding that relational paths in KGs encode rich interaction patterns, a concept that has been explored in various KGE applications (48). However, the interpretability of RNN-derived path saliency had practical limitations. CKGE (4) significantly advances this by introducing contextualized meta-graphs and a KG-based Transformer, offering more precise, motivation-aware recommendations and explainable paths. This reflects the field's broader shift

towards Transformer architectures (as discussed in Section 3.3) for their superior ability to model long-range dependencies and contextual information. The tension here lies between achieving highly personalized and explainable recommendations and the increased computational burden and data requirements of these advanced models. The "explainability" claimed by both RKGE and CKGE needs careful methodological scrutiny; while they identify influential paths, the *why* behind their influence might still require human interpretation, an evaluation gap that is often overlooked.

In essence, KGEs serve as a crucial bridge, transforming the symbolic, sparse nature of KGs into a continuous, semantically rich representation that is amenable to modern machine learning. This enables more intelligent, personalized, and interpretable user interactions in diverse application contexts. The evolution of these applications underscores a convergent research direction: the integration of KGEs with advanced deep learning models (like BERT and Transformers) and the increasing emphasis on explainability to build trust and provide actionable insights. However, persistent challenges include managing the computational cost of these hybrid systems, ensuring the generalizability of domain-specific solutions, and developing truly intuitive and actionable explanations for human users. Future research will likely focus on more seamless and efficient integration of KGEs with large language models, robust methods for handling noisy and dynamic data in these applications, and standardized metrics for evaluating the quality of explanations.

7.4 Domain-Specific Applications and Explainability

Building upon the discussion of Knowledge Graph Embeddings (KGEs) in general applications like Question Answering and Recommender Systems (Section 7.3), this subsection shifts focus to the specialized and often high-stakes domains where KGEs are not merely applied but meticulously tailored, validated, and, crucially, made interpretable. In fields such as biological systems, patent analysis, and drug discovery, the demand for verifiable and transparent solutions necessitates moving beyond generic performance metrics to deliver actionable insights and build trust. This represents a critical evolution in KGE research, where the emphasis is on practical utility and explainability rather than solely

on abstract embedding quality.

Table 13: Comparative Framework: KGE in Specialized Domains with Emphasis on Explainability and Validation

|p0.15|p0.2|p0.2|p0.2|p0.15|

Domain/Paper	Problem Addressed	KGE Tailoring/Innovation	Domain-Specific Evaluation/Validation	Explainability Focus
Biological Systems (63)	Scalability for complex biological systems, drug-target prediction, polypharmacy side effects.	Learning low-rank vector representations of biological entities/relations.	Predictive accuracy for drug-target interactions, polypharmacy side effects.	Implicit (preserving graph structure), but less explicit.
Patent Metadata Analysis (107)	Measuring knowledge proximity, explaining domain expansion profiles of inventors/assignees.	Training KGE on patent metadata (citations, inventors, assignees, classifications).	Predicting target entities, explaining inventor/assignee domain expansion profiles.	Explaining domain expansion profiles.
Drug Repurposing (COVID-19) (80)	Urgent drug discovery for COVID-19, need for verifiable predictions.	Ensemble KGE, deep neural network for prediction.	Molecular docking (novel for KGE-based repurposing), retrieval of in-trial drugs.	Rules extracted from KG, explanatory paths (post-hoc).
Specific Diseases (Multimodal) (32)	Discovering new reliable knowledge for specific diseases, multimodal reasoning.	Multimodal reasoning (structural, category, description embeddings), reverse-hyperplane projection.	Manual proofreading of predicted pairs (drug-gene, gene-disease, disease-drug), universality of embeddings.	Implicit through multimodal integration, manual proofreading for human-understandable validation.

KGE in Biological Systems and Drug Discovery

The complexity of biological systems, characterized by intricate networks of genes, proteins, drugs, and diseases, presents a formidable challenge for traditional analytical methods. Knowledge graphs naturally represent these interconnected biological entities, but their sheer scale often limits the scalability of graph exploratory approaches (63). KGE models offer a solution by learning low-dimensional vector representations that preserve the graph's inherent structure, enabling efficient predictive and analytical tasks.

- **General Biological Applications**

  1. **Context**: Traditional graph exploration methods for biological KGs, while accurate, struggle with scalability due to time-consuming path exploration. The need for efficient, high-accuracy predictive models in drug discovery and understanding disease mechanisms motivated the adoption of KGEs.

2. **Problem Solved**: Biological applications of knowledge graph embedding models (63) addresses the scalability limitations of traditional graph processing in biological systems and demonstrates KGE's utility for tasks like predicting drug-target interactions and polypharmacy side effects.

3. **Core Innovation**: The core innovation lies in applying existing KGE models (e.g., TransE, DistMult) to biological knowledge graphs, leveraging their ability to learn low-rank vector representations of biological entities (drugs, proteins, diseases) and relations. This transforms complex graph structures into a computationally tractable embedding space.

4. **Mechanism**: Biological entities and their relationships (e.g., "drug A targets protein B," "protein B is associated with disease C") are encoded into a KG. KGE models then embed these entities and relations, allowing for similarity-based predictions (e.g., if drug A is similar to drug D, and drug D targets protein E, then drug A might also target protein E).

5. **Conditions for Success**: Requires well-curated biological KGs. The success hinges on the KGE model's ability to capture nuanced biological relationships accurately, which can be challenging given the heterogeneity and complexity of biological data.

6. **Theoretical Limitations**: While KGEs improve scalability, the interpretability of the learned embeddings in a purely biological sense can be limited. It's often unclear *why* certain embeddings are close or far apart, making it difficult to derive novel biological hypotheses directly from the embedding space without further analysis.

7. **Practical Limitations**: The quality and completeness of biological KGs are paramount. Data sparsity in specific biological pathways or drug-disease interactions can lead to unreliable embeddings.

8. **Comparison**: KGE models offer superior scalability compared to traditional path-exploratory methods (63). However, they often lack the direct, human-readable reasoning paths that rule-based systems (as discussed in Section 4.2)

or symbolic reasoning approaches might provide, which is a significant draw-back in high-stakes fields.

- **Molecular-evaluated and Explainable Drug Repurposing for COVID-19**

  1. **Context**: The urgent need for effective treatments for diseases like COVID-19 highlights the value of drug repurposing. While KGEs can predict drug-disease associations, the high-stakes nature of medical applications demands not just predictions, but also verifiable evidence and clear explanations. This addresses a critical gap in generic KGE applications, which often rely on abstract metrics.

  2. **Problem Solved**: Molecular-evaluated and explainable drug repurposing for COVID-19 using ensemble knowledge graph embedding (80) tackles the challenge of identifying potential drug candidates for COVID-19 while providing molecular-level validation and human-interpretable explanations.

  3. **Core Innovation**: This work proposes an ensemble KGE approach to create robust latent representations, which are then fed into a deep neural network for drug discovery. Crucially, it introduces *molecular docking* as a novel, domain-specific evaluation metric for KGE-based drug repurposing, moving beyond standard link prediction metrics. Furthermore, it provides *explanations* through rules extracted from the KG and instantiated by explanatory paths, directly addressing the need for interpretability in high-stakes medical contexts.

  4. **Mechanism**: A COVID-19 centric KG is constructed. Ensemble embeddings are learned to capture diverse relational patterns, enhancing the robustness of representations. A deep neural network leverages these embeddings to predict drug-COVID-19 associations. Top predictions are then subjected to molecular docking simulations to assess their binding affinity to SARS-CoV-2 targets, providing a physical, verifiable validation. Explanatory paths are generated by tracing relevant relations in the KG that support the prediction, offering a transparent rationale.

5. **Conditions for Success**: Requires a comprehensive and high-quality COVID-19 KG. The effectiveness of molecular docking depends on accurate protein structures and docking algorithms. The quality of extracted rules for explanation is also vital.

6. **Theoretical Limitations**: While ensemble embeddings improve robustness, the underlying deep neural network remains a black box. The post-hoc nature of rule extraction for explanation, while valuable, may not fully capture the complex non-linear reasoning of the neural network. The generalizability of specific molecular docking results to clinical efficacy is also a known challenge in drug discovery.

7. **Practical Limitations**: Molecular docking is computationally intensive. Constructing and maintaining a high-quality, up-to-date domain-specific KG is a significant effort.

8. **Comparison**: This approach significantly advances beyond general biological KGE applications (63) by integrating a *verifiable, domain-specific evaluation metric* (molecular docking) and *explicit explainability mechanisms*. It directly addresses the need for trustworthiness in medical AI, a theme echoed in discussions on evaluation and reproducibility (Section 6.3). The emphasis on rules for explanation connects to the rule-based KGE methods discussed in Section 4.2, demonstrating their practical utility in providing interpretability.

- **Multimodal Reasoning for Specific Diseases**

  1. **Context**: Biomedical KGs are rich but often incomplete. Leveraging multimodal information (e.g., textual descriptions, categories) alongside structural data can enhance knowledge discovery, especially for specific diseases. The challenge is effectively integrating these diverse modalities.

  2. **Problem Solved**: Multimodal reasoning based on knowledge graph embedding for specific diseases (32) aims to discover new and reliable knowledge within Specific Disease Knowledge Graphs (SDKGs) by integrating structural, cate-

gory, and description embeddings.

3. **Core Innovation**: This work constructs SDKGs for various diseases and implements multimodal reasoning using reverse-hyperplane projection. It combines structural embeddings (from graph triples), category embeddings (from entity types), and description embeddings (from textual information) to enrich representations. Reliability is verified through *manual proofreading* of predicted drug-gene, gene-disease, and disease-drug pairs.

4. **Mechanism**: SDKGs are built by extracting triplets, standardizing entities, and linking relations. KGEs are learned for each modality. Reverse-hyperplane projection is used to integrate these multimodal embeddings, allowing for more comprehensive reasoning. The predictions are then manually validated by domain experts, providing a human-centric layer of trustworthiness.

5. **Conditions for Success**: Requires high-quality multimodal data (textual descriptions, categories) alongside structural data. The effectiveness of multimodal fusion is critical. Manual proofreading, while robust, is resource-intensive and relies on expert availability.

6. **Theoretical Limitations**: The reverse-hyperplane projection, while effective, might not fully capture all complex interactions between modalities. The "universality" of embeddings, while demonstrated for biomolecular interaction classification, might not extend to all downstream tasks without further fine-tuning.

7. **Practical Limitations**: Constructing and maintaining multimodal SDKGs is complex. Manual proofreading, while a strong validation, is not scalable for large-scale, continuous knowledge discovery.

8. **Comparison**: This approach highlights the power of multimodal KGEs (as discussed in Section 4.3) in specialized domains. Similar to (80), it emphasizes *verifiable reliability* through manual proofreading, reinforcing the need for human oversight in high-stakes applications. It shows that combining diverse information sources can lead to more robust and reliable knowledge discovery.

KGE in Patent Metadata Analysis

Beyond biomedical applications, KGEs are proving invaluable in intellectual property and innovation analysis, where understanding the relationships between patents, inventors, assignees, and technological domains is crucial for strategic decision-making.

1. **Embedding Knowledge Graph of Patent Metadata**

    (a) **Context**: Understanding "knowledge proximity" between entities (e.g., patents, inventors, companies) in large patent databases is essential for innovation management. Traditional methods often rely on simple co-occurrence or citation analysis, which may not capture nuanced semantic associations.

    (b) **Problem Solved**: Embedding knowledge graph of patent metadata to measure knowledge proximity (107) operationalizes knowledge proximity in the US Patent Database by building a knowledge graph (PatNet) from patent metadata and applying KGEs to measure associations.

    (c) **Core Innovation**: The work constructs PatNet using patent metadata (citations, inventors, assignees, domain classifications) and trains various KGE models on it. Knowledge proximity is then defined as the cosine similarity between the learned embeddings of entities. The approach is validated by its ability to predict target entities and *explain domain expansion profiles* of inventors and assignees.

    (d) **Mechanism**: Patent metadata is transformed into a KG where patents, inventors, assignees, and technology classifications are entities, and relationships like "inventor invents patent," "patent cites patent," "assignee owns patent" are relations. KGE models (e.g., TransE, DistMult) learn embeddings for these entities and relations. The proximity in the embedding space reflects knowledge proximity.

    (e) **Conditions for Success**: Requires a comprehensive patent database with

rich metadata. The quality of the constructed `PatNet` and the chosen KGE model are critical.

(f) **Theoretical Limitations**: While cosine similarity in embedding space is a common proxy for semantic proximity, its direct interpretation as "knowledge proximity" might be an oversimplification. The KGE models primarily capture structural patterns, and the richness of semantic proximity might require more sophisticated models or multimodal integration.

(g) **Practical Limitations**: Building and maintaining a large-scale patent KG is a significant data engineering task. The interpretation of "domain expansion profiles" from embeddings requires careful domain expertise.

(h) **Comparison**: This application demonstrates KGE's utility in a non-biomedical, high-value domain. Similar to the drug repurposing work, it moves beyond generic link prediction to provide *actionable insights* (knowledge proximity, domain expansion) that are directly relevant to industry problems. The focus on explaining domain expansion profiles aligns with the broader demand for explainability, albeit in a different context than medical applications.

Synthesis and Critical Analysis

The application of KGEs in specialized domains marks a significant maturation of the field, shifting from theoretical model development to delivering verifiable and transparent solutions for complex, real-world challenges. A clear pattern emerges: the high-stakes nature of these fields (e.g., medicine, intellectual property) necessitates a departure from sole reliance on generic KGE performance metrics (like Hits@K or MRR), which can be misleading or insufficient for building trust (62; 108). Instead, there is a strong demand for *domain-specific evaluation* and *explainability*.

The evolution from general biological applications (63) to highly specific tasks like COVID-19 drug repurposing (80) illustrates this trend. While (63) highlights KGE's scalability advantage over traditional graph methods, it implicitly assumes that pre-

serving graph structure is sufficient for biological insight. (80) directly questions this assumption by introducing *molecular docking* as a novel, independent, and physically verifiable evaluation metric. This is a crucial innovation, as it grounds abstract embedding predictions in concrete scientific evidence, addressing the methodological quality concern of evaluating KGEs solely on internal metrics. The integration of rule extraction and explanatory paths in (80) further exemplifies the growing demand for interpretable KGE models, moving beyond "what" is predicted to "why" it is predicted, which is paramount in fields where decisions have direct human impact. This directly connects to the rule-based KGE approaches discussed in Section 4.2, showcasing their practical utility in providing transparency.

Similarly, the multimodal reasoning for specific diseases (32) and patent metadata analysis (107) underscore the need for tailored KGEs. (32) leverages multimodal information (structural, category, description embeddings) to enrich representations, aligning with the broader trend of multi-modal KGEs (Section 4.3). Its reliance on *manual proofreading* by experts for validation is a strong, albeit resource-intensive, commitment to trustworthiness. This highlights an unstated assumption in many KGE papers: that standard benchmark evaluations are sufficient. In contrast, these domain-specific applications demonstrate that human expert validation is often indispensable for high-stakes scenarios. (107)'s focus on explaining "domain expansion profiles" in patents further emphasizes that explainability is not confined to medical applications but is a cross-domain requirement for actionable insights.

A critical tension exists between the desire for highly expressive and complex KGE models (often deep learning-based, as discussed in Section 3) and the imperative for interpretability and verifiable solutions in these high-stakes domains. While complex models might achieve higher predictive accuracy, their black-box nature can hinder adoption where trust and accountability are paramount. The approaches reviewed here attempt to bridge this gap through *post-hoc explanation mechanisms* (e.g., rule extraction, explanatory paths (80)) or *human-in-the-loop validation* (e.g., manual proofreading (32)). However, the theoretical limitations of post-hoc explanations,

which may not fully reflect the model's internal reasoning, remain an open challenge.

In conclusion, the application of KGEs in specialized domains represents a significant paradigm shift. It underscores the field's commitment to real-world impact by demanding models that are not only performant but also interpretable, verifiable, and tailored to specific industry problems. This necessitates a re-evaluation of standard KGE evaluation practices, pushing towards domain-specific metrics and robust explanation mechanisms to build trust and deliver transparent, actionable insights for complex, real-world challenges. This evolution aligns with the broader demand for trustworthy AI, a theme that will continue to shape future KGE research.

8   Conclusion and Future Directions

Having journeyed through the foundational theories, advanced architectures, and diverse real-world applications of Knowledge Graph Embeddings (KGEs), this concluding section offers a comprehensive synthesis of the field's intellectual trajectory and charts its future course. The preceding discussions, particularly in Section

Here, we first consolidate these key developments, highlighting the evolution from basic embedding paradigms to highly expressive, dynamic, and multi-modal solutions. However, despite these remarkable strides, the field of KGE is not without its persistent challenges. We will critically examine the open theoretical gaps and practical limitations that continue to demand rigorous investigation, including the delicate balance between model expressiveness and computational complexity, the imperative for enhanced interpretability, and the need for more robust evaluation methodologies (? ? ?). Looking forward, this section will then explore emerging trends, such as the increasingly symbiotic relationship between KGEs and large language models, the development of adaptive embedding spaces, and the growing importance of federated and privacy-preserving KGEs. Crucially, we will also address the ethical considerations inherent in the responsible development and deployment of KGE technologies, including issues of bias and transparency (? ?).

This forward-looking perspective aims not only to summarize the past but also to inspire new research directions and guide the responsible advancement of KGE technologies for a more intelligent and equitable future.

8.1 Summary of Key Developments

The field of knowledge graph embedding (KGE) has undergone a remarkable evolution, transforming the static, symbolic representation of knowledge into dynamic, continuous vector spaces amenable to machine learning. This progression has been driven by a continuous quest for enhanced expressiveness, efficiency, and robustness, culminating in sophisticated models capable of extracting actionable insights for diverse AI applications (50; 57).

The initial wave of KGE research was dominated by *foundational geometric and algebraic models*, primarily translation-based approaches. Pioneering models like TransE (**?**), TransH (60), TransR (**?**), and TransD (78) established the paradigm of representing relations as transformations between entity embeddings. These early models addressed the challenge of efficiently capturing relational semantics by modeling relations as simple translations in Euclidean space. TransE, for instance, represented relations as vectors $h + r \approx t$. However, its simple additive nature struggled with complex mapping properties (e.g., one-to-many, many-to-one, one-to-one, many-to-many) because a single relation vector could not adequately distinguish between multiple tail entities for a given head, or vice-versa (60). To mitigate this, TransH (60) innovated by projecting entities onto relation-specific hyperplanes, allowing for more flexible representations of entities under different relations. Similarly, TransD (78) introduced dynamic mapping matrices to account for the diversity of both relations and entities, reducing parameters compared to its predecessors. While efficient and scalable, a fundamental limitation of these translational models lay in their inability to fully capture intricate logical patterns like symmetry, antisymmetry, inversion, and composition, or to adequately represent polysemous relations, primarily because their linear transformations lacked the

necessary algebraic properties (76). This limitation spurred the development of *rotational and complex space embeddings.* In contrast to the additive nature of translational models, RotatE (77) modeled relations as rotations in complex vector spaces, demonstrating superior capability in inferring these complex patterns due to the inherent multiplicative properties of complex numbers. Building on this, Rotate3D (64) extended rotations to three-dimensional space, leveraging non-commutative properties for multi-hop reasoning. Further algebraic innovations included HousE (107) with Householder transformations and CompoundE (73) and CompoundE3D (104), which combined translation, rotation, and scaling operations, aiming for a more generalized and expressive framework. Other geometric approaches, such as TorusE (**?**) and CyclE (**?**), explored embedding entities on Lie groups or using different metric choices to enhance expressiveness and circumvent regularization problems. These models continuously pushed the boundary of expressiveness by exploring richer mathematical operations and embedding spaces, with recent theoretical advancements ensuring properties like closure under composition (99) and resolving theoretical deficiencies like the "Z-paradox" in existing models (106). The practical limitation of these geometric models often involves a trade-off between increased expressiveness and higher computational complexity or parameter count, making them more challenging to scale than their simpler predecessors.

A significant paradigm shift occurred with the adoption of *deep learning architectures* for KGE, which addressed the problem of capturing highly complex, hierarchical, and non-linear relational patterns that simpler geometric models often missed. Convolutional Neural Networks (CNNs) were introduced to extract local features and model intricate interactions between entity and relation embeddings. Models like AcrE (74) utilized atrous convolutions and residual learning for efficient feature interactions, specifically designed to overcome the "reduced feature resolution" and "information forgotten" issues prevalent in standard deep CNNs, while maintaining parameter efficiency. In contrast, ReInceptionE (70) integrated an Inception network with attention for joint local-global structural information. More

advanced CNN-based approaches like M-DCN (72), CNN-ECFA (59), and SEConv (53) further refined multi-scale feature aggregation and entity-specific common feature learning, demonstrating CNNs' power in capturing non-linear patterns. PConvKB (**?**) further enhanced CNN-based KGE by incorporating relation paths, both locally via attention and globally via a novel DIPF measure, distinguishing itself from models that only utilize triple information. Graph Neural Networks (GNNs) and attention mechanisms also became central, leveraging the graph's topology to learn context-dependent embeddings. Early work like Logic Attention-based Neighborhood Aggregation (84) provided inductive capabilities by aggregating neighbor information. DisenKGAT (7) introduced disentangled graph attention networks for more diverse and independent component representations, while GAATs (41) incorporated graph-attenuated attention to better extract features from neighbor nodes and relation paths. However, a key limitation of many GNN-based KGC models is their "data dependence," where performance is sensitive to local graph structures, and their message functions primarily operate in Euclidean space, failing to capture richer intrinsic structural information (115). To address this, MGTCA (115) proposed a Mixed Geometry Message Function (MGMF) integrating hyperbolic, hypersphere, and Euclidean spaces, alongside a Trainable Convolutional Attention Network (TCAN) that adaptively switches between GNN types (GCN, GAT, KG-CAT) to overcome data dependence. The emergence of Transformer architectures further revolutionized KGE, with models like CoKE (54), Knowformer (109), and TGformer (66) adapting self-attention mechanisms to capture long-range dependencies and contextualized representations. While these deep learning models achieve state-of-the-art accuracy by automatically extracting features and flexibly modeling interactions, their core practical limitations include increased computational cost, higher parameter counts, and reduced interpretability compared to their geometric counterparts.

Beyond purely structural modeling, the field has increasingly focused on *enriching KGE with auxiliary information, rules, and multi-modality*. This addresses the

inherent incompleteness and semantic ambiguity of raw knowledge graphs. Approaches like TransET (81) and TaKE (11) incorporated entity type information to provide semantic guidance, leading to more discriminative embeddings for KG completion. Similarly, AEKE (112) leveraged entity attributes to make KGE models robust against noisy or erroneous triples, a critical practical concern, while HINGE (?) extended beyond triplets to hyper-relational facts. The integration of *logical rules and constraints* also gained prominence, with models like RUGE (13) and RulE (44) iteratively guiding embeddings with soft rules to enforce consistency and enhance reasoning capabilities. These methods solve the problem of injecting prior knowledge and ensuring logical coherence, moving beyond purely data-driven learning. However, a critical assumption here is the availability and quality of such explicit rules, which can be challenging to extract or define for large, dynamic KGs. Furthermore, *multi-modal and cross-domain KGE* emerged to tackle data sparsity and enrich semantic understanding. SSP (95) projected textual descriptions into semantic space, while more recent works like "Joint Language Semantic and Structure Embedding" (22) fine-tuned pre-trained language models to jointly learn from text and structure. Cross-Domain Knowledge Graph Chiasmal Embedding (102) and multimodal reasoning for specific diseases (32) exemplify the application of multi-modal KGE to complex, specialized domains. These approaches collectively aim for a more comprehensive and nuanced knowledge representation by leveraging diverse, complementary information sources. A common practical limitation across all these enrichment strategies is the availability and quality of such auxiliary or multi-modal data, and the inherent complexity of effectively fusing heterogeneous information without introducing new biases or noise.

The field has also made significant strides in addressing *practical considerations* such as efficiency, robustness, and adaptability to dynamic environments. To counter the computational and storage demands of large KGs, techniques like knowledge distillation (DualDE (61)), embedding compression (30; 81), and parameter-efficient learning (EARL (3)) have been developed. System-level optimizations like GE2

(99) further accelerate KGE training. For instance, DualDE reduces model size and inference time by distilling knowledge from a larger teacher model, whereas LightKG focuses on compressing the embedding vectors themselves. Robustness against noisy data and imbalanced distributions has been tackled through methods like confidence-aware negative sampling (9), reinforcement learning-based noise filtering (36), and weighted training schemes (96). The critical role of negative sampling in KGE training has been extensively studied, leading to sophisticated strategies (100; 47) and even non-sampling approaches (88) to improve accuracy and efficiency. The dynamic nature of real-world KGs spurred research into *inductive and continual KGE*. Approaches like Logic Attention-based Neighborhood Aggregation (84) and meta-learning strategies (14; 92) enable models to embed unseen entities and adapt to new facts efficiently. These continual learning models, however, face the persistent challenge of catastrophic forgetting, where acquiring new knowledge can degrade performance on previously learned information; techniques like incremental LoRA (106) aim to mitigate this by parameter-efficient adaptation. Finally, the growing concern for data privacy led to the emergence of *Federated KGE (FKGE)*, with solutions addressing communication efficiency (112), personalization for diverse client data (112), and security vulnerabilities like poisoning attacks (69). While FKGE offers significant privacy benefits, it introduces a trade-off with communication overhead and the complexity of aggregating disparate local models. A comprehensive empirical study by (62) further highlighted that reported KGE performance is highly sensitive to hyperparameter tuning and training configurations, often revealing reproducibility failures. This critique underscores the need for robust training optimization and standardized evaluation practices, as architectural innovation alone does not guarantee reliable real-world performance.

In summary, the narrative arc of KGE research traces a path from simple, elegant geometric models to highly sophisticated deep learning architectures, continuously integrating richer semantic context, logical constraints, and multi-modal information. This evolution reflects a persistent effort to enhance expressiveness, efficiency, and

robustness, while increasingly focusing on adaptability to dynamic, inductive, and distributed knowledge environments. The significant strides made in transforming symbolic knowledge into actionable insights have positioned KGE as a cornerstone for modern AI systems, enabling tasks from link prediction and question answering to personalized recommendation and scientific discovery.

8.2   Open Challenges and Theoretical Gaps

Despite the significant advancements in knowledge graph embedding (KGE) research, as highlighted in the preceding sections, several critical open challenges and theoretical gaps persist, representing fertile ground for future investigation. These challenges often stem from the inherent complexity of knowledge graphs, the trade-offs in model design, and the practical demands of real-world applications.

One of the most fundamental and recurring tensions in KGE research is **balancing model expressiveness with computational complexity and interpretability**. As discussed in Section 2 and 3, the field has progressed from simpler translational models like TransH (60) to highly expressive rotational models such as RotatE (77) and advanced deep learning architectures including GNNs (41; 55) and Transformers. While models like GoldE (120) and Fully Hyperbolic Rotation (33) push the boundaries of expressiveness by capturing intricate logical patterns and hierarchical structures, they often introduce higher computational costs and increased parameter counts. For instance, deep learning models, while powerful in automatically extracting features and modeling non-linear interactions, are notoriously difficult to interpret. This "black-box" nature is a significant practical limitation, especially in high-stakes domains like medicine or finance, where understanding *why* a prediction is made is as crucial as the prediction itself. Although some efforts, such as SpherE (120) and Contextualized KGE for Explainable Recommendation (CKGE) (4), attempt to imbue models with interpretability through specific geometric designs or path-based explanations, these are often post-hoc or limited in scope. The challenge lies in developing models that are both highly expressive and inherently

interpretable, without sacrificing efficiency. The work on knowledge distillation like DualDE (61) offers a partial solution by creating smaller, faster student models from larger teachers, but the interpretability of the underlying teacher model often remains unaddressed.

A second major challenge revolves around **data quality, rule integration, and the 'true' negative distribution in training**. Real-world knowledge graphs are inherently incomplete, noisy, and constantly evolving. As explored in Section 4.2, integrating logical rules (e.g., RUGE (13), RulE (44)) can inject prior knowledge and enhance reasoning. However, the efficient extraction of high-quality, consistent rules from massive and often noisy KGs remains a significant practical hurdle. Furthermore, balancing strict adherence to rules with the flexibility to capture exceptions is a delicate act. Theoretically, the most profound gap in training KGE models lies in resolving issues related to the 'true' negative distribution. Most KGE models rely on negative sampling strategies to create contrastive learning signals, as KGs typically only store positive facts. While methods like NSCaching (100) and Modality-Aware Negative Sampling (MANS) (96) have improved sampling efficiency and quality, and non-sampling approaches like NS-KGE (88) attempt to circumvent the problem entirely, the fundamental issue persists: the true distribution of false facts is unknown (47; 31). Current negative sampling methods are heuristics, implicitly assuming certain distributions of false facts, which may not hold in real-world scenarios. This can lead to suboptimal performance or models that are brittle to noise. For instance, Confidence-Aware Negative Sampling (9) attempts to mitigate noise, but its effectiveness is still bounded by the quality of confidence scores. The challenge is to develop more theoretically grounded and robust training paradigms that are less reliant on such uncertain heuristics, or to devise mechanisms that can reliably infer the 'true' negative distribution.

The **scalability for extremely large and dynamic knowledge graphs** presents another formidable obstacle. As KGs grow to web-scale, traditional KGE training and inference become computationally prohibitive. Section 6.1 highlighted efforts

in efficiency and compression, such as parameter-efficient learning (EARL (3)), embedding compression (30; 81), and optimized systems like GE2 (99). However, these often involve trade-offs. For example, while CPa-WAC (94) uses graph partitioning to improve scalability for GNN-based KGEs, partitioning can inherently lead to a loss of global graph structure, potentially impacting the quality of embeddings that rely on long-range dependencies. The challenge is to devise methods that can scale to billions of entities and relations without sacrificing the richness of the learned representations. Moreover, real-world KGs are rarely static; new entities, relations, and facts emerge constantly. This necessitates the development of **truly generalizable inductive and continual models**. While meta-learning approaches like MorsE (14) and MetaHG (92) and parameter-efficient adaptation techniques like incremental LoRA (FastKGE (106)) have shown promise in handling unseen entities and efficiently updating models, they still grapple with the problem of catastrophic forgetting, where new knowledge acquisition degrades performance on previously learned information. A truly generalizable inductive model should be able to embed entirely novel entities or relations with minimal or no prior context, a capability that remains largely elusive. Furthermore, the complexities of Federated KGE (FKGE), as discussed in Section 5.3, introduce additional scalability and privacy-preserving challenges, such as communication efficiency (112) and personalization across diverse client data (112), while also guarding against adversarial attacks (69).

Finally, the field continues to grapple with the **need for more robust evaluation metrics and improved reproducibility**. As highlighted in Section 6.3, studies like (62) and (108) have exposed significant reproducibility failures and biases in standard evaluation practices, such as the over-representation of certain entities in benchmarks. The high sensitivity of KGE performance to hyperparameter tuning, which varies across datasets (6), further complicates fair comparisons and hinders scientific progress. The current metrics (e.g., MRR, Hits@K) may not fully capture the nuances of complex relational patterns, the quality of explanations, or the utility in specific downstream applications (e.g., set retrieval (120), molecular docking

for drug repurposing (80)). The challenge is twofold: first, to develop standardized, robust, and domain-aware evaluation protocols that accurately reflect model performance across diverse tasks and real-world scenarios; and second, to foster a culture of rigorous reproducibility by providing comprehensive code, configurations, and detailed experimental setups, as advocated by initiatives like LibKGE (82).

In summary, future research must push the boundaries of KGE capabilities by addressing these multifaceted challenges. This includes developing hybrid architectures that intelligently balance expressiveness, efficiency, and interpretability; designing more theoretically sound training paradigms that overcome the limitations of heuristic negative sampling; creating truly scalable and adaptable models for dynamic, web-scale KGs; and establishing more rigorous and comprehensive evaluation frameworks to ensure reliable and trustworthy advancements in the field.

8.3   Emerging Trends and Ethical Considerations

Building upon the persistent open challenges and theoretical gaps identified in the preceding subsection, the field of knowledge graph embedding (KGE) is now witnessing a convergence of cutting-edge technological trends and a growing imperative to address crucial ethical considerations. These emerging directions are poised to redefine the landscape of KGE research, pushing towards more intelligent, robust, and socially responsible AI systems.

One of the most significant emerging trends is the **deeper integration with pretrained language models (PLMs)** for richer semantic understanding. Traditional KGE models, as discussed in Sections 2 and 3, primarily derive embeddings from the structural patterns of knowledge graphs (KGs). While effective, they often struggle with entities or relations that are sparse in the graph but have rich textual descriptions. PLMs, on the other hand, excel at capturing nuanced semantic information from vast text corpora. The synergy between these two paradigms promises to create KGEs that are not only structurally informed but also semanti-

cally grounded. Early efforts in this direction, such as type-augmented frameworks like TaKE (11) or systems like Marie and BERT (16) for chemistry question answering, have demonstrated the benefits of integrating textual features. These approaches, however, often treat text as auxiliary information rather than deeply fusing it with structural learning. The next generation of KGE models will likely move towards joint learning frameworks where PLMs are fine-tuned with KG structures, or where KGE models directly leverage contextualized embeddings from PLMs to enhance entity and relation representations. This integration aims to solve the problem of data sparsity and enhance the zero-shot or few-shot learning capabilities of KGEs, allowing them to generalize to unseen entities with only textual descriptions. However, this trend introduces practical limitations, particularly in computational cost, as large PLMs are resource-intensive. Furthermore, it raises theoretical concerns about how to effectively align the continuous vector spaces of PLMs with the discrete, relational logic of KGs, and how to prevent the propagation of biases inherent in large text corpora into the KGEs.

Another promising direction involves the development of **more adaptive multi-curvature embeddings**. As highlighted in Section 2.3 and 5.1, KGs often exhibit heterogeneous structures: some parts are hierarchical (best modeled by hyperbolic spaces (118; 33)), others are cyclic or grid-like (suited for Euclidean spaces), and some might benefit from spherical geometries. The assumption that a single geometric space can optimally represent all facets of a complex KG is increasingly being questioned. Recent work, such as the Mixed Geometry Message and Trainable Convolutional Attention Network (MGTCA) (115), explicitly addresses this by integrating messages and scoring functions from hyperbolic, hypersphere, and Euclidean spaces. The core innovation here is the ability to adaptively model diverse local structures within a single framework. This approach claims to solve the problem of representational rigidity inherent in single-geometry models, offering higher fidelity for complex KGs. The conditions for its success typically involve KGs with clearly identifiable mixed structural patterns. However, practical limitations include

increased model complexity and potential training instability due to the need to optimize parameters across multiple geometric spaces. Compared to fully hyperbolic models like (33) which commit to a single non-Euclidean space, mixed-geometry models offer greater flexibility but demand more sophisticated optimization strategies. This trend signifies a shift towards more nuanced and geometrically expressive KGEs, moving beyond the "one-size-fits-all" assumption of embedding spaces.

The increasing demand for data privacy and distributed computing is driving advancements in **federated and privacy-preserving KGE**. As discussed in Section 5.3, traditional KGE models often require centralizing sensitive KG data, which is problematic for privacy-conscious applications. Federated KGE (FKGE) addresses this by enabling collaborative model training across distributed KGs without direct data sharing. Innovations in this area focus on overcoming critical challenges such as communication efficiency and data heterogeneity. For instance, Communication-Efficient Federated KGE (112) proposes Entity-Wise Top-K Sparsification to reduce the volume of transmitted parameters, achieving significant efficiency gains with negligible performance degradation. Similarly, Personalized Federated KGE (112) tackles semantic disparities among clients by learning personalized supplementary knowledge, moving beyond a universal global model. These methods aim to solve the dual problem of privacy preservation and collaborative learning. However, the FKGE paradigm introduces new practical and theoretical limitations. Communication overhead, even with sparsification, remains a challenge, and ensuring fair and effective aggregation across heterogeneous client data is complex. Moreover, FKGE systems are vulnerable to adversarial attacks, as demonstrated by the Poisoning Attack on Federated KGE (69), where malicious clients can inject poisoned data to force victim clients to predict false facts. This highlights a critical, often unstated, assumption in FKGE: that participating clients are trustworthy. The development of robust defense mechanisms against such attacks is an urgent, unresolved debate in this emerging field.

Beyond technological advancements, the growing deployment of KGE in real-world

applications necessitates a rigorous examination of **ethical considerations**. A primary concern is the **potential for biases in learned representations**. KGE models, by learning from existing KGs, inevitably inherit and can amplify societal biases present in the training data. These biases, whether related to gender, race, socioeconomic status, or other sensitive attributes, can manifest as unfair or discriminatory outcomes in downstream applications. For example, a KGE-powered recommender system might perpetuate stereotypes, or a KGE used in a hiring context could inadvertently discriminate. As highlighted in Section 6.3, the quality and characteristics of datasets significantly impact KGE performance (6), and biases in these datasets are implicitly encoded into the embeddings. The problem is particularly insidious because these biases are often implicit and difficult to detect or quantify within high-dimensional vector spaces, unlike explicit biases in rule-based systems. This challenges the assumption that KGEs are neutral representations of knowledge. Future research must focus on developing fairness-aware KGE models, bias detection techniques, and debiasing strategies that can mitigate these risks without unduly compromising model utility.

This leads directly to the imperative for the **responsible use of KGE in sensitive applications**. KGEs are increasingly deployed in high-stakes domains such as healthcare (e.g., drug repurposing for COVID-19 (80)), finance, legal systems, and talent management (4). In these contexts, erroneous or biased predictions can have severe societal, economic, or even life-threatening consequences. The methodological quality of KGE models, including their robustness to noise and generalizability, becomes paramount. For instance, the molecular-evaluated and explainable drug repurposing approach (80) explicitly integrates domain-specific validation (molecular docking) and explanation mechanisms, recognizing the critical need for reliability and transparency in medical applications. Similarly, the Contextualized KGE for Explainable Talent Training Course Recommendation (4) emphasizes providing interpretable recommendations. The unstated assumption that KGE models, once validated on standard benchmarks, are fit for sensitive real-world deployment is fun-

damentally flawed. There is a pressing need for domain-specific ethical guidelines, robust auditing mechanisms, and perhaps even regulatory frameworks to ensure that KGE systems are deployed responsibly, with human oversight and accountability.

Finally, the discussion on responsible use underscores the **imperative for transparent and explainable AI systems**. Many advanced KGE models, particularly those leveraging deep learning architectures (Section 3), operate as "black boxes." While they achieve high predictive performance, they offer little insight into *why* a particular prediction or embedding was generated. This lack of transparency hinders trust, makes debugging difficult, and impedes accountability, especially when KGEs are used in critical decision-making processes. As discussed in the previous subsection, achieving inherent interpretability without sacrificing expressiveness remains an open challenge. While post-hoc explanation methods exist, they can sometimes be misleading or incomplete. Models like SpherE (120) aim for inherent interpretability through their geometric design, and path-based explanations in recommendation systems (77; 4) offer valuable insights. However, these are often specific to certain model types or applications. The field must move towards developing KGE models that are inherently more transparent, allowing stakeholders to understand the underlying reasoning. This involves research into intrinsically interpretable architectures, robust and faithful explanation techniques, and user-centric interfaces that present explanations in an understandable and actionable manner.

In conclusion, the next generation of KGE research is characterized by a dual mandate: to push the boundaries of technological innovation through deeper integration with PLMs, adaptive geometries, and distributed learning, while simultaneously upholding societal responsibility by rigorously addressing biases, ensuring responsible deployment in sensitive applications, and prioritizing transparency and explainability. These intertwined trends and ethical considerations will shape the research agenda, ensuring that KGE not only advances AI capabilities but also serves humanity ethically and equitably.

References

[1] Zequn Sun, Wei Hu, Qingheng Zhang, et al. (2018). *Bootstrapping Entity Alignment with Knowledge Graph Embedding.* International Joint Conference on Artificial Intelligence.

[2] S. Dasgupta, Swayambhu Nath Ray, and P. Talukdar (2018). *HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding.* Conference on Empirical Methods in Natural Language Processing.

[3] Mingyang Chen, Wen Zhang, Zhen Yao, et al. (2023). *Entity-Agnostic Representation Learning for Parameter-Efficient Knowledge Graph Embedding.* AAAI Conference on Artificial Intelligence.

[4] Yang Yang, Chubing Zhang, Xin Song, et al. (2023). *Contextualized Knowledge Graph Embedding for Explainable Talent Training Course Recommendation.* ACM Trans. Inf. Syst..

[5] Yantao Jia, Yuanzhuo Wang, Hailun Lin, et al. (2015). *Locally Adaptive Translation for Knowledge Graph Embedding.* AAAI Conference on Artificial Intelligence.

[6] Oliver Lloyd, Yi Liu, and T. Gaunt (2022). *Assessing the effects of hyperparameters on knowledge graph embedding quality.* Journal of Big Data.

[7] Junkang Wu, Wentao Shi, Xuezhi Cao, et al. (2021). *DisenKGAT: Knowledge Graph Embedding with Disentangled Graph Attention Network.* International Conference on Information and Knowledge Management.

[8] Chengjin Xu, M. Nayyeri, Fouad Alkhoury, et al. (2019). *Temporal Knowledge Graph Embedding Model based on Additive Time Series Decomposition.* arXiv.org.

[9] Yingchun Shan, Chenyang Bu, Xiaojian Liu, et al. (2018). *Confidence-Aware Negative Sampling Method for Noisy Knowledge Graph Embedding*. International Conference on Big Knowledge.

[10] Zhuoxun Zheng, Baifan Zhou, Hui Yang, et al. (2024). *Knowledge graph embedding closed under composition*. Data mining and knowledge discovery.

[11] Peng He, Gang Zhou, Yao Yao, et al. (2023). *A type-augmented knowledge graph embedding framework for knowledge graph completion*. Scientific Reports.

[12] Han Xiao, Minlie Huang, and Xiaoyan Zhu (2015). *TransG : A Generative Model for Knowledge Graph Embedding*. Annual Meeting of the Association for Computational Linguistics.

[13] Shu Guo, Quan Wang, Lihong Wang, et al. (2017). *Knowledge Graph Embedding with Iterative Guidance from Soft Rules*. AAAI Conference on Artificial Intelligence.

[14] Mingyang Chen, Wen Zhang, Yushan Zhu, et al. (2021). *Meta-Knowledge Transfer for Inductive Knowledge Graph Embedding*. Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

[15] Guang-pu Li, Zequn Sun, Wei Hu, et al. (2023). *Position-Aware Relational Transformer for Knowledge Graph Embedding*. IEEE Transactions on Neural Networks and Learning Systems.

[16] Xiaochi Zhou, Shaocong Zhang, Mehal Agarwal, et al. (2023). *Marie and BERT—A Knowledge Graph Embedding Based Question Answering System for Chemistry*. ACS Omega.

[17] Yuejia Xiang, Ziheng Zhang, Jiaoyan Chen, et al. (2021). *OntoEA: Ontology-guided Entity Alignment via Joint Knowledge Graph Embedding*. Findings.

[18] Jiahang Cao, Jinyuan Fang, Zaiqiao Meng, et al. (2022). *Knowledge Graph Embedding: A Survey from the Perspective of Representation Spaces*. ACM Computing Surveys.

[19] Peng Wang, Jing Zhou, Yuzhang Liu, et al. (2021). *TransET: Knowledge Graph Embedding with Entity Types*. Electronics.

[20] Shu Guo, Lin Li, Zhen Hui, et al. (2020). *Knowledge Graph Embedding Preserving Soft Logical Regularity*. International Conference on Information and Knowledge Management.

[21] Xiaoxiong Zhang, Zhiwei Zeng, Xin Zhou, et al. (2024). *Communication-Efficient Federated Knowledge Graph Embedding with Entity-Wise Top-K Sparsification*. Knowledge-Based Systems.

[22] Jianhao Shen, Chenguang Wang, Linyuan Gong, et al. (2022). *Joint Language Semantic and Structure Embedding for Knowledge Graph Completion*. International Conference on Computational Linguistics.

[23] Kairong Hu, Xiaozhi Zhu, Hai Liu, et al. (2024). *Convolutional Neural Network-Based Entity-Specific Common Feature Aggregation for Knowledge Graph Embedding Learning*. IEEE transactions on consumer electronics.

[24] Yang Liu, Huang Fang, Yunfeng Cai, et al. (2024). *MQuinE: a Cure for "Z-paradox" in Knowledge Graph Embedding*. Conference on Empirical Methods in Natural Language Processing.

[25] Yongqi Zhang, Quanming Yao, Wenyuan Dai, et al. (2019). *AutoSF: Searching Scoring Functions for Knowledge Graph Embedding*. IEEE International Conference on Data Engineering.

[26] Shihui Yang, Jidong Tian, Honglun Zhang, et al. (2019). *TransMS: Knowledge Graph Embedding for Complex Relations by Multidirectional Semantics*. International Joint Conference on Artificial Intelligence.

[27] Zhiwen Xie, Runjie Zhu, Jin Liu, et al. (2023). *TARGAT: A Time-Aware Relational Graph Attention Model for Temporal Knowledge Graph Embedding.* IEEE/ACM Transactions on Audio Speech and Language Processing.

[28] Jiapu Wang, Boyue Wang, Junbin Gao, et al. (2024). *MADE: Multicurvature Adaptive Embedding for Temporal Knowledge Graph Completion.* IEEE Transactions on Cybernetics.

[29] Han Xiao, Yidong Chen, and X. Shi (2019). *Knowledge Graph Embedding Based on Multi-View Clustering Framework.* IEEE Transactions on Knowledge and Data Engineering.

[30] Mrinmaya Sachan (2020). *Knowledge Graph Embedding Compression.* Annual Meeting of the Association for Computational Linguistics.

[31] Tiroshan Madushanka, and R. Ichise (2024). *Negative Sampling in Knowledge Graph Representation Learning: A Review.* arXiv.org.

[32] Chaoyu Zhu, Zhihao Yang, Xiaoqiong Xia, et al. (2022). *Multimodal reasoning based on knowledge graph embedding for specific diseases.* Bioinform..

[33] Qiuyu Liang, Weihua Wang, F. Bao, et al. (2024). *Fully Hyperbolic Rotation for Knowledge Graph Embedding.* European Conference on Artificial Intelligence.

[34] Li, Yuyi Ao, and Jingrui He (2024). *SpherE: Expressive and Interpretable Knowledge Graph Embedding for Set Retrieval.* Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

[35] Takuma Ebisu, and R. Ichise (2017). *TorusE: Knowledge Graph Embedding on a Lie Group.* AAAI Conference on Artificial Intelligence.

[36] Zhao Zhang, Fuzhen Zhuang, Hengshu Zhu, et al. (2021). *Towards Robust Knowledge Graph Embedding via Multi-Task Reinforcement Learning.* IEEE Transactions on Knowledge and Data Engineering.

[37] Xiao Huang, Jingyuan Zhang, Dingcheng Li, et al. (2019). *Knowledge Graph Embedding Based Question Answering*. Web Search and Data Mining.

[38] Yun Tang, Jing Huang, Guangtao Wang, et al. (2019). *Orthogonal Relation Transforms with Graph Context Modeling for Knowledge Graph Embedding*. Annual Meeting of the Association for Computational Linguistics.

[39] Zhu Sun, Jie Yang, Jie Zhang, et al. (2018). *Recurrent knowledge graph embedding for effective recommendation*. ACM Conference on Recommender Systems.

[40] Xiou Ge, Yun Cheng Wang, Bin Wang, et al. (2023). *Knowledge Graph Embedding: An Overview*. APSIPA Transactions on Signal and Information Processing.

[41] Rui Wang, Bicheng Li, Shengwei Hu, et al. (2020). *Knowledge Graph Embedding via Graph Attenuated Attention Networks*. IEEE Access.

[42] Rui Li, Jianan Zhao, Chaozhuo Li, et al. (2022). *HousE: Knowledge Graph Embedding with Householder Parameterization*. International Conference on Machine Learning.

[43] Qingheng Zhang, Zequn Sun, Wei Hu, et al. (2019). *Multi-view Knowledge Graph Embedding for Entity Alignment*. International Joint Conference on Artificial Intelligence.

[44] Xiaojuan Tang, Song-Chun Zhu, Yitao Liang, et al. (2022). *RulE: Knowledge Graph Reasoning with Rule Embedding*. Annual Meeting of the Association for Computational Linguistics.

[45] Xin Lv, Lei Hou, Juan-Zi Li, et al. (2018). *Differentiating Concepts and Instances for Knowledge Graph Embedding*. Conference on Empirical Methods in Natural Language Processing.

[46] Jie Chen, Yinlong Wang, Shu Zhao, et al. (2025). *Contextualized Quaternion Embedding Towards Polysemy in Knowledge Graph for Link Prediction*. ACM Trans. Asian Low Resour. Lang. Inf. Process..

[47] Jing Qian, Gangmin Li, Katie Atkinson, et al. (2021). *Understanding Negative Sampling in Knowledge Graph Embedding*. International Journal of Artificial Intelligence Applications.

[48] Yuanfei Dai, Shiping Wang, N. Xiong, et al. (2020). *A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks*. Electronics.

[49] Hao Ji, Li Yan, and Z. Ma (2024). *FSTRE: Fuzzy Spatiotemporal RDF Knowledge Graph Embedding Using Uncertain Dynamic Vector Projection and Rotation*. IEEE transactions on fuzzy systems.

[50] Qi Yan, Jiaxin Fan, Mohan Li, et al. (2022). *A Survey on Knowledge Graph Embedding*. International Conference on Data Science in Cyberspace.

[51] Yichi Zhang, Mingyang Chen, and Wen Zhang (2023). *Modality-Aware Negative Sampling for Multi-modal Knowledge Graph Embedding*. IEEE International Joint Conference on Neural Network.

[52] Ren Li, Yanan Cao, Qiannan Zhu, et al. (2021). *How Does Knowledge Graph Embedding Extrapolate to Unseen Data: a Semantic Evidence View*. AAAI Conference on Artificial Intelligence.

[53] Qingqing Yang, Min He, Zhongwen Li, et al. (2025). *A Semantic Enhanced Knowledge Graph Embedding Model With AIGC Designed for Healthcare Prediction*. IEEE transactions on consumer electronics.

[54] Quan Wang, Pingping Huang, Haifeng Wang, et al. (2019). *CoKE: Contextualized Knowledge Graph Embedding*. arXiv.org.

[55] Shimin Di, and Lei Chen (2023). *Message Function Search for Knowledge Graph Embedding*. The Web Conference.

[56] Yantao Jia, Yuanzhuo Wang, Xiaolong Jin, et al. (2017). *Knowledge Graph Embedding.* ACM Transactions on the Web.

[57] Shivani Choudhary, Tarun Luthra, Ashima Mittal, et al. (2021). *A Survey of Knowledge Graph Embedding and Their Applications.* arXiv.org.

[58] Han Xiao, Minlie Huang, and Xiaoyan Zhu (2015). *From One Point to a Manifold: Knowledge Graph Embedding for Precise Link Prediction.* International Joint Conference on Artificial Intelligence.

[59] Lei Hu, Wenwen Li, Jun Xu, et al. (2024). *GeoEntity-type constrained knowledge graph embedding for predicting natural-language spatial relations.* International Journal of Geographical Information Science.

[60] Zhen Wang, Jianwen Zhang, Jianlin Feng, et al. (2014). *Knowledge Graph Embedding by Translating on Hyperplanes.* AAAI Conference on Artificial Intelligence.

[61] Yushan Zhu, Wen Zhang, Mingyang Chen, et al. (2020). *DualDE: Dually Distilling Knowledge Graph Embedding for Faster and Cheaper Reasoning.* Web Search and Data Mining.

[62] Mehdi Ali, M. Berrendorf, Charles Tapley Hoyt, et al. (2020). *Bringing Light Into the Dark: A Large-Scale Evaluation of Knowledge Graph Embedding Models Under a Unified Framework.* IEEE Transactions on Pattern Analysis and Machine Intelligence.

[63] Sameh K. Mohamed, A. Nounu, and V. Nováček (2020). *Biological applications of knowledge graph embedding models.* Briefings Bioinform..

[64] Chang Gao, Chengjie Sun, Lili Shan, et al. (2020). *Rotate3D: Representing Relations as Rotations in Three-Dimensional Space for Knowledge Graph Embedding.* International Conference on Information and Knowledge Management.

[65] Xutan Peng, Guanyi Chen, Chenghua Lin, et al. (2021). *Highly Efficient Knowledge Graph Embedding Learning with Orthogonal Procrustes Analysis.* North American Chapter of the Association for Computational Linguistics.

[66] Fobo Shi, Duantengchuan Li, Xiaoguang Wang, et al. (2025). *TGformer: A Graph Transformer Framework for Knowledge Graph Embedding.* IEEE Transactions on Knowledge and Data Engineering.

[67] Xiaoxiong Zhang, Zhiwei Zeng, Xin Zhou, et al. (2024). *Personalized Federated Knowledge Graph Embedding with Client-Wise Relation Graph.* Applied intelligence (Boston).

[68] Paolo Rosso, Dingqi Yang, and P. Cudré-Mauroux (2020). *Beyond Triplets: Hyper-Relational Knowledge Graph Embedding for Link Prediction.* The Web Conference.

[69] Enyuan Zhou, Song Guo, Zhixiu Ma, et al. (2024). *Poisoning Attack on Federated Knowledge Graph Embedding.* The Web Conference.

[70] Zhiwen Xie, Guangyou Zhou, Jin Liu, et al. (2020). *ReInceptionE: Relation-Aware Inception Network with Joint Local-Global Structural Information for Knowledge Graph Embedding.* Annual Meeting of the Association for Computational Linguistics.

[71] Tengwei Song, Jie Luo, and Lei Huang (2021). *Rot-Pro: Modeling Transitivity by Projection in Knowledge Graph Embedding.* Neural Information Processing Systems.

[72] Zhaoli Zhang, Zhifei Li, Hai Liu, et al. (2020). *Multi-Scale Dynamic Convolutional Network for Knowledge Graph Embedding.* IEEE Transactions on Knowledge and Data Engineering.

[73] Xiou Ge, Yun Cheng Wang, Bin Wang, et al. (2022). *CompoundE: Knowledge Graph Embedding with Translation, Rotation and Scaling Compound Operations.* arXiv.org.

[74] Feiliang Ren, Jucheng Li, Huihui Zhang, et al. (2020). *Knowledge Graph Embedding with Atrous Convolution and Residual Learning.* International Conference on Computational Linguistics.

[75] Jun Yuan, Neng Gao, and Ji Xiang (2019). *TransGate: Knowledge Graph Embedding with Shared Gate Structure.* AAAI Conference on Artificial Intelligence.

[76] Han Xiao, Minlie Huang, Yu Hao, et al. (2015). *TransA: An Adaptive Approach for Knowledge Graph Embedding.* arXiv.org.

[77] Zhiqing Sun, Zhihong Deng, Jian-Yun Nie, et al. (2018). *RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space.* International Conference on Learning Representations.

[78] Guoliang Ji, Shizhu He, Liheng Xu, et al. (2015). *Knowledge Graph Embedding via Dynamic Mapping Matrix.* Annual Meeting of the Association for Computational Linguistics.

[79] Lifan Lin, and Kun She (2020). *Tensor Decomposition-Based Temporal Knowledge Graph Embedding.* IEEE International Conference on Tools with Artificial Intelligence.

[80] M. Islam, Diego Amaya-Ramirez, B. Maigret, et al. (2023). *Molecular-evaluated and explainable drug repurposing for COVID-19 using ensemble knowledge graph embedding.* Scientific Reports.

[81] Haoyu Wang, Yaqing Wang, Defu Lian, et al. (2021). *A Lightweight Knowledge Graph Embedding Framework for Efficient Inference and Storage.* International Conference on Information and Knowledge Management.

[82] Samuel Broscheit, Daniel Ruffinelli, Adrian Kochsiek, et al. (2020). *LibKGE - A knowledge graph embedding library for reproducible research.* Conference on Empirical Methods in Natural Language Processing.

[83] N. Fanourakis, Vasilis Efthymiou, D. Kotzinos, et al. (2022). *Knowledge graph embedding methods for entity alignment: experimental review.* Data mining and knowledge discovery.

[84] Peifeng Wang, Jialong Han, Chenliang Li, et al. (2018). *Logic Attention Based Neighborhood Aggregation for Inductive Knowledge Graph Embedding.* AAAI Conference on Artificial Intelligence.

[85] Pedro Tabacof, and Luca Costabello (2019). *Probability Calibration for Knowledge Graph Embedding Models.* International Conference on Learning Representations.

[86] Shichao Pei, Lu Yu, R. Hoehndorf, et al. (2019). *Semi-Supervised Entity Alignment via Knowledge Graph Embedding with Awareness of Degree Difference.* The Web Conference.

[87] Yongqi Zhang, Quanming Yao, Yingxia Shao, et al. (2018). *NSCaching: Simple and Efficient Negative Sampling for Knowledge Graph Embedding.* IEEE International Conference on Data Engineering.

[88] Zelong Li, Jianchao Ji, Zuohui Fu, et al. (2021). *Efficient Non-Sampling Knowledge Graph Embedding.* The Web Conference.

[89] Guangtong Li, L. Siddharth, and Jianxi Luo (2022). *Embedding knowledge graph of patent metadata to measure knowledge proximity.* J. Assoc. Inf. Sci. Technol..

[90] Boyang Ding, Quan Wang, Bin Wang, et al. (2018). *Improving Knowledge Graph Embedding Using Simple Constraints.* Annual Meeting of the Association for Computational Linguistics.

[91] Xuanyu Zhang, Qing Yang, and Dongliang Xu (2022). *TranS: Transition-based Knowledge Graph Embedding with Synthetic Relation Representation.* Conference on Empirical Methods in Natural Language Processing.

[92] Hongliang Sun, Jinlan Liu, Can Wang, et al. (2024). *Learning Dynamic Knowledge Graph Embedding in Evolving Service Ecosystems via Meta-Learning.* 2024 IEEE International Conference on Web Services (ICWS).

[93] Jiapu Wang, Zheng Cui, Boyue Wang, et al. (2024). *IME: Integrating Multi-curvature Shared and Specific Embedding for Temporal Knowledge Graph Completion.* The Web Conference.

[94] S. Modak, Aakarsh Malhotra, Sarthak Malik, et al. (2024). *CPa-WAC: Constellation Partitioning-based Scalable Weighted Aggregation Composition for Knowledge Graph Embedding.* International Joint Conference on Artificial Intelligence.

[95] Han Xiao, Minlie Huang, Lian Meng, et al. (2016). *SSP: Semantic Space Projection for Knowledge Graph Embedding with Text Descriptions.* AAAI Conference on Artificial Intelligence.

[96] Zhao Zhang, Zhanpeng Guan, Fuwei Zhang, et al. (2023). *Weighted Knowledge Graph Embedding.* Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

[97] Shu Guo, Quan Wang, Bin Wang, et al. (2015). *Semantically Smooth Knowledge Graph Embedding.* Annual Meeting of the Association for Computational Linguistics.

[98] Chengjin Xu, M. Nayyeri, Fouad Alkhoury, et al. (2020). *TeRo: A Time-aware Knowledge Graph Embedding via Temporal Rotation.* International Conference on Computational Linguistics.

[99] Chenguang Zheng, Guanxian Jiang, Xiao Yan, et al. (2024). *GE2: A General and Efficient Knowledge Graph Embedding Learning System.* Proc. ACM Manag. Data.

[100] Zhao Zhang, Fuzhen Zhuang, Meng Qu, et al. (2018). *Knowledge Graph Embedding with Hierarchical Relation Structure*. Conference on Empirical Methods in Natural Language Processing.

[101] Beibei Zhu, Ruolin Wang, Junyi Wang, et al. (2024). *A survey: knowledge graph entity alignment research based on graph embedding*. Artificial Intelligence Review.

[102] Jia Liu, Wei Huang, Tianrui Li, et al. (2023). *Cross-Domain Knowledge Graph Chiasmal Embedding for Multi-Domain Item-Item Recommendation*. IEEE Transactions on Knowledge and Data Engineering.

[103] S. Choi, Hyun-Je Song, and Seong-Bae Park (2020). *An Approach to Knowledge Base Completion by a Committee-Based Knowledge Graph Embedding*. Applied Sciences.

[104] Xiou Ge, Yun Cheng Wang, Bin Wang, et al. (2023). *Knowledge Graph Embedding with 3D Compound Geometric Transformations*. APSIPA Transactions on Signal and Information Processing.

[105] A. Sadeghian, Mohammadreza Armandpour, Anthony Colas, et al. (2021). *ChronoR: Rotation Based Temporal Knowledge Graph Embedding*. AAAI Conference on Artificial Intelligence.

[106] Jiajun Liu, Wenjun Ke, Peng Wang, et al. (2024). *Fast and Continual Knowledge Graph Embedding via Incremental LoRA*. International Joint Conference on Artificial Intelligence.

[107] Yizhi Li, Wei Fan, Chaochun Liu, et al. (2022). *TranSHER: Translating Knowledge Graph Embedding with Hyper-Ellipsoidal Restriction*. Conference on Empirical Methods in Natural Language Processing.

[108] Andrea Rossi, D. Firmani, Antonio Matinata, et al. (2020). *Knowledge Graph Embedding for Link Prediction*. ACM Transactions on Knowledge Discovery from Data.

[109] Jiang Li, Xiangdong Su, and Guanglai Gao (2023). *TeAST: Temporal Knowledge Graph Embedding via Archimedean Spiral Timeline*. Annual Meeting of the Association for Computational Linguistics.

[110] Yanhui Peng, and Jing Zhang (2020). *LineaRE: Simple but Powerful Knowledge Graph Embedding for Link Prediction*. Industrial Conference on Data Mining.

[111] Hao Ji, Li Yan, and Z. Ma (2024). *Multihop Fuzzy Spatiotemporal RDF Knowledge Graph Query via Quaternion Embedding*. IEEE transactions on fuzzy systems.

[112] Qinggang Zhang, Junnan Dong, Qiaoyu Tan, et al. (2024). *Integrating Entity Attributes for Error-Aware Knowledge Graph Embedding*. IEEE Transactions on Knowledge and Data Engineering.

[113] Adrian Kochsiek (2021). *Parallel Training of Knowledge Graph Embedding Models: A Comparison of Techniques*. Proceedings of the VLDB Endowment.

[114] Han Yang, Leilei Zhang, Bingning Wang, et al. (2021). *Cycle or Minkowski: Which is More Appropriate for Knowledge Graph Embedding?*. International Conference on Information and Knowledge Management.

[115] Bin Shang, Yinliang Zhao, Jun Liu, et al. (2024). *Mixed Geometry Message and Trainable Convolutional Attention Network for Knowledge Graph Completion*. AAAI Conference on Artificial Intelligence.

[116] S. M. Asmara, N. A. Sahabudin, Nor Syahidatul Nadiah Ismail, et al. (2023). *A Review of Knowledge Graph Embedding Methods of TransE, TransH and TransR for Missing Links*. International Conference on Software Engineering and Computer Systems.

[117] Cosimo Gregucci, M. Nayyeri, D. Hern'andez, et al. (2023). *Link Prediction with Attention Applied on Multiple Knowledge Graph Embedding Models*. The Web Conference.

[118] Zhe Pan, and Peng Wang (2021). *Hyperbolic Hierarchy-Aware Knowledge Graph Embedding for Link Prediction*. Conference on Empirical Methods in Natural Language Processing.

[119] Hee-Geun Yoon, Hyun-Je Song, Seong-Bae Park, et al. (2016). *A Translation-Based Knowledge Graph Embedding Preserving Logical Property of Relations*. North American Chapter of the Association for Computational Linguistics.

[120] Rui Li, Chaozhuo Li, Yanming Shen, et al. (2024). *Generalizing Knowledge Graph Embedding with Universal Orthogonal Parameterization*. International Conference on Machine Learning.

[121] Chenyan Xiong, Russell Power, and Jamie Callan (2017). *Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding*. The Web Conference.

[122] Fan Gong, Meng Wang, Haofen Wang, et al. (2020). *SMR: Medical Knowledge Graph Embedding for Safe Medicine Recommendation*. Big Data Research.

[123] Bin Zhou, Xingwang Shen, Yuqian Lu, et al. (2022). *Semantic-aware event link reasoning over industrial knowledge graph embedding time series data*. International Journal of Production Research.

[124] Thanh-Binh Le, N. Le, and H. Le (2022). *Knowledge graph embedding by relational rotation and complex convolution for link prediction*. Expert systems with applications.

[125] Zhehui Zhou, Can Wang, Yan Feng, et al. (2022). *JointE: Jointly utilizing 1D and 2D convolution for knowledge graph embedding*. Knowledge-Based Systems.

[126] Da Xu, Chuanwei Ruan, Evren Körpeoglu, et al. (2019). *Product Knowledge Graph Embedding for E-commerce*. Web Search and Data Mining.

[127] Haithem Mezni, D. Benslimane, and Ladjel Bellatreche (2021). *Context-Aware Service Recommendation Based on Knowledge Graph Embedding*. IEEE Transactions on Knowledge and Data Engineering.

[128] P. Do, and Truong H. V. Phan (2021). *Developing a BERT based triple classification model using knowledge graph embedding for question answering system*. Applied intelligence (Boston).

[129] Gengchen Mai, K. Janowicz, Ling Cai, et al. (2020). *SE-KGE: A location-aware Knowledge Graph Embedding model for Geographic Question Answering and Spatial Semantic Lifting*. Trans. GIS.

[130] Jiarui Zhang, Jian Huang, Jialong Gao, et al. (2022). *Knowledge graph embedding by logical-default attention graph convolution neural network for link prediction*. Information Sciences.

[131] Daniel N. Sosa, Alexander Derry, Margaret Guo, et al. (2019). *A Literature-Based Knowledge Graph Embedding Method for Identifying Drug Repurposing Opportunities in Rare Diseases*. bioRxiv.

[132] Niannian Guan, Dandan Song, and L. Liao (2019). *Knowledge graph embedding with concepts*. Knowledge-Based Systems.

[133] M. Fan, Qiang Zhou, E. Chang, et al. (2014). *Transition-based Knowledge Graph Embedding with Relational Mapping Properties*. Pacific Asia Conference on Language, Information and Computation.

[134] Hengtong Zhang, T. Zheng, Jing Gao, et al. (2019). *Data Poisoning Attack against Knowledge Graph Embedding*. International Joint Conference on Artificial Intelligence.

[135] Qi Chen, Wei Wang, Kaizhu Huang, et al. (2022). *Zero-Shot Text Classification via Knowledge Graph Embedding for Social Media Data*. IEEE Internet of Things Journal.

[136] Xin Wang, Shengfei Lyu, Xiangyu Wang, et al. (2022). *Temporal knowledge graph embedding via sparse transfer matrix.* Information Sciences.

[137] Mingyang Chen, Wen Zhang, Zonggang Yuan, et al. (2022). *Federated knowledge graph completion via embedding-contrastive learning.* Knowledge-Based Systems.

[138] Bilal Abu-Salih, Marwan Al-Tawil, Ibrahim Aljarah, et al. (2020). *Relational Learning Analysis of Social Politics using Knowledge Graph Embedding.* Data mining and knowledge discovery.

[139] Haichuan Fang, Youwei Wang, Zhen Tian, et al. (2022). *Learning knowledge graph embedding with a dual-attention embedding network.* Expert systems with applications.

[140] R. Çelebi, Hüseyin Uyar, Erkan Yasar, et al. (2019). *Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings.* BMC Bioinformatics.

[141] Xiao Sha, Zhu Sun, and Jie Zhang (2019). *Hierarchical attentive knowledge graph embedding for personalized recommendation.* Electronic Commerce Research and Applications.

[142] Zhifei Li, Hai Liu, Zhaoli Zhang, et al. (2021). *Recalibration convolutional networks for learning interaction knowledge graph embedding.* Neurocomputing.

[143] Han Xiao, Minlie Huang, Yu Hao, et al. (2015). *TransG : A Generative Mixture Model for Knowledge Graph Embedding.* arXiv.org.

[144] Fei Zhang, Bo Sun, Xiaolin Diao, et al. (2021). *Prediction of adverse drug reactions based on knowledge graph embedding.* BMC Medical Informatics and Decision Making.

[145] Guanying Wang, Wen Zhang, Ruoxu Wang, et al. (2018). *Label-Free Distant*

Supervision for Relation Extraction via Knowledge Graph Embedding. Conference on Empirical Methods in Natural Language Processing.

[146] Xinyu Li, P. Zheng, Jinsong Bao, et al. (2021). *Achieving cognitive mass personalization via the self-X cognitive manufacturing network: An industrial-knowledge-graph- and graph-embedding-enabled pathway.* Engineering.

[147] Xin Wang, Xiao Liu, Jin Liu, et al. (2021). *A novel knowledge graph embedding based API recommendation method for Mashup development.* World wide web (Bussum).

[148] Víctor Gutiérrez-Basulto, and S. Schockaert (2018). *From Knowledge Graph Embedding to Ontology Embedding? An Analysis of the Compatibility between Vector Space Representations and Rules.* International Conference on Principles of Knowledge Representation and Reasoning.

[149] Jan Portisch, Nicolas Heist, and Heiko Paulheim (2022). *Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction - two sides of the same coin?.* Semantic Web.

[150] Fuwei Zhang, Zhao Zhang, Xiang Ao, et al. (2022). *Along the Time: Timeline-traced Embedding for Temporal Knowledge Graph Completion.* International Conference on Information and Knowledge Management.

[151] Jun Feng, Minlie Huang, Mingdong Wang, et al. (2016). *Knowledge Graph Embedding by Flexible Translation.* International Conference on Principles of Knowledge Representation and Reasoning.

[152] Jia Liu, Tianrui Li, Shenggong Ji, et al. (2021). *Urban Flow Pattern Mining Based on Multi-Source Heterogeneous Data Fusion and Knowledge Graph Embedding.* IEEE Transactions on Knowledge and Data Engineering.

[153] Shengtian Sang, Zhihao Yang, Xiaoxia Liu, et al. (2019). *GrEDeL: A Knowledge Graph Embedding Based Method for Drug Discovery From Biomedical Literatures.* IEEE Access.

[154] M. Wang, Mengyue Liu, Jun Liu, et al. (2017). *Safe Medicine Recommendation via Medical Knowledge Graph Embedding.* arXiv.org.

[155] Dan Jiang, Ronggui Wang, Juan Yang, et al. (2021). *Kernel multi-attention neural network for knowledge graph embedding.* Knowledge-Based Systems.

[156] Yang Liu, Zequn Sun, Guang-pu Li, et al. (2022). *I Know What You Do Not Know: Knowledge Graph Embedding via Co-distillation Learning.* International Conference on Information and Knowledge Management.

[157] Nasrullah Khan, Zongmin Ma, Aman Ullah, et al. (2022). *Similarity attributed knowledge graph embedding enhancement for item recommendation.* Information Sciences.

[158] Haithem Mezni (2021). *Temporal Knowledge Graph Embedding for Effective Service Recommendation.* IEEE Transactions on Services Computing.

[159] Qianjin Zhang, Ronggui Wang, Juan Yang, et al. (2021). *Structural context-based knowledge graph embedding for link prediction.* Neurocomputing.

[160] Xuqian Huang, Jiuyang Tang, Zhen Tan, et al. (2021). *Knowledge graph embedding by relational and entity rotation.* Knowledge-Based Systems.

[161] Aleksandar Pavlovic, and Emanuel Sallinger (2022). *ExpressivE: A Spatio-Functional Embedding For Knowledge Graph Completion.* International Conference on Learning Representations.

[162] Shensi Wang, Kun Fu, Xian Sun, et al. (2021). *Hierarchical-aware relation rotational knowledge graph embedding for link prediction.* Neurocomputing.

[163] Shuai Zhang, Yi Tay, Lina Yao, et al. (2019). *Quaternion Knowledge Graph Embedding.* arXiv.org.

[164] Gengchen Mai, Bo Yan, K. Janowicz, et al. (2019). *Relaxing Unanswerable Geographic Questions Using A Spatially Explicit Knowledge Graph Embedding Model.* Agile Conference.

[165] Zhuobing Han, Xiaohong Li, Hongtao Liu, et al. (2018). *DeepWeak: Reasoning common software weaknesses via knowledge graph embedding*. IEEE International Conference on Software Analysis, Evolution, and Reengineering.

[166] Feiyang Wang, Zhongbao Zhang, Li Sun, et al. (2022). *DiriE: Knowledge Graph Embedding with Dirichlet Distribution*. The Web Conference.

[167] Ilaria Ferrari, Giacomo Frisoni, Paolo Italiani, et al. (2022). *Comprehensive Analysis of Knowledge Graph Embedding Techniques Benchmarked on Link Prediction*. Electronics.

[168] Guirong Fu, Zhao Meng, Zhen Han, et al. (2022). *TempCaps: A Capsule Network-based Embedding Model for Temporal Knowledge Graph Completion*. SPNLP.

[169] Yanrong Wu, and Zhichun Wang (2018). *Knowledge Graph Embedding with Numeric Attributes of Entities*. Rep4NLP@ACL.

[170] Qianjin Zhang, Ronggui Wang, Juan Yang, et al. (2021). *Knowledge graph embedding by reflection transformation*. Knowledge-Based Systems.

[171] Sameh K. Mohamed, V. Nováček, P. Vandenbussche, et al. (2019). *Loss Functions in Knowledge Graph Embedding Models*. DL4KG@ESWC.

[172] Kexuan Xin, Zequn Sun, Wen Hua, et al. (2022). *Large-scale Entity Alignment via Knowledge Graph Merging, Partitioning and Embedding*. International Conference on Information and Knowledge Management.

[173] Binling Nie, and Shouqian Sun (2019). *Knowledge graph embedding via reasoning over entities, relations, and text*. Future generations computer systems.

[174] Yang Liu, Qingguo Zeng, Huanrui Yang, et al. (2018). *Stock Price Movement Prediction from Financial News with Deep Learning and Knowledge Graph Embedding*. Pacific Rim Knowledge Acquisition Workshop.

[175] Chien-Chun Ni, Kin Sum Liu, and Nicolas Torzec (2020). *Layered Graph Embedding for Entity Recommendation using Wikipedia in the Yahoo! Knowledge Graph.* The Web Conference.

[176] Chen Li, Xutan Peng, Yuhang Niu, et al. (2021). *Learning graph attention-aware knowledge graph embedding.* Neurocomputing.

[177] S. Yu, Sujit Rokka Chhetri, A. Canedo, et al. (2019). *Pykg2vec: A Python Library for Knowledge Graph Embedding.* Journal of machine learning research.

[178] Bahare Fatemi, Siamak Ravanbakhsh, and D. Poole (2018). *Improved Knowledge Graph Embedding using Background Taxonomic Information.* AAAI Conference on Artificial Intelligence.

[179] Zhuo Chen, Mi-Yen Yeh, and Tei-Wei Kuo (2021). *PASSLEAF: A Pool-bAsed Semi-Supervised LEArning Framework for Uncertain Knowledge Graph Embedding.* AAAI Conference on Artificial Intelligence.

[180] Sicong Dong, Xupeng Miao, Peng Liu, et al. (2022). *HET-KG: Communication-Efficient Knowledge Graph Embedding Training via Hotness-Aware Cache.* IEEE International Conference on Data Engineering.

[181] Fengyuan Lu, Peijin Cong, and Xinli Huang (2020). *Utilizing Textual Information in Knowledge Graph Embedding: A Survey of Methods and Applications.* IEEE Access.

[182] Weidong Li, Rong Peng, and Zhi Li (2022). *Improving knowledge graph completion via increasing embedding interactions.* Applied intelligence (Boston).

[183] Yuanfei Luo, Quan Wang, Bin Wang, et al. (2015). *Context-Dependent Knowledge Graph Embedding.* Conference on Empirical Methods in Natural Language Processing.

[184] Xiaohan Zhou, Yunhui Yi, and Geng Jia (2021). *Path-RotatE: Knowledge*

Graph Embedding by Relational Rotation of Path in Complex Space. International Conference on Innovative Computing and Cloud Computing.

[185] Feng Zhao, Haoran Sun, Langjunqing Jin, et al. (2020). *Structure-augmented knowledge graph embedding for sparse data with rule learning.* Computer Communications.

[186] Yantao Jia, Yuanzhuo Wang, Xiaolong Jin, et al. (2018). *Path-specific knowledge graph embedding.* Knowledge-Based Systems.

[187] Gengchen Mai, K. Janowicz, and Bo Yan (2018). *Combining Text Embedding and Knowledge Graph Embedding Techniques for Academic Search Engines.* Semdeep/NLIWoD@ISWC.

[188] Dingcheng Li, Siamak Zamani, Jingyuan Zhang, et al. (2019). *Integration of Knowledge Graph Embedding Into Topic Modeling with Hierarchical Dirichlet Process.* North American Chapter of the Association for Computational Linguistics.

[189] Xiaoli Tang, Rui Yuan, Qianyu Li, et al. (2020). *Timespan-Aware Dynamic Knowledge Graph Embedding by Incorporating Temporal Evolution.* IEEE Access.

[190] Lingbing Guo, Qiang Zhang, Zequn Sun, et al. (2022). *Understanding and Improving Knowledge Graph Embedding for Entity Alignment.* International Conference on Machine Learning.

[191] Dan Jiang, Ronggui Wang, Lixia Xue, et al. (2022). *Multiview feature augmented neural network for knowledge graph embedding.* Knowledge-Based Systems.

[192] Yu Liu, Wen Hua, Kexuan Xin, et al. (2019). *Context-Aware Temporal Knowledge Graph Embedding.* WISE.

[193] Qianjin Zhang, Ronggui Wang, Juan Yang, et al. (2020). *Knowledge graph embedding by translating in time domain space for link prediction.* Knowledge-Based Systems.

[194] Liang Chang, Manli Zhu, T. Gu, et al. (2017). *Knowledge Graph Embedding by Dynamic Translation.* IEEE Access.

[195] Yeon-Chang Lee, and Sang-Wook Kim (2022). *THOR: Self-Supervised Temporal Knowledge Graph Embedding via Three-Tower Graph Convolutional Networks.* Industrial Conference on Data Mining.

[196] Yongqi Zhang, Zhanke Zhou, Quanming Yao, et al. (2022). *Efficient Hyperparameter Search for Knowledge Graph Embedding.* Annual Meeting of the Association for Computational Linguistics.

[197] Chang Liu, Lun Li, Xiaolu Yao, et al. (2019). *A Survey of Recommendation Algorithms Based on Knowledge Graph Embedding.* 2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI).

[198] Wei Song, Jingjin Guo, Ruiji Fu, et al. (2021). *A Knowledge Graph Embedding Approach for Metaphor Processing.* IEEE/ACM Transactions on Audio Speech and Language Processing.

[199] László Grad-Gyenge, A. Kiss, and P. Filzmoser (2017). *Graph Embedding Based Recommendation Techniques on the Knowledge Graph.* User Modeling, Adaptation, and Personalization.

[200] Xiaofei Zhou, Lingfeng Niu, Qiannan Zhu, et al. (2021). *Knowledge Graph Embedding by Double Limit Scoring Loss.* IEEE Transactions on Knowledge and Data Engineering.

[201] Yao Chen, Jiangang Liu, Zhe Zhang, et al. (2021). *MöbiusE: Knowledge Graph Embedding on Möbius Ring.* Knowledge-Based Systems.

[202] Yongqi Zhang, Quanming Yao, and Lei Chen (2020). *Interstellar: Searching Recurrent Architecture for Knowledge Graph Embedding.* Neural Information Processing Systems.

[203] Armand Boschin (2020). *TorchKGE: Knowledge Graph Embedding in Python and PyTorch.* arXiv.org.

[204] P. Wang, D. Dou, Fangzhao Wu, et al. (2019). *Logic Rules Powered Knowledge Graph Embedding.* arXiv.org.

[205] E. B. Myklebust, Ernesto Jiménez-Ruiz, Jiaoyan Chen, et al. (2019). *Knowledge Graph Embedding for Ecotoxicological Effect Prediction.* International Workshop on the Semantic Web.

[206] Miriyala Kartheek, and G. Sajeev (2021). *Building Semantic Based Recommender System Using Knowledge Graph Embedding.* International Conference on Intelligent Information Processing.

[207] Xiao Sha, Zhu Sun, and Jie Zhang (2019). *Attentive Knowledge Graph Embedding for Personalized Recommendation.* arXiv.org.

[208] Haonan Lu, and Hailin Hu (2020). *DensE: An Enhanced Non-Abelian Group Representation for Knowledge Graph Embedding.* arXiv.org.

[209] Siheng Zhang, Zhengya Sun, and Wensheng Zhang (2020). *Improve the translational distance models for knowledge graph embedding.* Journal of Intelligence and Information Systems.

[210] Mingda Li, Zhengya Sun, Siheng Zhang, et al. (2020). *Enhancing Knowledge Graph Embedding with Relational Constraints.* 2020 IEEE International Conference on Knowledge Graph (ICKG).

[211] Jian Li, Zhuoming Xu, Yan Tang, et al. (2020). *Deep Hybrid Knowledge Graph Embedding for Top-N Recommendation.* Web Information System and Application Conference.

[212] Kuekyeng Kim, Yuna Hur, Gyeongmin Kim, et al. (2020). *GREG: A Global Level Relation Extraction with Knowledge Graph Embedding.* Applied Sciences.

[213] Jizhao Zhu, Yantao Jia, Jun Xu, et al. (2018). *Modeling the Correlations of Relations for Knowledge Graph Embedding.* Journal of Computational Science and Technology.

[214] Kien Do, T. Tran, and S. Venkatesh (2018). *Knowledge Graph Embedding with Multiple Relation Projections.* International Conference on Pattern Recognition.

[215] Yunpu Ma, Volker Tresp, Liming Zhao, et al. (2019). *Variational Quantum Circuit Model for Knowledge Graph Embedding.* Advanced Quantum Technologies.

[216] Yuhang Zhang, Jun Wang, and Jie Luo (2020). *Knowledge Graph Embedding Based Collaborative Filtering.* IEEE Access.

[217] Wen Zhang, Shumin Deng, Han Wang, et al. (2019). *XTransE: Explainable Knowledge Graph Embedding for Link Prediction with Lifestyles in e-Commerce.* Joint International Conference of Semantic Technology.

[218] Zhihao Wang, and Xin Li (2019). *Hybrid-TE: Hybrid Translation-Based Temporal Knowledge Graph Embedding.* IEEE International Conference on Tools with Artificial Intelligence.

[219] Hung Nghiep Tran, and A. Takasu (2019). *Analyzing Knowledge Graph Embedding Methods from a Multi-Embedding Interaction Perspective.* EDBT/ICDT Workshops.

[220] Shengwu Xiong, Weitao Huang, and P. Duan (2018). *Knowledge Graph Embedding via Relation Paths and Dynamic Mapping Matrix.* ER Workshops.

[221] Wessel Radstok, M. Chekol, and M. Schäfer (2021). *Are Knowledge Graph*

Embedding Models Biased, or Is it the Data That They Are Trained on?. Wikidata@ISWC.

[222] Ling Zhao, Hanhan Deng, L. Qiu, et al. (2020). *Urban Multi-Source Spatio-Temporal Data Analysis Aware Knowledge Graph Embedding*. Symmetry.

[223] Maoyuan Zhang, Qi Wang, Wukui Xu, et al. (2018). *Discriminative Path-Based Knowledge Graph Embedding for Precise Link Prediction*. European Conference on Information Retrieval.

[224] Ningning Jia, Xiang Cheng, and Sen Su (2020). *Improving Knowledge Graph Embedding Using Locally and Globally Attentive Relation Paths*. European Conference on Information Retrieval.

[225] Qiannan Zhu, Xiaofei Zhou, P. Zhang, et al. (2019). *A neural translating general hyperplane for knowledge graph embedding*. Journal of Computer Science.

[226] Shen Wang, Xiaokai Wei, C. D. Santos, et al. (2021). *Knowledge Graph Representation via Hierarchical Hyperbolic Neural Graph Embedding*. 2021 IEEE International Conference on Big Data (Big Data).

[227] Zhiyuan Ning, Ziyue Qiao, Hao Dong, et al. (2021). *LightCAKE: A Lightweight Framework for Context-Aware Knowledge Graph Embedding*. Pacific-Asia Conference on Knowledge Discovery and Data Mining.

[228] Nasrullah Sheikh, Xiao Qin, B. Reinwald, et al. (2021). *Knowledge Graph Embedding using Graph Convolutional Networks with Relation-Aware Attention*. arXiv.org.

[229] Wiem Ben Rim, Carolin (Haas) Lawrence, Kiril Gashteovski, et al. (2021). *Behavioral Testing of Knowledge Graph Embedding Models for Link Prediction*. Conference on Automated Knowledge Base Construction.

[230] Chunhong Zhang, Miao Zhou, Xiao Han, et al. (2017). *Knowledge Graph Embedding for Hyper-Relational Data*. Unpublished manuscript.

[231] R. Çelebi, Erkan Yasar, Hüseyin Uyar, et al. (2018). *Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction using Linked Open Data*. Workshop on Semantic Web Applications and Tools for Life Sciences.

[232] Martina Garofalo, Maria Angela Pellegrino, Abdulrahman Altabba, et al. (2018). *Leveraging Knowledge Graph Embedding Techniques for Industry 4.0 Use Cases*. arXiv.org.

[233] Kai Wang, Yu Liu, Xiujuan Xu, et al. (2018). *Knowledge Graph Embedding with Entity Neighbors and Deep Memory Network*. arXiv.org.

[234] Chanyoung Chung, and Joyce Jiyoung Whang (2021). *Knowledge Graph Embedding via Metagraph Learning*. Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

[235] Hung Nghiep Tran, and A. Takasu (2019). *Exploring Scholarly Data by Semantic Query on Knowledge Graph Embedding Space*. International Conference on Theory and Practice of Digital Libraries.

[236] Jun Shi, Huan Gao, G. Qi, et al. (2017). *Knowledge Graph Embedding with Triple Context*. International Conference on Information and Knowledge Management.

[237] Wen Zhang (2017). *Knowledge Graph Embedding with Diversity of Structures*. The Web Conference.

[238] Ming-Yi Zhu, De-sheng Zhen, Ran Tao, et al. (2019). *Top-N Collaborative Filtering Recommendation Algorithm Based on Knowledge Graph Embedding*. International Conference on Knowledge Management in Organizations.

[239] Natthawut Kertkeidkachorn, Xin Liu, and R. Ichise (2019). *GTransE: Generalizing Translation-Based Model on Uncertain Knowledge Graph Embedding*. JSAI.

[240] Jia Zhu, Zetao Zheng, Min Yang, et al. (2019). *A semi-supervised model for knowledge graph embedding.* Data mining and knowledge discovery.

[241] Hengtong Zhang, T. Zheng, Jing Gao, et al. (2019). *Towards Data Poisoning Attack against Knowledge Graph Embedding.* arXiv.org.

[242] Wenqiang Liu, Hongyun Cai, Xu Cheng, et al. (2019). *Learning High-order Structural and Attribute information by Knowledge Graph Attention Networks for Enhancing Knowledge Graph Embedding.* Knowledge-Based Systems.

[243] Vibhor Kanojia, Hideyuki Maeda, Riku Togashi, et al. (2017). *Enhancing Knowledge Graph Embedding with Probabilistic Negative Sampling.* The Web Conference.

[244] Huan Gao, Jun Shi, G. Qi, et al. (2018). *Triple Context-Based Knowledge Graph Embedding.* IEEE Access.

[245] Gengchen Mai, K. Janowicz, and Bo Yan (2018). *Support and Centrality: Learning Weights for Knowledge Graph Embedding Models.* International Conference Knowledge Engineering and Knowledge Management.

[246] Han Xiao, Minlie Huang, and Xiaoyan Zhu (2016). *Knowledge Semantic Representation: A Generative Model for Interpretable Knowledge Graph Embedding.* arXiv.org.

[247] Peifeng Liu, Lu Qian, Xingwei Zhao, et al. (2024). *Joint Knowledge Graph and Large Language Model for Fault Diagnosis and Its Application in Aviation Assembly.* IEEE Transactions on Industrial Informatics.

[248] Jin-cheng Zhang, A. Zain, Kai Zhou, et al. (2024). *A review of recommender systems based on knowledge graph embedding.* Expert systems with applications.

[249] Xiao-Rui Su, Zhuhong You, Deshuang Huang, et al. (2023). *Biomedical Knowledge Graph Embedding With Capsule Network for Multi-Label Drug-Drug Interaction Prediction.* IEEE Transactions on Knowledge and Data Engineering.

[250] Xiangrong Zhu, Guang-pu Li, and Wei Hu (2023). *Heterogeneous Federated Knowledge Graph Embedding Learning and Unlearning.* The Web Conference.

[251] Jiajun Liu, Wenjun Ke, Peng Wang, et al. (2024). *Towards Continual Knowledge Graph Embedding via Incremental Distillation.* AAAI Conference on Artificial Intelligence.

[252] Wei Wang, Xiaoxuan Shen, Baolin Yi, et al. (2024). *Knowledge-aware fine-grained attention networks with refined knowledge graph embedding for personalized recommendation.* Expert systems with applications.

[253] Duantengchuan Li, Tao Xia, Jing Wang, et al. (2024). *SDFormer: A shallow-to-deep feature interaction for knowledge graph embedding.* Knowledge-Based Systems.

[254] Jaejun Lee, Chanyoung Chung, and Joyce Jiyoung Whang (2023). *InGram: Inductive Knowledge Graph Embedding via Relation Graphs.* International Conference on Machine Learning.

[255] Zeinab Shokrzadeh, M. Feizi-Derakhshi, M. Balafar, et al. (2023). *Knowledge graph-based recommendation system enhanced by neural collaborative filtering and knowledge graph embedding.* Ain Shams Engineering Journal.

[256] Weibo Gao, Hao Wang, Qi Liu, et al. (2023). *Leveraging Transferable Knowledge Concept Graph Embedding for Cold-Start Cognitive Diagnosis.* Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

[257] Yufeng Li, Wenchao Zhao, Bo Dang, et al. (2024). *Research on Adverse Drug Reaction Prediction Model Combining Knowledge Graph Embedding and Deep*

Learning. 2024 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE).

[258] Zengcan Xue, Zhao Zhang, Hai Liu, et al. (2023). *Learning knowledge graph embedding with multi-granularity relational augmentation network.* Expert systems with applications.

[259] Pengbo Duan, Kuo Yang, Xin Su, et al. (2024). *HTINet2: herb–target prediction via knowledge graph embedding and residual-like graph neural network.* Briefings Bioinform..

[260] Zhen Chen, Dalin Zhang, Shanshan Feng, et al. (2024). *KGTS: Contrastive Trajectory Similarity Learning over Prompt Knowledge Graph Embedding.* AAAI Conference on Artificial Intelligence.

[261] Jia Zhu, Changqin Huang, and P. D. Meo (2022). *DFMKE: A dual fusion multi-modal knowledge graph embedding framework for entity alignment.* Information Fusion.

[262] Katerina Mitropoulou, Panagiotis C. Kokkinos, P. Soumplis, et al. (2023). *Anomaly Detection in Cloud Computing using Knowledge Graph Embedding and Machine Learning Mechanisms.* Journal of Grid Computing.

[263] Harry Shomer, Wei Jin, Wentao Wang, et al. (2023). *Toward Degree Bias in Embedding-Based Knowledge Graph Completion.* The Web Conference.

[264] Mingjie Wang, Zijie Li, Jun Wang, et al. (2024). *TracKGE: Transformer with Relation-pattern Adaptive Contrastive Learning for Knowledge Graph Embedding.* Knowledge-Based Systems.

[265] Zhifei Li, Wei Huang, Xuchao Gong, et al. (2024). *Decoupled semantic graph neural network for knowledge graph embedding.* Neurocomputing.

[266] Mingqi Li, Wenming Ma, and Zihao Chu (2024). *KGIE: Knowledge graph*

convolutional network for recommender system with interactive embedding. Knowledge-Based Systems.

[267] Yan Jia, Mengqi Lin, Yechen Wang, et al. (2023). *Extrapolation over temporal knowledge graph via hyperbolic embedding.* CAAI Transactions on Intelligence Technology.

[268] Wei Huang, Jia Liu, Tianrui Li, et al. (2023). *FedCKE: Cross-Domain Knowledge Graph Embedding in Federated Learning.* IEEE Transactions on Big Data.

[269] Ruoxin Wang, and C. F. Cheung (2023). *Knowledge graph embedding learning system for defect diagnosis in additive manufacturing.* Computers in industry (Print).

[270] Xiangning Hou, Ruizhe Ma, Li Yan, et al. (2023). *T-GAE: A Timespan-aware Graph Attention-based Embedding Model for Temporal Knowledge Graph Completion.* Information Sciences.

[271] Dan Jiang, Ronggui Wang, Lixia Xue, et al. (2023). *Multisource hierarchical neural network for knowledge graph embedding.* Expert systems with applications.

[272] H. Lu, Hailin Hu, and Xiaodong Lin (2022). *DensE: An enhanced non-commutative representation for knowledge graph embedding with adaptive semantic hierarchy.* Neurocomputing.

[273] W. Djeddi, Khalil Hermi, S. Yahia, et al. (2023). *Advancing drug–target interaction prediction: a comprehensive graph-based approach integrating knowledge graph embedding and ProtBert pretraining.* BMC Bioinformatics.

[274] Yuchao Zhang, Xiangjie Kong, Zhehui Shen, et al. (2024). *A survey on temporal knowledge graph embedding: Models and applications.* Knowledge-Based Systems.

[275] Thanh-Binh Le, Huy Tran, and H. Le (2023). *Knowledge graph embedding with the special orthogonal group in quaternion space for link prediction.* Knowledge-Based Systems.

[276] Zhen Yao, Wen Zhang, Mingyang Chen, et al. (2023). *Analogical Inference Enhanced Knowledge Graph Embedding.* AAAI Conference on Artificial Intelligence.

[277] Zhipeng Li, Shanshan Feng, Jun Shi, et al. (2023). *Future Event Prediction Based on Temporal Knowledge Graph Embedding.* Computer systems science and engineering.

[278] Shihan Yang, Weiya Zhang, R. Tang, et al. (2022). *Approximate inferring with confidence predicting based on uncertain knowledge graph embedding.* Information Sciences.

[279] Debayan Banerjee, Pranav Ajit Nair, Ricardo Usbeck, et al. (2023). *GETT-QA: Graph Embedding based T2T Transformer for Knowledge Graph Question Answering.* Extended Semantic Web Conference.

[280] Yuke Hu, Wei Liang, Ruofan Wu, et al. (2023). *Quantifying and Defending against Privacy Threats on Federated Knowledge Graph Embedding.* The Web Conference.

[281] Daiyi Li, Li Yan, Xiaowen Zhang, et al. (2023). *EventKGE: Event knowledge graph embedding with event causal transfer.* Knowledge-Based Systems.

[282] Xinkun Hao, Qingfeng Chen, Haiming Pan, et al. (2022). *Enhancing drug–drug interaction prediction by three-way decision and knowledge graph embedding.* Granular Computing.

[283] Nasrullah Khan, Z. Ma, Li Yan, et al. (2022). *Hashing-based semantic relevance attributed knowledge graph embedding enhancement for deep probabilistic recommendation.* Applied intelligence (Boston).

[284] Thanh-Binh Le, Ngoc Huynh, and Bac Le (2022). *Knowledge graph embedding by projection and rotation on hyperplanes for link prediction.* Applied intelligence (Boston).

[285] Shuang Liang (2023). *Knowledge Graph Embedding Based on Graph Neural Network.* IEEE International Conference on Data Engineering.

[286] Nasrullah Khan, Zongmin Ma, Aman Ullah, et al. (2022). *DCA-IoMT: Knowledge-Graph-Embedding-Enhanced Deep Collaborative Alert Recommendation Against COVID-19.* IEEE Transactions on Industrial Informatics.

[287] Peng He, Gang Zhou, Mengli Zhang, et al. (2022). *Improving temporal knowledge graph embedding using tensor factorization.* Applied intelligence (Boston).

[288] Linshan Shen, Rongbo He, and Shaobin Huang (2022). *Entity alignment with adaptive margin learning knowledge graph embedding.* Data Knowledge Engineering.

[289] Shimin Di, Quanming Yao, Yongqi Zhang, et al. (2021). *Efficient Relation-aware Scoring Function Search for Knowledge Graph Embedding.* IEEE International Conference on Data Engineering.

[290] Guanglin Niu, Bo Li, Yongfei Zhang, et al. (2020). *AutoETER: Automated Entity Type Representation with Relation-Aware Attention for Knowledge Graph Embedding.* Findings.

[291] H. Nie, Xiangguo Zhao, Xin Bi, et al. (2023). *Correlation embedding learning with dynamic semantic enhanced sampling for knowledge graph completion.* World wide web (Bussum).

[292] Jiayi Li, and Yujiu Yang (2022). *STaR: Knowledge Graph Embedding by Scaling, Translation and Rotation.* Autonomous Infrastructure, Management and Security.

[293] A. Daruna, Devleena Das, and S. Chernova (2022). *Explainable Knowledge Graph Embedding: Inference Reconciliation for Knowledge Inferences Supporting Robot Actions.* IEEE/RJS International Conference on Intelligent RObots and Systems.

[294] Xing-Chun Zhou, Peng Wang, Qi Luo, et al. (2021). *Multi-hop Knowledge Graph Reasoning Based on Hyperbolic Knowledge Graph Embedding and Reinforcement Learning.* IJCKG.

[295] Kong Wei Kun, Xin Liu, Teeradaj Racharak, et al. (2023). *WeExt: A Framework of Extending Deterministic Knowledge Graph Embedding Models for Embedding Weighted Knowledge Graphs.* IEEE Access.

[296] Yao Dong, Lei Wang, Ji Xiang, et al. (2022). *RotateCT: Knowledge Graph Embedding by Rotation and Coordinate Transformation in Complex Space.* International Conference on Computational Linguistics.

[297] Hidetaka Kamigaito, and Katsuhiko Hayashi (2021). *Unified Interpretation of Softmax Cross-Entropy and Negative Sampling: With Case Study for Knowledge Graph Embedding.* Annual Meeting of the Association for Computational Linguistics.

[298] Franziska Krause (2022). *Dynamic Knowledge Graph Embeddings via Local Embedding Reconstructions.* Extended Semantic Web Conference.

[299] Zhao Zhang, Fuzhen Zhuang, Meng Qu, et al. (2021). *Knowledge graph embedding with shared latent semantic units.* Neural Networks.

[300] Guang-pu Li, Zequn Sun, Lei Qian, et al. (2021). *Rule-based data augmentation for knowledge graph embedding.* AI Open.

[301] Kai Wang, Yu Liu, Xiujuan Xu, et al. (2020). *Enhancing knowledge graph embedding by composite neighbors for link prediction.* Computing.

[302] Yuyang Wei, Wei Chen, Zhixu Li, et al. (2021). *Incremental Update of Knowledge Graph Embedding by Rotating on Hyperplanes*. 2021 IEEE International Conference on Web Services (ICWS).

[303] Yongqi Zhang, Quanming Yao, and Lei Chen (2021). *Simple and automated negative sampling for knowledge graph embedding*. The VLDB journal.

[304] Nasrullah Sheikh, Xiao Qin, B. Reinwald, et al. (2022). *Scaling knowledge graph embedding models for link prediction*. EuroMLSys@EuroSys.

[305] Chao Ren, Le Zhang, Lintao Fang, et al. (2021). *Ontological Concept Structure Aware Knowledge Transfer for Inductive Knowledge Graph Embedding*. IEEE International Joint Conference on Neural Network.

[306] Victoria Eyharabide, I. E. I. Bekkouch, and Nicolae Dragoș Constantin (2021). *Knowledge Graph Embedding-Based Domain Adaptation for Musical Instrument Recognition*. De Computis.

[307] Y. Hong, Chenyang Bu, and Tingting Jiang (2020). *Rule-enhanced Noisy Knowledge Graph Embedding via Low-quality Error Detection*. 2020 IEEE International Conference on Knowledge Graph (ICKG).

[308] Yan Huang, Haili Sun, Xu Ke, et al. (2020). *CoRelatE: Learning the correlation in multi-fold relations for knowledge graph embedding*. Knowledge-Based Systems.

[309] M. Kurokawa (2021). *Explainable Knowledge Reasoning Framework Using Multiple Knowledge Graph Embedding*. IJCKG.

[310] Sameh K. Mohamed, Emir Muñoz, and V. Nováček (2021). *On Training Knowledge Graph Embedding Models*. Inf..

[311] Thomas Gebhart, J. Hansen, and Paul Schrater (2021). *Knowledge Sheaves: A Sheaf-Theoretic Framework for Knowledge Graph Embedding*. International Conference on Artificial Intelligence and Statistics.

[312] Weibin Deng, Yiteng Zhang, Hong Yu, et al. (2024). *Knowledge graph embedding based on dynamic adaptive atrous convolution and attention mechanism for link prediction.* Information Processing Management.

[313] Jin Liu, Hao Du, R. Guo, et al. (2024). *MMGK: Multimodality Multiview Graph Representations and Knowledge Embedding for Mild Cognitive Impairment Diagnosis.* IEEE Transactions on Computational Social Systems.

[314] Chengcheng Zhang, Tianyi Zang, and Tianyi Zhao (2024). *KGE-UNIT: toward the unification of molecular interactions prediction based on knowledge graph and multi-task learning on drug discovery.* Briefings Bioinform..

[315] Mingsheng He, Lin Zhu, and Luyi Bai (2024). *ConvTKG: A query-aware convolutional neural network-based embedding model for temporal knowledge graph completion.* Neurocomputing.

[316] Dong Zhang, Zhe Rong, Chengyuan Xue, et al. (2024). *SimRE: Simple contrastive learning with soft logical rule for knowledge graph embedding.* Information Sciences.

[317] Dong Zhang, Wenlong Feng, Zonghang Wu, et al. (2024). *CDRGN-SDE: Cross-Dimensional Recurrent Graph Network with neural Stochastic Differential Equation for temporal knowledge graph embedding.* Expert systems with applications.

[318] Yanzhen Jing, Guanghui Zhou, Chao Zhang, et al. (2024). *XMKR: Explainable manufacturing knowledge recommendation for collaborative design with graph embedding learning.* Advanced Engineering Informatics.

[319] Pengcheng Jiang, Lang Cao, Cao Xiao, et al. (2024). *KG-FIT: Knowledge Graph Fine-Tuning Upon Open-World Knowledge.* Neural Information Processing Systems.

[320] Zhulin Han, and Jian Wang (2024). *Knowledge enhanced graph inference network based entity-relation extraction and knowledge graph construction for industrial domain.* Frontiers of Engineering Management.

[321] Huafeng Quan, Yiting Li, Dashuai Liu, et al. (2024). *Protection of Guizhou Miao batik culture based on knowledge graph and deep learning.* Heritage Science.

[322] Bufan Liu, Chun-Hsien Chen, and Zuoxu Wang (2024). *A multi-hierarchical aggregation-based graph convolutional network for industrial knowledge graph embedding towards cognitive intelligent manufacturing.* Journal of manufacturing systems.

[323] Nour Hello, P. Lorenzo, and E. Strinati (2024). *Semantic Communication Enhanced by Knowledge Graph Representation Learning.* International Workshop on Signal Processing Advances in Wireless Communications.

[324] Jinpeng Li, Hang Yu, Xiangfeng Luo, et al. (2024). *COSIGN: Contextual Facts Guided Generation for Knowledge Graph Completion.* North American Chapter of the Association for Computational Linguistics.

[325] Qun Yan, Juan Zhao, Linfu Xue, et al. (2024). *Mineral Prospectivity Mapping Based on Spatial Feature Classification with Geological Map Knowledge Graph Embedding: Case Study of Gold Ore Prediction at Wulonggou, Qinghai Province (Western China).* Natural Resources Research.

[326] Jhih-Chen Liu, Chiao-Ting Chen, Chi Lee, et al. (2024). *Evolving Knowledge Graph Representation Learning with Multiple Attention Strategies for Citation Recommendation System.* ACM Transactions on Intelligent Systems and Technology.

[327] Chuanghui Wang, Yunqing Yang, Jinshuai Song, et al. (2024). *Research Progresses and Applications of Knowledge Graph Embedding Technique in Chemistry.* Journal of Chemical Information and Modeling.

[328] Xiao Long, Liansheng Zhuang, Aodi Li, et al. (2024). *KGDM: A Diffusion Model to Capture Multiple Relation Semantics for Knowledge Graph Embedding.* AAAI Conference on Artificial Intelligence.

[329] Qihui Zhou, Peiqi Yin, Xiao Yan, et al. (2024). *Atom: An Efficient Query Serving System for Embedding-based Knowledge Graph Reasoning with Operator-level Batching.* Proc. ACM Manag. Data.

[330] Chen Huang, Deshan Chen, Tengze Fan, et al. (2024). *Incorporating environmental knowledge embedding and spatial-temporal graph attention networks for inland vessel traffic flow prediction.* Engineering applications of artificial intelligence.

[331] Ming Lu, Yancong Li, Jiangxiao Zhang, et al. (2024). *Deep hyperbolic convolutional model for knowledge graph embedding.* Knowledge-Based Systems.

[332] Qi Liu, Qinghua Zhang, Fan Zhao, et al. (2024). *Uncertain knowledge graph embedding: an effective method combining multi-relation and multi-path.* Frontiers Comput. Sci..

[333] Nasrullah Khan, Zongmin Ma, Ruizhe Ma, et al. (2024). *Continual knowledge graph embedding enhancement for joint interaction-based next click recommendation.* Knowledge-Based Systems.

[334] Zengcan Xue, Zhaoli Zhang, Hai Liu, et al. (2025). *MHRN: A multi-perspective hierarchical relation network for knowledge graph embedding.* Knowledge-Based Systems.

[335] Xiao Long, Liansheng Zhuang, Aodi Li, et al. (2024). *Fact Embedding through Diffusion Model for Knowledge Graph Completion.* The Web Conference.

[336] Chen Huang, Fei Yu, Zhiguo Wan, et al. (2024). *Knowledge graph confidence-aware embedding for recommendation.* Neural Networks.

[337] Yuzhuo Wang, Hongzhi Wang, Xianglong Liu, et al. (2024). *GFedKG: GNN-based federated embedding model for knowledge graph completion.* Knowledge-Based Systems.

[338] Xinyan Wang, Kuo Yang, Ting Jia, et al. (2024). *KDGene: knowledge graph completion for disease gene prediction using interactional tensor decomposition.* Briefings Bioinform..

[339] Yuhan Liu, Zelin Cao, Xing Gao, et al. (2024). *Bridging the Space Gap: Unifying Geometry Knowledge Graph Embedding with Optimal Transport.* The Web Conference.

[340] Yongfang Li, and Chunhua Zhu (2024). *TransE-MTP: A New Representation Learning Method for Knowledge Graph Embedding with Multi-Translation Principles and TransE.* Electronics.

[341] Qianjin Zhang, and Yandan Xu (2024). *Knowledge graph embedding with inverse function representation for link prediction.* Engineering applications of artificial intelligence.

[342] Hao Wang, Dandan Song, Zhijing Wu, et al. (2024). *A collaborative learning framework for knowledge graph embedding and reasoning.* Knowledge-Based Systems.

[343] Shengkun Ni, Xiangtai Kong, Yingying Zhang, et al. (2024). *Identifying compound-protein interactions with knowledge graph embedding of perturbation transcriptomics.* Cell Genomics.

[344] Jixuan Nie, Xia Hou, Wenfeng Song, et al. (2024). *Knowledge Graph Efficient Construction: Embedding Chain-of-Thought into LLMs.* VLDB Workshops.

[345] Jingchao Wang, Weimin Li, Fangfang Liu, et al. (2024). *ConeE: Global and local context-enhanced embedding for inductive knowledge graph completion.* Expert systems with applications.

[346] Yuren Mao, Yu Hao, Xin Cao, et al. (2024). *Dynamic Graph Embedding via Meta-Learning.* IEEE Transactions on Knowledge and Data Engineering.

[347] Parastoo Jafarzadeh, F. Ensan, Mahdiyar Ali Akbar Alavi, et al. (2024). *A Knowledge Graph Embedding Model for Answering Factoid Entity Questions.* ACM Trans. Inf. Syst..

[348] Yalin Wang, Yubin Peng, and Jingyu Guo (2024). *Enhancing knowledge graph embedding with structure and semantic features.* Applied intelligence (Boston).

[349] Yuhuan Lu, Weijian Yu, Xin Jing, et al. (2024). *HyperCL: A Contrastive Learning Framework for Hyper-Relational Knowledge Graph Embedding with Hierarchical Ontology.* Annual Meeting of the Association for Computational Linguistics.

[350] Yadan Han, Guangquan Lu, Shichao Zhang, et al. (2024). *A Temporal Knowledge Graph Embedding Model Based on Variable Translation.* Tsinghua Science and Technology.

[351] Bingchen Liu, Shifu Hou, Weiyi Zhong, et al. (2024). *Enhancing Temporal Knowledge Graph Alignment in News Domain With Box Embedding.* IEEE Transactions on Computational Social Systems.

[352] Yunjie He, Daniel Hernández, M. Nayyeri, et al. (2024). *Generating $SROI^-$ Ontologies via Knowledge Graph Query Embedding Learning.* Unpublished manuscript.

[353] Yan Fang, Xiaodong Liu, Wei Lu, et al. (2024). *Knowledge graph completion with low-dimensional gated hierarchical hyperbolic embedding.* Knowledge-Based Systems.

[354] Mingtao Zhang, Guoli Yang, Yi Liu, et al. (2024). *Knowledge graph accuracy evaluation: an LLM-enhanced embedding approach.* International Journal of Data Science and Analysis.

[355] Yicong Li, Yu Yang, Jiannong Cao, et al. (2024). *Toward Structure Fairness in Dynamic Graph Embedding: A Trend-aware Dual Debiasing Approach.* Knowledge Discovery and Data Mining.

[356] Dibo Dong, Shangwei Wang, Qiaoying Guo, et al. (2024). *Short-Term Marine Wind Speed Forecasting Based on Dynamic Graph Embedding and Spatiotemporal Information.* Journal of Marine Science and Engineering.

[357] Tao Wang, Bo Shen, Jinglin Zhang, et al. (2024). *Knowledge Graph Embedding via Triplet Component Interactions.* Neural Processing Letters.

[358] Pengfei Zhang, Xiaoxue Zhang, Yang Fang, et al. (2024). *Knowledge Graph Embedding for Hierarchical Entities Based on Auto-Embedding Size.* Mathematics.

[359] K. Liang, Yue Liu, Hao Li, et al. (2024). *Clustering then Propagation: Select Better Anchors for Knowledge Graph Embedding.* Neural Information Processing Systems.

[360] Qi Liu, Yuanyuan Jin, Xuefei Cao, et al. (2024). *An Entity Ontology-Based Knowledge Graph Embedding Approach to News Credibility Assessment.* IEEE Transactions on Computational Social Systems.

[361] H. V. Pham, Trung Tuan Nguyen, Luu Minh Tuan, et al. (2024). *IDGCN: A Proposed Knowledge Graph Embedding With Graph Convolution Network For Context-Aware Recommendation Systems.* Journal of Organizational Computing and Electronic Commerce.

[362] Yu Li, Zhu-Hong You, Shu-Min Wang, et al. (2024). *Attention-Based Learning for Predicting Drug-Drug Interactions in Knowledge Graph Embedding Based on Multisource Fusion Information.* International Journal of Intelligent Systems.

[363] Nan Li, Zhihao Yang, Jian Wang, et al. (2024). *Drug–target interaction prediction using knowledge graph embedding.* iScience.

[364] Liming Bao, Yan Wang, Xiaoyu Song, et al. (2024). *HGCGE: hyperbolic graph convolutional networks-based knowledge graph embedding for link prediction.* Knowledge and Information Systems.

[365] Guoshun Xu, Guozheng Rao, Li Zhang, et al. (2024). *Entity-relation aggregation mechanism graph neural network for knowledge graph embedding.* Applied intelligence (Boston).

[366] Qiuyu Liang, Weihua Wang, Jie Yu, et al. (2024). *Effective Knowledge Graph Embedding with Quaternion Convolutional Networks.* Natural Language Processing and Chinese Computing.

[367] Jie Liu, Lizheng Zu, Yunbin Yan, et al. (2024). *Multi-Filter soft shrinkage network for knowledge graph embedding.* Expert systems with applications.

[368] Jie Dong, Cuiping Chen, Chi Zhang, et al. (2025). *Knowledge Graph Embedding With Graph Convolutional Network and Bidirectional Gated Recurrent Unit for Fault Diagnosis of Industrial Processes.* IEEE Sensors Journal.

[369] Sensen Zhang, Xun Liang, Simin Niu, et al. (2025). *Integrating Large Language Models and Möbius Group Transformations for Temporal Knowledge Graph Embedding on the Riemann Sphere.* AAAI Conference on Artificial Intelligence.

[370] Xinyue Liu, Jianan Zhang, Chi Ma, et al. (2024). *Temporal Knowledge Graph Reasoning with Dynamic Hypergraph Embedding.* International Conference on Language Resources and Evaluation.

[371] Ruiyi Yang, Flora D. Salim, and Hao Xue (2024). *SSTKG: Simple Spatio-Temporal Knowledge Graph for Intepretable and Versatile Dynamic Information Embedding.* The Web Conference.

[372] Bo Li, Haowei Quan, Jiawei Wang, et al. (2024). *Neural Library Recommendation by Embedding Project-Library Knowledge Graph.* IEEE Transactions on Software Engineering.

[373] Xiaojian Liu, Xinwei Guo, and Wen Gu (2024). *SecKG2vec: A novel security knowledge graph relational reasoning method based on semantic and structural fusion embedding*. Computers security.

[374] Bin Chen, Hongyi Li, Di Zhao, et al. (2024). *Quality assessment of cyber threat intelligence knowledge graph based on adaptive joining of embedding model*. Complex amp; Intelligent Systems.

[375] Deng Chen, Weiwei Zhang, and Zuohua Ding (2024). *Embedding dynamic graph attention mechanism into Clinical Knowledge Graph for enhanced diagnostic accuracy*. Expert systems with applications.

[376] Quan Wang, Zhendong Mao, Bin Wang, et al. (2017). *Knowledge Graph Embedding: A Survey of Approaches and Applications*. IEEE Transactions on Knowledge and Data Engineering.

[377] Zhifei Li, Hai Liu, Zhaoli Zhang, et al. (2021). *Learning Knowledge Graph Embedding With Heterogeneous Relation Attention Networks*. IEEE Transactions on Neural Networks and Learning Systems.