

# 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 rapid proliferation of digital information across diverse domains has ushered in an era where the ability to effectively organize, interpret, and leverage vast quantities of data is paramount for the advancement of Artificial Intelligence (AI) systems. At the forefront of this organizational paradigm are Knowledge Graphs (KGs), which have emerged as a foundational technology for structuring complex, interconnected information into a machine-readable format. KGs represent real-world entities and their relationships as a collection of triples (head entity, relation, tail entity), thereby providing a rich, semantic layer over raw data. This structured representation facilitates a deeper understanding of information, enabling AI systems to move beyond pattern recognition to more sophisticated reasoning, inference, and decision-making capabilities. From powering search engines and recommendation systems to facilitating drug discovery and scientific research, KGs are increasingly indispensable components of modern AI architectures.

However, the inherent symbolic nature of KGs, while offering precision and interpretability, presents significant challenges when confronted with the demands of modern AI. Traditional symbolic knowledge representation methods, relying on discrete symbols and explicit logical rules, often struggle with the scale, sparsity, and inherent fuzziness of real-world knowledge. They are computationally expensive for large graphs, brittle in the face of noise or incompleteness, and fundamentally incompatible with the continuous vector spaces that underpin most contemporary machine learning (ML) models. This fundamental mismatch necessitates a transformative approach: the conversion of KGs into dense, continuous vector spaces, a process known as Knowledge Graph Embedding (KGE).

The core motivation behind KGE is to overcome the limitations of symbolic representation by learning low-dimensional, continuous vector representations (embeddings) for entities and relations within a KG. These embeddings aim to capture the latent semantic and structural properties of the graph, allowing for efficient computation, enhanced scalability, and seamless integration with a wide array of machine learning algorithms.

By transforming discrete symbols into a continuous space, KGE models can perform tasks such as link prediction (inferring missing relationships), entity alignment (identifying equivalent entities across different KGs), and question answering with unprecedented efficiency and accuracy. For instance, models like ? leverage KGE for bootstrapping entity alignment, demonstrating how embedding-based approaches can effectively overcome data scarcity challenges in integrating KGs. Similarly, the work by ? highlights the critical need for parameter-efficient KGE models to ensure scalability and deployability in resource-constrained environments, directly addressing the computational burden often associated with large KGs.

This comprehensive literature review aims to synthesize and critically analyze the evolution of KGE research, providing a pedagogical progression from its foundational principles to its cutting-edge applications and future directions. We will explore how KGE models have evolved to capture increasingly complex aspects of knowledge, including temporal dynamics ??, multi-modal information, and hierarchical structures. The review will not merely enumerate existing models but will delve into the underlying theoretical assumptions, comparative strengths and weaknesses, and the persistent challenges that continue to drive innovation in the field. By constructing explicit comparative frameworks and identifying recurring patterns, tensions, and unresolved debates, we seek to provide a holistic understanding of the intellectual trajectory of KGE research, highlighting both its profound successes and its enduring frontiers.

## 1.1 1.1. Background: The Rise of Knowledge Graphs

The landscape of modern AI is increasingly defined by its capacity to process and reason over vast, intricate datasets. In this context, Knowledge Graphs (KGs) have emerged as a pivotal technology, offering a structured and semantic framework for organizing information that transcends the limitations of unstructured text or tabular data ????. A KG fundamentally represents knowledge as a directed graph, where nodes denote real-world entities (e.g., "Albert Einstein," " $E=mc^2$ ") and edges represent semantic relations between them (e.g., "discovered," "is\_a") ?. Each piece of information is typically ex-

pressed as a triple (head entity, relation, tail entity), such as (Albert Einstein, discovered,  $E=mc^2$ ). This explicit, machine-readable structure allows AI systems to not only store facts but also to understand the relationships and context surrounding those facts, which is crucial for higher-level cognitive tasks.

The prevalence of KGs has surged due to their ability to provide a rich, interconnected web of facts that can power a multitude of AI applications. For instance, KGs are instrumental in enhancing search engine relevance by understanding user intent and providing direct answers rather than just links [1]. In recommendation systems, KGs enable more personalized and explainable recommendations by modeling complex user-item interactions and underlying reasons for preferences [2]. For example, [3] introduces Contextualized Knowledge Graph Embedding (CKGE) for explainable talent training course recommendation, demonstrating how integrating motivation-aware information and high-order connections within a KG can yield superior and interpretable results. This work highlights a critical tension in recommendation systems: the trade-off between predictive accuracy and explainability, which KGE-based approaches like CKGE aim to resolve by leveraging the structured nature of KGs.

Beyond these common applications, KGs are increasingly vital in specialized domains. In biomedicine, they facilitate drug discovery, repurposing, and adverse drug reaction prediction by mapping complex biological pathways and drug-target interactions [4]. For instance, [5] evaluates KGE approaches for drug-drug interaction prediction, underscoring the practical utility of KGs in identifying critical medical insights. Similarly, KGs support entity alignment across heterogeneous data sources, a crucial task for integrating disparate knowledge bases [6]. The work by [7], "Bootstrapping Entity Alignment with Knowledge Graph Embedding," exemplifies this by proposing a novel bootstrapping approach that iteratively labels likely entity alignments to expand training data for alignment-oriented KG embeddings. This method addresses the significant challenge of limited prior alignment, which often plagues entity alignment tasks, by using a global optimal labeling strategy based on max-weighted matching and an alignment editing method to reduce error propagation. This directly contrasts with earlier

methods like IPTransE [1] which relied on local optimal distance measures and were highly sensitive to initial alignment precision, leading to error accumulation.

Despite their immense utility, KGs are not without their inherent limitations. A primary challenge is their incompleteness; real-world KGs are often sparse, with many potential facts missing [2]. This sparsity is exacerbated by the dynamic nature of knowledge, where facts evolve over time, necessitating temporal awareness [3]. Furthermore, KGs can be noisy, containing erroneous or inconsistent information, especially when constructed automatically. The sheer scale of many KGs, comprising millions of entities and billions of facts, also poses significant computational and storage challenges. For example, the need for parameter-efficient models, as highlighted by [4], becomes critical when dealing with KGs that can have hundreds of millions of parameters, making deployment on resource-constrained devices or in federated learning scenarios impractical. These limitations underscore that while KGs provide an unparalleled framework for knowledge organization, their full potential can only be realized through advanced representation techniques that can overcome these intrinsic hurdles. The evolution of KGE research is a direct response to these challenges, aiming to transform KGs into a more flexible, scalable, and computationally amenable format.

## 1.2 1.2. The Imperative for Knowledge Graph Embedding

The intrinsic value of Knowledge Graphs (KGs) in structuring complex information for AI systems is undeniable, yet their traditional symbolic representation presents a formidable bottleneck to their full utilization. Symbolic methods, which represent entities and relations as discrete tokens and rely on explicit logical rules for inference, are inherently limited in several critical ways. Firstly, they suffer from **sparsity and incompleteness** [5]. Real-world KGs are never exhaustive; many valid facts are simply unknown or unrecorded. Symbolic systems struggle to infer these missing links without explicit rules, leading to brittle and incomplete knowledge bases. Secondly, symbolic representations are **computationally expensive and lack scalability** for large-scale KGs. Performing complex logical inference over millions of entities and billions of relations can lead

to combinatorial explosions, making real-time reasoning impractical ?. Thirdly, they are **sensitive to noise and lack robustness**. A single erroneous fact or rule can propagate errors throughout the system, leading to incorrect inferences. Finally, and perhaps most critically for modern AI, symbolic representations are **fundamentally incompatible with the continuous vector spaces** that underpin the vast majority of contemporary machine learning (ML) models, particularly deep learning architectures. This incompatibility creates a chasm between structured knowledge and powerful data-driven learning paradigms, hindering the seamless integration of knowledge into ML systems ?.

This growing chasm necessitated a paradigm shift, giving rise to the imperative for Knowledge Graph Embedding (KGE). KGE addresses these limitations by transforming discrete symbolic entities and relations into dense, low-dimensional, continuous vector representations (embeddings) in a latent semantic space ??????. The core innovation lies in capturing the latent semantic and structural properties of the KG within these vector spaces, allowing for geometric interpretations of relationships. For instance, translational models like TransE ? conceptualize relations as translation vectors between entity embeddings (i.e.,  $h + r \approx t$ ), while rotational models like RotatE ? represent relations as rotations in complex vector spaces. This transformation offers several profound benefits that are critical for the advancement of AI.

Firstly, KGE enables **efficient computation and scalability**. Once entities and relations are embedded as vectors, operations like similarity calculations or relation predictions become simple vector arithmetic, which can be performed efficiently on GPUs. This drastically reduces the computational cost compared to symbolic inference, allowing KGE models to handle massive KGs with millions of entities and billions of triples ??. The work by ? on "Entity-Agnostic Representation Learning for Parameter-Efficient Knowledge Graph Embedding" directly tackles the scalability challenge by proposing a method that learns embeddings without the parameter count scaling linearly with the number of entities. This is a crucial innovation, as conventional KGE models can require colossal parameter counts (e.g., 123 million for RotatE on YAGO3-10), making them impractical for deployment on edge devices or in federated learning scenarios. By focusing on entity-

agnostic encoding, ? offers a path toward more sustainable and deployable KGE models, demonstrating that significant parameter reduction can be achieved without sacrificing performance. This highlights a critical, often unstated, assumption in KGE research: that entity embeddings must be unique and directly learned for each entity. ? challenges this by showing that rich representations can be learned compositionally from local context.

Secondly, KGE significantly enhances **generalization and robustness**. By learning continuous representations, KGE models can infer missing links (link prediction) by identifying plausible relationships in the embedding space, even if they have never been explicitly observed in the KG ?. This addresses the inherent incompleteness of KGs. For example, methods like ?’s TransA, which introduces "Locally Adaptive Translation," improve generalization by adaptively determining optimal margins for the loss function based on the specific structure and locality of the knowledge graph. This is a critical advancement over prior methods like TransE and TransH, which relied on global, fixed margins that often led to suboptimal performance across diverse KG subsets. TransA’s innovation lies in its ability to theoretically justify and empirically demonstrate that different graph localities require different optimal loss functions, thereby making the embedding process more sensitive and robust to the nuances of KG structure. This directly addresses the implicit assumption that a single, globally tuned hyperparameter (like the margin) is sufficient for all parts of a heterogeneous KG.

Thirdly, KGE ensures **seamless integration with machine learning models**. The output of KGE models—dense vector embeddings—can be directly fed as features into various downstream ML tasks, including natural language processing, computer vision, and recommendation systems. This bridges the gap between structured knowledge and statistical learning, allowing ML models to leverage rich semantic information that was previously inaccessible or difficult to incorporate. This integration is exemplified in applications like explainable recommendation systems ? or question answering systems ??, where KGE provides the semantic backbone for intelligent agents.

Finally, KGE models have evolved to capture increasingly complex aspects of knowledge, such as **temporal dynamics**. Early KGE models treated KGs as static, ignoring



the time-varying nature of facts. However, the recognition that facts have temporal validity (e.g., "Barack Obama is President of the USA" is only true for a specific period) led to the development of Temporal Knowledge Graph Embedding (TKGE) models [1, 2]. For instance, [3]’s HyTE (Hyperplane-based Temporally aware Knowledge Graph Embedding) explicitly incorporates time by associating each timestamp with a hyperplane, allowing for temporally guided inference and prediction of temporal scopes for missing annotations. This is a significant advancement over static KGEs, which would treat facts like (Obama, presidentOf, USA, 2010) and (Obama, presidentOf, USA, 2020) identically, leading to inaccurate reasoning. Building on this, [4] proposes ATiSE (Additive Time Series Embedding), which models entity and relation representations as multi-dimensional Gaussian distributions evolving via additive time series decomposition (trend, seasonal, random components). This novel approach explicitly captures *temporal uncertainty* in knowledge evolution, a critical limitation ignored by many prior TKGE models that assumed deterministic changes. The tension here lies between the simplicity and efficiency of static embeddings versus the expressive power and accuracy required for dynamic, real-world knowledge.

The field also grapples with the challenge of **disentangled representations**, where entities might have multiple facets, and relations might focus on distinct aspects. [5]’s DisenKGAT (Disentangled Knowledge Graph Attention Network) addresses this by learning disentangled entity representations, incorporating both micro-disentanglement (relation-aware aggregation) and macro-disentanglement (mutual information regularization) to ensure independence between learned components. This innovation directly confronts the limitation of single, static entity representations that fail to capture the complex, multi-faceted nature of entities and the context-dependency of relations, a problem that even advanced GNN-based KGC models often overlook. The implicit assumption that a single vector can adequately represent all aspects of an entity is systematically questioned and addressed by DisenKGAT.

In summary, the imperative for KGE stems from the inherent limitations of symbolic knowledge representation in the face of massive, dynamic, and often incomplete real-world

KGs. KGE provides a powerful framework for transforming these symbolic structures into a computationally efficient, scalable, generalizable, and ML-compatible format, thereby unlocking the full potential of KGs in modern AI systems. The evolution of KGE models reflects a continuous effort to address these challenges, pushing the boundaries of what knowledge-driven AI can achieve.

### 1.3 1.3. Scope and Structure of the Review

Having established the foundational importance of Knowledge Graphs (KGs) in modern AI and the compelling imperative for Knowledge Graph Embedding (KGE) to overcome the limitations of symbolic representation, this literature review now delineates its scope and pedagogical structure. The primary objective is not merely to catalog existing KGE models but to provide a comprehensive, critical, and synthetic analysis of the field’s intellectual evolution, identifying key paradigms, recurring challenges, and future trajectories. This review aims to serve as a structured guide for researchers and practitioners, offering insights into the underlying mechanisms, comparative performance, and theoretical underpinnings of various KGE approaches.

The pedagogical progression of this review is designed to build knowledge systematically, starting from the foundational principles and gradually advancing to more complex and specialized topics. We will begin by categorizing KGE models into major families, such as translational models, semantic matching models, and graph neural network-based approaches. This initial categorization will establish a comparative framework, highlighting how different model families (e.g., TransE ?, DistMult, ComplEx, RotatE ?) share the common goal of embedding entities and relations but diverge significantly in their scoring functions and geometric interpretations. For instance, while translational models focus on vector addition, rotational models leverage complex-space rotations, each offering distinct advantages in capturing specific relational patterns like symmetry, anti-symmetry, and compositionality ?. A critical analysis will compare these foundational approaches on dimensions such as expressiveness, computational complexity, and their ability to model different relation types, drawing on empirical evaluations like those in ? which assess

the effects of hyperparameters on KGE quality across various datasets. The work by [1] is particularly insightful as it empirically demonstrates that hyperparameter sensitivities vary substantially between knowledge graphs, suggesting that optimal tuning strategies are dataset-specific. This finding implicitly questions the common assumption in many KGE papers that a single set of hyperparameters can generalize well across diverse KG structures.

Following this foundational overview, the review will delve into advanced KGE techniques that address specific challenges and extend the capabilities of initial models. This includes the integration of auxiliary information, such as entity types [2], textual descriptions [3], and logical rules [4], to enrich embeddings and improve performance. We will critically examine how these extensions enhance model expressiveness and robustness, often by mitigating the sparsity problem inherent in KGs. For example, methods that incorporate textual information (e.g., [5]) aim to leverage the rich semantics embedded in natural language descriptions to compensate for sparse relational data, thereby improving embedding quality, especially for long-tail entities. A key tension here is the trade-off between the complexity introduced by integrating diverse data sources and the resulting gains in performance and interpretability.

A significant portion of the review will be dedicated to the evolution of KGE models in handling dynamic and complex knowledge. This includes Temporal Knowledge Graph Embedding (TKGE) models, which explicitly account for the time-varying nature of facts [6]. We will analyze how models like HyTE [7] and ATiSE [8] represent time, comparing their approaches to capturing temporal validity and uncertainty. HyTE’s use of hyperplanes for timestamps offers a geometric interpretation of temporal facts, while ATiSE’s Gaussian distributions and additive time series decomposition provide a novel way to model the evolution and inherent uncertainty of entity/relation representations over time. This comparison will highlight the shift from static to dynamic knowledge representation, a critical paradigm change driven by the need for more realistic and context-aware AI systems. Furthermore, the review will explore models designed for hyper-relational data [9] and those that learn disentangled representations [10], addressing

the multi-faceted nature of entities and relations. DisenKGAT ?, for example, challenges the assumption of single, static entity representations by learning independent components, thereby enhancing interpretability and robustness, especially for complex relation types.

The synthesis and critical analysis throughout the review will focus on identifying recurring trade-offs (e.g., expressiveness vs. efficiency, interpretability vs. performance), methodological trends (e.g., from simple linear models to deep neural architectures), and unresolved debates (e.g., the optimal geometry for embedding spaces, the best negative sampling strategies ???). We will systematically question assumptions, such as the generalizability of evaluation metrics or the implicit assumption of uniform confidence in KG facts, as addressed by works like ? which proposes confidence-aware negative sampling for noisy KGs. The review will also highlight evaluation gaps, such as the lack of consistent benchmarks for certain complex KGE tasks or the limited focus on real-world deployment challenges, including parameter efficiency ? and federated learning scenarios ??????.

Finally, the review will conclude by outlining cutting-edge applications of KGE across various domains and discussing promising future directions. These include the integration of KGE with Large Language Models (LLMs) ??, advancements in explainable KGE ???, robust KGE against adversarial attacks ???, and the development of more adaptive and efficient learning systems ?. By providing a structured narrative that connects foundational principles to emerging trends, this review aims to offer a profound understanding of how KGE has transformed the utility of knowledge graphs in AI and where the field is headed next.

## 2. Foundational Concepts and Early Geometric Models

The transition from symbolic knowledge representation to continuous vector spaces marks a pivotal paradigm shift in the field of Knowledge Graphs (KGs), driven by the imperative to overcome the inherent limitations of discrete symbols, as discussed in Section 1.2 ??.

This section lays the groundwork for understanding Knowledge Graph Embedding (KGE) by introducing its fundamental concepts, particularly the representation of entities and relations in low-dimensional vector spaces. It then delves into the pioneering 'translational distance models,' which established a foundational geometric paradigm for KGE. These early models, including TransE, TransH, and TransR, significantly advanced the field by offering efficient ways to capture basic relational patterns, thereby enabling scalable link prediction and other downstream tasks. We will critically analyze their core mechanisms, comparative strengths, and inherent limitations, underscoring the continuous drive for enhanced representational power in KGE.

The core idea behind KGE is to embed entities and relations into a continuous vector space, typically  $\mathbb{R}^k$ , where  $k$  is the dimension of the embedding. Each entity  $e \in \mathcal{E}$  (the set of entities) is represented by a vector  $\mathbf{e} \in \mathbb{R}^k$ , and each relation  $r \in \mathcal{R}$  (the set of relations) is also represented by a vector  $\mathbf{r} \in \mathbb{R}^k$  or a matrix/tensor. The goal is to learn these embeddings such that the structural and semantic properties of the KG are preserved. Specifically, for a given triple  $(h, r, t)$  where  $h$  is the head entity,  $r$  is the relation, and  $t$  is the tail entity, a scoring function  $f_r(h, t)$  is defined. This function measures the plausibility of the triple, aiming for a low score for valid triples and a high score for invalid ones. The embeddings are learned by minimizing a margin-based ranking loss or a logistic loss over observed (positive) and corrupted (negative) triples. Negative sampling, where valid triples are corrupted by replacing the head or tail entity, is a crucial technique in this process, though its effectiveness and biases have been subjects of ongoing research. For instance, [1] highlights that most KGE models implicitly assume uniform confidence in KG facts, which is often violated in noisy real-world KGs. They propose a confidence-aware negative sampling method to address this, demonstrating a critical tension between idealized model assumptions and practical data characteristics.

The early geometric models, particularly the translational family, conceptualized relations as operations that transform entity embeddings in the vector space. This geometric intuition provided a simple yet powerful framework for modeling relational patterns. TransE [2] pioneered this approach by proposing that for a valid triple  $(h, r, t)$ , the embed-

ding of the head entity  $\mathbf{h}$  plus the embedding of the relation  $\mathbf{r}$  should be approximately equal to the embedding of the tail entity  $\mathbf{t}$  (i.e.,  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ ). This elegant formulation offered a computationally efficient way to capture basic relational patterns, making it a cornerstone for subsequent KGE research. However, its simplicity also introduced limitations, particularly in handling complex relation types such as one-to-many (1-N), many-to-one (N-1), and many-to-many (N-N) relations, as well as symmetric relations. These limitations spurred immediate extensions like TransH [1] and TransR [2], which sought to enhance representational power by introducing more flexible geometric transformations, such as projecting entities onto relation-specific hyperplanes or spaces.

The continuous drive for enhanced representational power led to further innovations, moving beyond strict translational assumptions. TransD [3] introduced dynamic mapping matrices, allowing entities to have distinct projection vectors when acting as head or tail entities for different relations, thereby increasing expressiveness. ManifoldE [4] took a different approach by expanding entity representations from single points to manifolds, offering a more flexible geometric form to capture richer semantic information and address the limitations of rigid point-based embeddings. This evolution from simple vector addition to more complex geometric transformations—hyperplanes, dynamic matrices, and manifolds—reflects a fundamental tension in KGE research: balancing model simplicity and computational efficiency with the need to capture the intricate, diverse, and often ambiguous semantics of real-world knowledge graphs. The parameter efficiency of these models also became a concern, as highlighted by [5], where conventional KGEs can lead to colossal parameter counts, hindering deployment in resource-constrained environments. This section will critically examine these foundational models, their contributions, and the challenges they posed, setting the stage for more advanced KGE paradigms.

## 2.1 Core Principles of Knowledge Graph Embedding

The foundational principle of Knowledge Graph Embedding (KGE) is to represent entities and relations as continuous, low-dimensional vectors (embeddings) in a latent semantic space, typically  $\mathbb{R}^k$  [6]. This transformation from discrete symbols to dense vectors

is crucial for bridging the gap between symbolic knowledge representation and modern machine learning algorithms, which operate most effectively on continuous data. The primary objective is to learn these embeddings such that the inherent structure and semantics of the Knowledge Graph (KG) are preserved, enabling computational efficiency and generalization.

At its heart, KGE operates on the assumption that a valid triple  $(h, r, t)$  (head entity, relation, tail entity) should exhibit a particular geometric or algebraic relationship between their respective embeddings  $(\mathbf{h}, \mathbf{r}, \mathbf{t})$  in the vector space. This relationship is quantified by a **scoring function**,  $f_r(h, t)$ , which assigns a plausibility score to each triple. The design of this scoring function is a defining characteristic of different KGE models. For instance, in translational models, the score is often based on a distance metric, while in semantic matching models, it might involve dot products or bilinear forms. The lower the score for a positive triple and the higher for a negative (corrupted) triple, the better the embeddings are considered to be.

The learning process for these embeddings typically involves minimizing a **loss function**, which aims to maximize the scores of positive triples while minimizing the scores of negative triples. A common approach is the **margin-based ranking loss**, exemplified by:  $L = \sum_{(h, r, t) \in S} \sum_{(h', r', t') \in S'} \max(0, \gamma + f_r(h, t) - f_r(h', t'))$  where  $S$  is the set of positive triples,  $S'$  is the set of negative triples, and  $\gamma$  is a margin hyperparameter. The selection of an appropriate margin is critical, and as highlighted by [1] with their TransA model, a globally fixed margin can be suboptimal for heterogeneous KGs, leading to the need for adaptive margin determination. This demonstrates a tension between model simplicity and the need to capture local structural nuances.

A crucial component of training KGE models is **negative sampling** [2]. Since KGs are typically sparse, explicitly enumerating all negative triples is infeasible. Instead, negative triples are generated by corrupting positive triples, usually by replacing either the head or tail entity with a randomly chosen entity from the KG. For example, for a positive triple  $(h, r, t)$ , negative samples could be  $(h', r, t)$  or  $(h, r, t')$ . The quality of negative samples significantly impacts the learning process. Naive uniform negative sampling can be

inefficient, as many sampled negative triples might be easily distinguishable from positive ones. This led to more sophisticated strategies, such as **self-adversarial negative sampling** [1], which samples negative triples that are harder to distinguish, thereby forcing the model to learn finer distinctions. [2] introduces  $\epsilon$ -truncated uniform negative sampling, limiting the sampling scope to  $s$ -nearest neighbors to generate more challenging negatives, directly addressing the inefficiency of arbitrary sampling. This evolution in negative sampling techniques reflects a recurring pattern in KGE research: the refinement of training methodologies to improve learning efficiency and embedding quality.

The embeddings themselves are typically initialized randomly and then iteratively updated using optimization algorithms like Stochastic Gradient Descent (SGD) or Adam. The dimension  $k$  of the embedding space is a crucial hyperparameter, balancing representational capacity with computational cost and the risk of overfitting. A higher dimension allows for more information to be encoded but requires more parameters and computational resources [3]. The work by [4] empirically demonstrates the substantial variability in hyperparameter sensitivities, including embedding dimension, across different KGs, suggesting that optimal tuning strategies are highly dataset-specific. This finding challenges the implicit assumption that a universal set of hyperparameters exists for all KGE tasks and datasets.

The learned embeddings offer several advantages:

1. **Generalization and Link Prediction:** By capturing latent semantic relationships, KGE models can predict missing links (i.e., infer new facts) by identifying plausible connections in the embedding space [5]. This addresses the inherent incompleteness of KGs.
2. **Computational Efficiency:** Vector operations are highly optimized, allowing KGE models to handle large-scale KGs more efficiently than symbolic reasoning systems [6]. However, parameter efficiency remains a concern, as highlighted by [7], where conventional KGE models can lead to colossal parameter counts (e.g., 123 million for RotatE on YAGO3-10). Their proposed Entity-Agnostic Representation Learning (EARL) directly addresses this by learning representations that do not



scale linearly with the number of entities, challenging the implicit assumption that each entity requires a unique, directly learned embedding.

3. **Integration with Machine Learning:** The continuous vector representations serve as rich features for various downstream ML tasks, such as question answering [10], recommendation systems [11], and entity alignment [12]. For example, [13]’s CKGE leverages contextualized KGE for explainable talent training course recommendation, demonstrating how embeddings can capture motivation-aware information for more nuanced and interpretable outputs.

In essence, the core principles of KGE revolve around transforming discrete knowledge into a continuous, geometrically interpretable space, enabling efficient learning and inference. This foundational shift paved the way for the development of various model families, starting with the pioneering translational models.

## 2.2 Translational Models: From TransE to TransH and TransR

The translational distance models represent a seminal family of Knowledge Graph Embedding (KGE) approaches, establishing a powerful geometric intuition for modeling relations. This family, beginning with TransE, conceptualizes relations as "translations" between entity embeddings in a low-dimensional vector space. This section traces the evolution within this family, from the foundational TransE to its more expressive successors, TransH and TransR, highlighting their innovations, capabilities, and limitations.

**TransE (Translating Embeddings) [14]:** \* **Context and Problem Solved:** TransE was one of the first and most influential KGE models, addressing the problem of efficiently learning embeddings for large-scale KGs to perform link prediction. It aimed to capture the basic relational patterns (e.g., "capitalOf" as a translation from "France" to "Paris") in a simple and scalable manner. Prior methods often struggled with scalability or lacked a clear geometric interpretation of relations. \* **Core Innovation:** Its core innovation lies in its elegant and intuitive scoring function:  $f(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{L_1/L_2}$ . For a valid triple  $(h, r, t)$ , it posits that the embedding of the head entity  $\mathbf{h}$  plus the embedding of the relation  $\mathbf{r}$  should be close to the embedding of the tail entity  $\mathbf{t}$ . This makes relations

act as translation vectors in the embedding space. \* **Conditions for Success:** TransE performs well on datasets where relations primarily exhibit one-to-one (1-1) mapping properties. Its simplicity makes it computationally efficient and scalable for very large KGs. \* **Theoretical Limitations:** The primary theoretical limitation of TransE is its inability to effectively model complex relation types, specifically one-to-many (1-N), many-to-one (N-1), and many-to-many (N-N) relations, as well as symmetric relations. For instance, if "Obama" has two children, "Malia" and "Sasha," and "hasChild" is the relation, TransE would try to make **Obama + hasChild  $\approx$  Malia** and **Obama + hasChild  $\approx$  Sasha**. This forces **Malia** and **Sasha** to be very close in the embedding space, which is often semantically incorrect. Similarly, for symmetric relations (e.g., "friendOf"),  **$\mathbf{h} + \mathbf{r} \approx \mathbf{t}$**  implies  **$\mathbf{t} + \mathbf{r} \approx \mathbf{h}$** , which means  **$\mathbf{r} \approx -\mathbf{r}$** , forcing  **$\mathbf{r}$**  to be a zero vector, losing its meaning. \* **Practical Limitations:** While efficient, its representational power is limited, leading to suboptimal performance on KGs rich in complex relations. The choice of distance metric ( $L_1$  or  $L_2$ ) and the margin hyperparameter are crucial and often require extensive tuning ?.

**TransH (Translating on Hyperplanes) ?:** \* **Context and Problem Solved:** TransH was proposed to address TransE's limitations with 1-N, N-1, and N-N relations. It recognized that a single vector for an entity might be insufficient when it participates in different relations with different "aspects." \* **Core Innovation:** Instead of representing relations as direct translation vectors between entity embeddings, TransH models each relation  $r$  as a **hyperplane** (defined by its normal vector  $\mathbf{w}_r$ ) and a **translation vector  $\mathbf{d}_r$**  on that hyperplane. Entities are first projected onto this relation-specific hyperplane before the translation operation. The scoring function becomes  $f(h, r, t) = \|\mathbf{h}_{\perp} + \mathbf{d}_r - \mathbf{t}_{\perp}\|_{L_1/L_2}$ , where  $\mathbf{h}_{\perp} = \mathbf{h} - \mathbf{w}_r^{\top} \mathbf{h} \mathbf{w}_r$  and  $\mathbf{t}_{\perp} = \mathbf{t} - \mathbf{w}_r^{\top} \mathbf{t} \mathbf{w}_r$ . This allows an entity to have different "projected" representations for different relations, mitigating the issue of forcing multiple tail entities to be close. \* **Conditions for Success:** TransH significantly improves performance on KGs with complex relations compared to TransE. It offers a more flexible geometric interpretation without a drastic increase in computational complexity. \* **Theoretical Limitations:**

While better than TransE, TransH still assumes that entities and relations lie in the same semantic space. The projection mechanism might not be expressive enough to capture highly diverse semantic contexts, especially for relations with very different characteristics. The normal vector  $\mathbf{w}_r$  and translation vector  $\mathbf{d}_r$  are still learned independently for each relation, which can lead to issues with sparse relations. \* **Practical Limitations:** It introduces more parameters than TransE (a normal vector and a translation vector per relation), slightly increasing training complexity. The projection operation adds a small computational overhead.

**TransR (Translating in Relation Space) ?:** \* **Context and Problem Solved:** Building on TransH, TransR further enhances expressiveness by recognizing that entities and relations should not necessarily reside in the same semantic space. It addresses the problem that entities and relations are fundamentally different types of objects and should be embedded in distinct spaces. \* **Core Innovation:** TransR proposes to embed entities in an entity space  $\mathbb{R}^k$  and relations in a separate relation space  $\mathbb{R}^d$ . For each relation  $r$ , a **relation-specific projection matrix**  $\mathbf{M}_r \in \mathbb{R}^{d \times k}$  is introduced. Head and tail entities are first projected from the entity space to the relation space:  $\mathbf{h}_r = \mathbf{h}\mathbf{M}_r$  and  $\mathbf{t}_r = \mathbf{t}\mathbf{M}_r$ . The translation then occurs in this relation-specific space:  $f(h, r, t) = \|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_{L_1/L_2}$ . This allows for more flexible and context-dependent entity representations for each relation. \* **Conditions for Success:** TransR generally outperforms both TransE and TransH, especially on KGs with diverse and complex relations, as it provides a more nuanced way to model how entities interact under different relational contexts. \* **Theoretical Limitations:** The introduction of a projection matrix for each relation significantly increases the number of parameters, potentially leading to overfitting for sparse relations. The projection operation itself is linear, which might still limit its ability to capture highly non-linear semantic transformations. \* **Practical Limitations:** The increased parameter count (a  $d \times k$  matrix per relation) and the matrix multiplication operations make TransR computationally more expensive and memory-intensive than TransE and TransH. This can be a significant bottleneck for very large KGs or KGs with a vast number of unique relations.

**Comparative Framework and Evolution:** The translational models represent a clear evolutionary path driven by the need for greater expressiveness to handle complex relational patterns.

- **TransE** (Papers like ? build on this) is the simplest, assuming entities and relations exist in a single, unified space, and relations are direct translations. It excels in efficiency and 1-1 relations but struggles with 1-N/N-1/N-N and symmetric relations due to its rigid geometric assumption.
- **TransH** ? introduces relation-specific hyperplanes, allowing entities to have different "roles" or "aspects" depending on the relation. This addresses the 1-N/N-1/N-N problem by projecting entities, thereby relaxing the strict point-to-point translation.
- **TransR** ? takes this a step further by embedding entities and relations in *separate* spaces and using relation-specific projection matrices. This provides the most flexible entity representation within this family, as entities are transformed into a relation-specific semantic space before translation. This directly addresses the implicit assumption in TransE and TransH that entities and relations are semantically comparable in the same vector space.

This evolution showcases a recurring trade-off: increasing model complexity and parameter count for enhanced representational power. While TransE is highly efficient, its limitations spurred the development of more expressive, albeit more computationally demanding, models. The empirical evidence consistently shows TransR outperforming TransE and TransH on link prediction tasks across various benchmarks, demonstrating the value of its increased expressiveness ?. However, this comes at the cost of higher training time and memory footprint, a practical limitation that subsequent research often seeks to mitigate.

## 2.3 Early Extensions: Dynamic Mappings and Manifold Embeddings

The foundational translational models, while groundbreaking, revealed inherent limitations in capturing the full spectrum of relational semantics within Knowledge Graphs. This spurred immediate research into extensions that sought to enhance representational power by introducing more flexible geometric transformations. Two notable directions in this early phase were models incorporating dynamic mapping matrices and those expanding entity representations to manifolds.

**TransD (Translating with Dynamic Mapping Matrix) ?:** \* **Context and Problem Solved:** TransD emerged as a direct response to the limitations of TransR, particularly its large number of parameters (a full projection matrix for each relation) and its potential for overfitting on sparse relations. TransR’s projection matrices  $\mathbf{M}_r$  were relation-specific but entity-agnostic, meaning the same matrix was applied regardless of the specific head or tail entity. TransD aimed to make these projections more dynamic and entity-aware. ***Core Innovation:** TransD introduces **dynamic mapping matrices** that are constructed from entity and relation vectors. Instead of a fixed projection matrix  $\mathbf{M}_r$  for each relation, TransD learns two projection vectors for each entity  $(h, t)$  and relation  $(r)$ :  $\mathbf{h}, \mathbf{t}, \mathbf{r} \in \mathbb{R}^k$  and  $\mathbf{h}_p, \mathbf{t}_p, \mathbf{r}_p \in \mathbb{R}^k$ . The projection matrix for a head entity  $\mathbf{M}_{rh}$  is dynamically constructed using  $\mathbf{r}_p$  and  $\mathbf{h}_p$ , and similarly for the tail entity  $\mathbf{M}_{rt}$  using  $\mathbf{r}_p$  and  $\mathbf{t}_p$ . Specifically, the projected entity embeddings are  $\mathbf{h}_r = \mathbf{h} + \mathbf{h}_p \mathbf{r}_p^\top \mathbf{h}$  and  $\mathbf{t}_r = \mathbf{t} + \mathbf{t}_p \mathbf{r}_p^\top \mathbf{t}$ . The scoring function remains translational in the projected space:  $f(h, r, t) = \|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_{L_1/L_2}$ . This allows the projection to be specific to both the relation and\* the entity involved, making it more expressive than TransR while using fewer parameters.* \* **Conditions for Success:** TransD excels in scenarios where entities exhibit diverse roles across different relations, requiring highly context-dependent projections. It offers a good balance between expressiveness and parameter efficiency compared to TransR, making it more robust to sparse relations. \* **Theoretical Limitations:** While more flexible, the dynamic construction of projection matrices is still a linear transformation. It might not capture highly non-linear

or complex interactions between entities and relations. The assumption that projection vectors can effectively capture the "type" or "aspect" of an entity for a given relation is implicit. \* **Practical Limitations:** Although more parameter-efficient than TransR, it still introduces additional vectors per entity and relation. The matrix construction and multiplication operations add computational overhead compared to simpler models like TransE.

### **ManifoldE (Knowledge Graph Embedding for Precise Link Prediction) ?:**

\* **Context and Problem Solved:** ManifoldE challenged the fundamental assumption of point-based entity representations common in TransE, TransH, and TransR. It recognized that representing entities as single points in a vector space might be too restrictive to capture their inherent ambiguity, fuzziness, or multi-faceted nature. The problem it aimed to solve was to allow for more flexible and robust entity representations that could better capture semantic nuances and improve link prediction precision. \* **Core Innovation:** ManifoldE expands entity representations from discrete points to **manifolds** (specifically, spheres or hyperplanes) in the embedding space. For a triple  $(h, r, t)$ , the head entity  $h$  is represented by a manifold  $M_h$ , and the tail entity  $t$  by a manifold  $M_t$ . The relation  $r$  is still a translation vector  $\mathbf{r}$ . The core idea is that for a valid triple, the manifold  $M_h$  translated by  $\mathbf{r}$  should "intersect" or be "close" to the manifold  $M_t$ . The scoring function is defined as the distance between the translated manifold  $M_h + \mathbf{r}$  and  $M_t$ . For example, if entities are represented as spheres, the score could be the distance between the center of the translated head sphere and the center of the tail sphere, adjusted by their radii. This allows for a more "fuzzy" matching, where entities are not forced to align perfectly at a single point. \* **Conditions for Success:** ManifoldE can be particularly effective in KGs where entities have rich, multi-dimensional semantics that are not adequately captured by single points. It offers increased robustness to noise and ambiguity in the data by allowing for a region of plausibility rather than a single exact location. \* **Theoretical Limitations:** Representing entities as manifolds increases the complexity of the model and the number of parameters (e.g., center and radius for a sphere). The choice of manifold type (sphere, hyperplane) is a hyperparameter that needs

careful consideration and might not be universally optimal. The geometric interpretation of "translation of a manifold" can be more complex than point translation. \* **Practical Limitations:** The increased parameter count and the more complex distance calculations involving manifolds lead to higher computational costs and memory requirements compared to point-based models. This can make it less scalable for extremely large KGs.

**Comparative Framework and Evolution:** The development of TransD and ManifoldE illustrates a crucial evolutionary step beyond the initial translational models, driven by the need to overcome their inherent geometric rigidity.

- **TransD** ? directly addresses the parameter efficiency and expressiveness trade-off observed in TransR. By making projection matrices dynamic and entity-aware, it offers a more nuanced way to handle context-dependent entity roles without the prohibitive parameter count of TransR. This is a refinement within the "projection" paradigm, making it more adaptive.
- **ManifoldE** ? represents a more fundamental departure, questioning the very nature of entity representation. By moving from points to manifolds, it introduces a richer, more flexible geometric object for entities, allowing for inherent fuzziness and multi-faceted semantics. This addresses the limitation that a single point might not adequately capture all aspects of an entity, especially in ambiguous or noisy data. This approach implicitly acknowledges the "uncertainty" in knowledge representation, a theme later explored more explicitly by models like ATiSE ? which use Gaussian distributions to model temporal uncertainty.

Both TransD and ManifoldE exemplify the field's early recognition that simple, rigid geometric forms (points and fixed translations) were insufficient for the complexity of real-world KGs. They represent a shift towards more adaptive and flexible representations, laying the groundwork for later models that would explore even more complex geometries (e.g., hyperbolic spaces ??) or entirely different mathematical frameworks. This continuous drive for enhanced representational power, often at the cost of increased complexity, is a defining characteristic of KGE research.

## 2.4 Limitations of Early Geometric Models and the Drive for Enhanced Expressiveness

The early geometric models, particularly the translational family (TransE, TransH, TransR, TransD) and their immediate extensions like ManifoldE, were foundational in establishing the Knowledge Graph Embedding (KGE) paradigm. They demonstrated the feasibility and efficiency of representing symbolic knowledge in continuous vector spaces for tasks like link prediction. However, their inherent design choices and geometric assumptions also exposed significant limitations, which collectively fueled a continuous drive for enhanced expressiveness in subsequent KGE research.

One of the most pervasive limitations of these early models stems from their **geometric rigidity and simplified relational semantics**. TransE, for instance, assumes relations are simple translations, which inherently struggles with complex relational patterns such as 1-N, N-1, N-N, and symmetric/anti-symmetric relations `??`. As discussed, if `(h, r, t1)` and `(h, r, t2)` are both true, TransE forces `t1` and `t2` to be very close, which is often semantically inaccurate. While TransH and TransR introduced projections onto hyperplanes or relation-specific spaces, and TransD offered dynamic mappings, they still relied on linear transformations and distance-based scoring functions. These linear transformations, despite their improvements, may not be sufficient to capture highly non-linear and intricate semantic interactions present in real-world KGs `?`. The assumption of a single, fixed geometric operation (translation, projection) per relation limits their ability to model the diverse and context-dependent nature of relational semantics.

Another critical limitation, often implicitly assumed away by these early models, is the **static nature of knowledge**. Most early KGE models treated KGs as static snapshots, ignoring the temporal validity of facts `??`. A fact like `(Barack Obama, PresidentOf, USA)` is only true for a specific time interval. Ignoring this temporal dimension leads to inaccurate reasoning and an inability to predict future events or understand historical changes. This fundamental oversight became a significant bottleneck, prompting the development of Temporal KGE (TKGE) models. For example, `?'s` HyTE explicitly incorporates time by associating each timestamp with a hyperplane, allowing for temporally



guided inference. Building on this, ?’s ATiSE further addresses this by modeling entity and relation representations as multi-dimensional Gaussian distributions evolving via additive time series decomposition, explicitly capturing *temporal uncertainty*—a nuance that static models completely miss. This highlights a clear evolution: early models implicitly assumed static knowledge, leading to a recognized limitation, which then spurred the explicit modeling of time and even temporal uncertainty.

Furthermore, these models often struggled with **representational capacity versus parameter efficiency**. While TransR and TransD offered increased expressiveness, they did so by introducing more parameters (projection matrices or dynamic mapping vectors). For KGs with a vast number of relations, this can lead to a parameter explosion, making models computationally expensive and prone to overfitting, especially for sparse relations ?. The work by ? on "Entity-Agnostic Representation Learning" directly questions the implicit assumption that each entity requires a unique, directly learned embedding. They demonstrate that conventional KGEs can lead to colossal parameter counts (e.g., 123 million for RotatE on YAGO3-10), hindering deployment. Their solution, EARL, which learns entity representations from distinguishable information rather than direct lookup, highlights a critical practical limitation of these early models’ parameter scaling. This tension between needing more parameters for expressiveness and the practical constraints of memory and computation is a recurring theme in KGE research.

The reliance on **point-based entity representations** was also a limitation. ManifoldE ? was an early attempt to address this by representing entities as manifolds (e.g., spheres), allowing for a more flexible and "fuzzy" representation that could better capture ambiguity or multi-faceted semantics. However, even ManifoldE’s geometric forms were somewhat constrained. The strictness of representing entities as single points or simple geometric shapes often failed to capture the rich, multi-aspect nature of real-world entities, where different relations might highlight different facets of the same entity. This inability to capture disentangled representations was a significant gap, later addressed by models like DisenKGAT ?, which explicitly learns independent components for entities, moving beyond the implicit assumption that a single vector can adequately represent all

aspects of an entity.

Finally, the **handling of noise and uncertainty** in KGs was largely overlooked by these early geometric models. Most assumed that triples were uniformly confident and error-free. However, real-world KGs are often noisy due to automatic extraction processes ?. This led to issues where models could overfit to erroneous facts. The development of confidence-aware negative sampling methods, as proposed by ?, directly addresses this by acknowledging that not all negative samples are equally informative, especially in noisy KGs. This highlights a critical methodological gap in the training procedures of early models.

In summary, the limitations of early geometric models can be categorized along several dimensions:

1. **Geometric Rigidity:** Inability to model complex relational patterns (1-N, N-1, N-N, symmetric) due to simple linear transformations.
2. **Static Knowledge Assumption:** Failure to account for the temporal dynamics and evolution of facts.
3. **Parameter Inefficiency:** Increased expressiveness often came with a prohibitive increase in parameters, leading to scalability and overfitting issues.
4. **Limited Entity Representation:** Representing entities as single points or simple manifolds was insufficient for capturing multi-faceted or ambiguous semantics.
5. **Ignorance of Noise and Uncertainty:** Implicit assumption of perfect, uniformly confident triples.

These limitations collectively underscored the need for KGE models with enhanced expressiveness. The field’s trajectory shifted towards exploring more complex geometric spaces (e.g., hyperbolic embeddings), non-linear transformations (e.g., neural networks), and the integration of auxiliary information, all aimed at developing more robust, flexible, and semantically rich representations for knowledge graphs. This continuous drive for enhanced representational power, often involving trade-offs between simplicity, efficiency, and accuracy, remains a central theme in KGE research.

### 3. Advancing Expressiveness: Rotational, Compound, and Higher-Dimensional Geometries

Building upon the foundational translational models discussed in Section 2, the field of Knowledge Graph Embedding (KGE) has progressively moved towards more sophisticated geometric and algebraic approaches to capture the intricate and diverse relational patterns inherent in real-world knowledge graphs (KGs). The initial simplicity of translation-based models, while efficient, proved insufficient for handling complex relational properties such as symmetry, antisymmetry, composition, and hierarchical structures [??](#). This section explores a significant paradigm shift, detailing the emergence of models that leverage rotations in complex and quaternion spaces, compound geometric operations, and embeddings in non-Euclidean geometries. These advancements collectively aim to overcome the geometric rigidity and limited expressiveness of earlier models, offering a more faithful and nuanced representation of knowledge.

The limitations of translational models, particularly their struggle with 1-N, N-1, N-N relations and symmetric properties (as highlighted in Section 2.2), spurred the development of rotational models. These models, exemplified by RotatE [?](#), conceptualize relations not as simple translations but as rotations in complex vector spaces. This algebraic shift provides an elegant mechanism to inherently model symmetry (a relation is its own inverse), antisymmetry (a relation implies a distinct inverse), and composition (relations can be chained), which were significant challenges for TransE and its direct extensions. The move to complex numbers and quaternions introduces a richer mathematical framework, allowing for more nuanced transformations that can naturally encode these relational characteristics.

Further advancing expressiveness, researchers explored compound geometric operations. Models like CompoundE [?](#) and CompoundE3D [?](#) recognize that a single, monolithic transformation (be it translation or rotation) might be too simplistic for the multifaceted nature of real-world relations. Instead, they propose combining multiple elementary transformations—such as translation, rotation, and scaling—to form a more versatile and adap-

tive relational mapping. This approach acknowledges that different aspects of a relation might require different geometric operations, thereby enabling a more fine-grained modeling of complex interactions. This represents an evolution from the fixed operations of earlier models towards more learnable and composite transformations, directly addressing the geometric rigidity identified in Section 2.3.

A more profound shift in KGE research involves moving beyond the conventional Euclidean embedding space. The implicit assumption that Euclidean space is optimal for all KG structures began to be questioned, particularly for representing hierarchical and complex topological patterns [??](#). Non-Euclidean geometries, such as hyperbolic and hyperspherical spaces, offer intrinsic properties that are better suited for specific data structures. Hyperbolic spaces, with their negative curvature, are inherently adept at modeling hierarchical data and tree-like structures, where distances grow exponentially. Conversely, hyperspherical spaces, with positive curvature, can be beneficial for cyclic or periodic patterns. This exploration culminates in multi-curvature embeddings, such as MADE [?](#) and IME [?](#), which attempt to adapt the local curvature of the embedding space to the specific characteristics of different parts of the knowledge graph. This adaptive geometry represents a significant leap, moving away from a one-size-fits-all embedding space to a more flexible, structure-aware representation.

The increasing complexity of these advanced geometric models, however, often comes with trade-offs. While they offer enhanced representational power, they can introduce greater computational costs, increased parameter counts, and potentially reduced interpretability. The parameter efficiency concerns highlighted by [?](#) for conventional KGEs become even more pertinent here, as complex transformations and higher-dimensional spaces can exacerbate the parameter explosion problem. This section will critically analyze these advanced approaches, comparing their theoretical underpinnings, practical implications, and the specific types of relational patterns they are designed to capture, ultimately illustrating the field’s continuous quest for more expressive and robust knowledge graph embeddings.

### Relational Rotations in Complex and Quaternion Spaces

The limitations of translational models in capturing complex relational patterns, particularly symmetry, antisymmetry, and composition, motivated a significant shift towards representing relations as rotations in higher-dimensional algebraic spaces. This paradigm leverages the inherent properties of complex numbers and quaternions to naturally encode these relational characteristics, offering a more expressive alternative to simple vector addition.

**RotatE (Knowledge Graph Embedding by Relational Rotation in Complex Space) ?:**

- \* **Context and Problem Solved:** RotatE was introduced to address the fundamental inability of translational models (like TransE, TransH, TransR) to model symmetric, antisymmetric, and compositional relations effectively. As discussed in Section 2.2, TransE struggles with symmetric relations by forcing the relation vector to be zero, and with compositional relations like  $(a, r_1, b)$  and  $(b, r_2, c)$  implying  $(a, r_1+r_2, c)$ . RotatE sought to provide a unified and elegant solution for these properties.
- \* **Core Innovation:** RotatE represents entities  $\mathbf{h}, \mathbf{t}$  as vectors in a complex vector space  $\mathbb{C}^k$  and relations  $\mathbf{r}$  as diagonal matrices where each diagonal element is a complex number with modulus 1 (i.e., a rotation). For a triple  $(h, r, t)$ , the scoring function is defined as  $f(h, r, t) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|_{L_1/L_2}$ , where  $\circ$  denotes the Hadamard (element-wise) product. The relation  $\mathbf{r}$  acts as an element-wise rotation from  $\mathbf{h}$  to  $\mathbf{t}$  in the complex plane.
- \* **Symmetry:** If  $r$  is symmetric, then  $\mathbf{r}$  should be close to a vector of ones (no rotation), meaning  $\mathbf{h} \approx \mathbf{t}$ .
- \* **Antisymmetry:** If  $r$  is antisymmetric, then  $\mathbf{r}$  should be far from a vector of ones, ensuring  $\mathbf{h} \circ \mathbf{r} \neq \mathbf{t}$  implies  $\mathbf{t} \circ \mathbf{r} \neq \mathbf{h}$ .
- \* **Composition:** For  $\mathbf{r}_1$  and  $\mathbf{r}_2$  such that  $\mathbf{r}_1 + \mathbf{r}_2$  is a composite relation, RotatE naturally models this as  $\mathbf{r}_1 \circ \mathbf{r}_2$ , meaning sequential rotations compose correctly. This directly addresses a major limitation of TransE’s additive composition.
- \* **Conditions for Success:** RotatE demonstrates strong performance on datasets rich in symmetric, antisymmetric, and compositional relations. Its elegant mathematical formulation provides a powerful inductive bias for these relational patterns.
- \* **Theoretical Limitations:** While powerful for rotation-based patterns, RotatE’s fixed rotation mechanism might still struggle with relations that are better modeled by translations or more complex non-linear transformations. The

assumption that all relations can be adequately represented as element-wise rotations in complex space might be overly simplistic for extremely diverse relation types. \* **Practical Limitations:** Complex number arithmetic can be slightly more computationally intensive than real vector operations. The parameter count is similar to TransE (two vectors per entity, one per relation), but the complex nature adds a factor of two in storage. ? notes that RotatE, among other models, sometimes failed to complete trials on larger datasets, suggesting potential computational demands or implementation challenges, which can be a practical bottleneck for very large KGs.

**Quaternion Embeddings (QuatE, HouseE, TorusE, SpherE):** Building on the idea of complex rotations, quaternion embeddings extend this concept to a 4-dimensional hypercomplex space, offering even richer rotational capabilities. Quaternions can represent 3D rotations more naturally and without gimbal lock issues, making them appealing for KGE. \* **QuatE (Quaternion Knowledge Graph Embedding) ?:** \* **Context and Problem Solved:** QuatE extends RotatE by using quaternions, which can capture more intricate relational patterns than complex numbers. It aims to model relations as rotations in 3D space, providing a more expressive representation for entities and relations. \* **Core Innovation:** Entities and relations are embedded as quaternion vectors. For a triple  $(h, r, t)$ , the scoring function is typically based on quaternion multiplication, e.g.,  $f(h, r, t) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|_{L_1/L_2}$  or similar, where  $\circ$  is quaternion multiplication. This allows for a more general form of rotation. \* **Conditions for Success:** QuatE often outperforms RotatE on benchmarks, demonstrating improved expressiveness, particularly for relations that benefit from richer rotational transformations. \* **Theoretical Limitations:** While more expressive, the interpretability of quaternion operations can be challenging. The increased dimensionality (4x real numbers per element) adds to parameter count and computational complexity. \* **HouseE (Knowledge Graph Embedding with Householder Parameterization) ?:** \* **Context and Problem Solved:** HouseE addresses the challenge of designing expressive and parameter-efficient rotational models. It leverages Householder transformations, which are reflections, to model relations. \* **Core Innovation:** Relations are parameterized using Householder matrices, which

are orthogonal and symmetric. This allows for modeling relations as reflections, which can be seen as a specific type of rotation. The scoring function involves applying this Householder transformation to the head entity embedding to approximate the tail entity.

\* **Comparison:** HousE offers a novel way to achieve rotational expressiveness with a potentially more parameter-efficient representation than full quaternion embeddings, as Householder matrices can be defined by a single vector. This directly addresses the parameter efficiency concerns raised by ?. \* **TorusE (Knowledge Graph Embedding on a Lie Group) ?:** \* **Context and Problem Solved:** TorusE explores embedding KGs on a Lie group, specifically a torus, which is a compact manifold. This aims to capture cyclic or periodic relational patterns that might not be well-represented in Euclidean or even complex spaces. \* **Core Innovation:** Embeds entities and relations as points on a torus, where relations are modeled as translations along the torus. The periodic nature of the torus naturally handles cyclic relations. \* **Comparison:** TorusE offers a unique geometric space that is distinct from complex or quaternion rotations but shares the goal of capturing specific algebraic properties. It provides an alternative to RotatE for certain types of cyclic relations, but its applicability might be more niche. \* **SpherE (Expressive and Interpretable Knowledge Graph Embedding for Set Retrieval) ?:** \* **Context and Problem Solved:** SpherE focuses on embedding knowledge graphs on a hypersphere, aiming for expressive and interpretable representations, particularly for set retrieval tasks. \* **Core Innovation:** Entities and relations are embedded on a unit hypersphere. Relations are modeled as rotations or transformations within this spherical space. The spherical geometry is particularly adept at capturing similarity based on angles, which can be useful for certain semantic relationships. \* **Comparison:** SpherE shares the rotational intuition with RotatE and QuatE but situates it within a spherical manifold. This offers a different inductive bias, potentially better for relations where angular distance is more meaningful than Euclidean distance.

**Comparative Framework and Evolution:** The evolution from translational to rotational models, and further to quaternion and other manifold-based embeddings, represents a clear trajectory towards increased algebraic and geometric expressiveness.

- **TransE** (Section 2.2) provides a simple additive model, failing for complex relations.
- **RotatE** ? introduces complex numbers, enabling inherent modeling of symmetry, antisymmetry, and composition through element-wise rotations. This is a direct algebraic enhancement that resolves many of TransE’s relational limitations. The work by ? (Rot-Pro) further explores transitivity modeling through projection in a similar vein.
- **QuatE** ? extends RotatE to quaternions, offering a richer 4D rotational space, potentially capturing more complex transformations. This represents a further step in leveraging higher-dimensional algebra for expressiveness.
- **HouseE** ? and **TorusE** ? and **SpherE** ? explore alternative mathematical frameworks (Householder reflections, Lie groups, spherical geometry) to achieve similar goals of enhanced rotational or geometrically constrained expressiveness. HouseE, in particular, highlights a tension between expressiveness and parameter efficiency, offering a potentially more compact way to represent transformations than full quaternion embeddings.

This family of models demonstrates a recurring pattern: leveraging advanced mathematical structures (complex numbers, quaternions, Lie groups) to imbue relations with richer algebraic properties. While these models significantly enhance the ability to capture specific relational patterns, they also introduce increased mathematical complexity and potentially higher computational demands. The choice among them often depends on the specific relational properties dominant in a given KG and the trade-off between expressiveness, efficiency, and interpretability.

### ### Compound Geometric Transformations

While rotational models significantly advanced the expressiveness of KGEs by introducing complex and quaternion algebra, they still largely relied on a single, albeit more sophisticated, geometric operation per relation. Recognizing that real-world relations often involve a confluence of different semantic transformations, a new class of models emerged that employs **compound geometric operations**, combining elements



like translation, rotation, and scaling. This approach aims to provide a more flexible and adaptive framework for modeling intricate relational patterns, moving beyond the monolithic transformations of earlier models.

**CompoundE (Knowledge Graph Embedding with Translation, Rotation and Scaling Compound Operations) ?:** \* **Context and Problem Solved:** CompoundE addresses the limitation that many relations cannot be adequately captured by a single type of geometric transformation (e.g., pure translation or pure rotation). For instance, a "part-of" relation might involve both a translation (moving from a whole to a part) and a scaling (the part is smaller than the whole). Previous models lacked the flexibility to combine these fundamental operations. \* **Core Innovation:** CompoundE represents each relation  $r$  as a combination of three distinct geometric operations: a **translation vector**  $\mathbf{t}_r$ , a **rotation matrix**  $\mathbf{R}_r$ , and a **scaling vector**  $\mathbf{s}_r$ . For a triple  $(h, r, t)$ , the scoring function is based on applying these compound operations to the head entity embedding:  $f(h, r, t) = \|\mathbf{s}_r \circ (\mathbf{R}_r \mathbf{h}) + \mathbf{t}_r - \mathbf{t}\|_{L_1/L_2}$ . This allows relations to simultaneously translate, rotate, and scale entity embeddings, providing a much richer set of transformations. The rotation matrix  $\mathbf{R}_r$  can be parameterized efficiently (e.g., using Euler angles or Rodrigues' rotation formula) to avoid excessive parameters. \* **Conditions for Success:** CompoundE is particularly effective on KGs where relations exhibit diverse semantic properties that require a blend of transformations, such as hierarchical relations (scaling), part-whole relations (translation + scaling), or abstract conceptual shifts (rotation). It offers superior performance over models relying on single operations by adapting to the multifaceted nature of relations. \* **Theoretical Limitations:** The interpretability of a relation as a combination of translation, rotation, and scaling can be complex. While more expressive, the model still operates within a Euclidean space, which may not be optimal for highly hierarchical or non-Euclidean structures. The parameterization of rotation matrices and the interaction between the three operations can be challenging to optimize. \* **Practical Limitations:** The introduction of three distinct sets of parameters per relation (translation vector, rotation matrix parameters, scaling vector) significantly increases the total parameter count compared to simpler models like

TransE or RotatE. This can lead to higher memory consumption and computational cost during training, especially for KGs with many relations. This exacerbates the parameter efficiency concerns raised by ?, where even single-operation models like RotatE already face parameter explosion on large KGs.

**CompoundE3D (Knowledge Graph Embedding with 3D Compound Geometric Transformations) ?:** \* **Context and Problem Solved:** CompoundE3D is an extension of CompoundE, specifically designed to operate in a 3D embedding space, aiming to further enhance the expressiveness of compound operations by leveraging the geometric properties of 3D rotations and transformations more directly. \* **Core Innovation:** Similar to CompoundE, it combines translation, rotation, and scaling, but explicitly within a 3D vector space. This might involve using quaternions for more stable and efficient 3D rotation representation, rather than less stable Euler angles or more parameter-heavy rotation matrices. The scoring function adapts to this 3D context. \* **Comparison:** CompoundE3D builds directly on CompoundE’s principle, focusing on optimizing the compound operations for a 3D space. This suggests an iterative refinement of the compound operation idea, seeking to improve stability and efficiency while maintaining expressiveness. Its performance is expected to be competitive with or superior to CompoundE, especially where 3D geometric intuition aligns well with relational semantics.

**STaR (Knowledge Graph Embedding by Scaling, Translation and Rotation) ?:** \* **Context and Problem Solved:** STaR also proposes a compound operation model, similar to CompoundE, but potentially with different parameterizations or optimization strategies. It aims to provide a robust framework for combining the three fundamental geometric transformations. \* **Core Innovation:** STaR models relations as a sequence of scaling, translation, and rotation operations. The specific order or parameterization might differ from CompoundE, but the core idea of composite transformations remains. \* **Comparison:** STaR is another manifestation of the compound operation paradigm. The existence of multiple models (CompoundE, CompoundE3D, STaR) employing similar compound operations highlights a convergent research direction, recognizing the need for

more flexible relational transformations. The differences likely lie in the specific mathematical parameterizations of these operations (e.g., how rotations are represented, how scaling is applied) and their optimization strategies.

**Comparative Framework and Evolution:** The emergence of compound geometric transformation models marks a significant evolution from both translational and purely rotational models.

- **Translational Models (e.g., TransE, TransH, TransR)** (Section 2.2) use a single, linear operation (translation, possibly with projection). They are simple but geometrically rigid.
- **Rotational Models (e.g., RotatE, QuatE)** (Section 3.1) introduce complex or quaternion algebra for rotations, addressing symmetry and composition, but still rely on a single type of transformation.
- **Compound Models (e.g., CompoundE ?, CompoundE3D ?, STaR ?)** represent a paradigm shift by combining multiple fundamental geometric operations (translation, rotation, scaling) into a single relational transformation. This directly addresses the limitation that real-world relations are often multifaceted and cannot be adequately captured by a singular geometric action. This approach acknowledges that relations are not just "one thing" but a composite of several semantic shifts.

A critical tension exists between the enhanced expressiveness of these compound models and their increased complexity and parameter count. While they achieve higher accuracy by better modeling complex relations, they are inherently more demanding computationally. This trade-off is a recurring challenge in KGE research, where the pursuit of higher representational power often necessitates more intricate models. The success of these models suggests that the field is moving towards a more nuanced understanding of relational semantics, where relations are viewed as dynamic, multi-component transformations rather than static, singular operations.

### Exploring Non-Euclidean and Multi-Curvature Embedding Spaces

A profound advancement in KGE research involves questioning the fundamental assumption that Euclidean space is the optimal manifold for embedding knowledge graphs. The recognition that real-world KGs often exhibit inherent hierarchical structures, power-law distributions, and complex topological patterns has led to the exploration of non-Euclidean geometries, particularly hyperbolic and hyperspherical spaces, and subsequently, multi-curvature embeddings. This shift represents a paradigm change, moving from a universally flat space to spaces whose intrinsic curvature better aligns with the underlying structure of knowledge.

**Hyperbolic Embedding Spaces:** \* **Context and Problem Solved:** Euclidean space struggles to efficiently embed hierarchical data, where the number of nodes grows exponentially with depth. Representing such structures in Euclidean space requires exponentially increasing embedding dimensions or leads to high distortion. Hyperbolic spaces, with their negative curvature, naturally expand exponentially, making them inherently well-suited for embedding tree-like and hierarchical structures with low distortion ???. This addresses a critical limitation of all previous Euclidean-based models (translational, rotational, compound) which implicitly assume a flat geometry. \* **Core Innovation:** Entities and relations are embedded in a hyperbolic manifold (e.g., Poincaré disk or hyperboloid model). Relations are typically modeled as transformations (e.g., Möbius transformations for translations/rotations) within this non-Euclidean space. The scoring function uses hyperbolic distance, which naturally captures hierarchical relationships: entities far apart in the hierarchy are exponentially distant in hyperbolic space, while entities close in the hierarchy are close in hyperbolic space. \* **Hyperbolic Hierarchy-Aware KGE for Link Prediction ?:** This work explicitly leverages hyperbolic geometry to capture hierarchical information within KGs for link prediction. It demonstrates how the intrinsic properties of hyperbolic space can better preserve the structural hierarchy of entities. \* **Fully Hyperbolic Rotation for KGE ?:** This model combines the benefits of hyperbolic geometry with rotational transformations, allowing for expressive modeling of relations (like symmetry/antisymmetry) within a hierarchy-aware space. This represents a convergence of rotational models (Section 3.1) with non-Euclidean geometry. \* **Conditions for**

**Success:** Hyperbolic embeddings excel on KGs that possess strong hierarchical structures, such as ontologies, taxonomies, or biological networks. They achieve significantly lower distortion and higher accuracy in link prediction and entity classification tasks on such datasets compared to Euclidean counterparts. \* **Theoretical Limitations:** While ideal for hierarchies, hyperbolic spaces might not be optimal for all types of relational patterns, particularly those that are not inherently tree-like or exhibit cyclic structures. The mathematical operations (e.g., addition, distance calculation) in hyperbolic space are more complex than in Euclidean space, requiring specialized algorithms. \* **Practical Limitations:** Implementing and optimizing models in hyperbolic space is more challenging due to the non-linear nature of the geometry and the need for specialized Riemannian optimization techniques. This can lead to higher computational costs and slower training times.

**Hyperspherical Embedding Spaces:** \* **Context and Problem Solved:** While hyperbolic spaces address hierarchies, some KGs or parts of KGs might exhibit cyclic, periodic, or clustered patterns that are better captured by positive curvature. Hyperspherical spaces (e.g., unit spheres) offer such a geometry. \* **Core Innovation:** Entities and relations are embedded on the surface of a hypersphere. Relations are modeled as rotations or transformations on this sphere. The distance metric is typically geodesic distance (arc length). \* **SpherE ?** (also mentioned in 3.1) is a prime example, leveraging spherical geometry for expressive and interpretable embeddings, particularly for set retrieval. It highlights how the angular relationships on a sphere can capture certain semantic similarities. \* **Conditions for Success:** Hyperspherical embeddings are beneficial for KGs where relations exhibit periodicity, or where entities naturally form clusters on a spherical manifold. \* **Theoretical Limitations:** Similar to hyperbolic spaces, hyperspherical spaces are not universally optimal. They might struggle with deep hierarchies or relations that require significant expansion. \* **Practical Limitations:** Optimization on spheres also requires specialized techniques, adding to computational complexity.

**Multi-Curvature Adaptive Embedding Spaces (MADE, IME):** \* **Context and Problem Solved:** The realization that different parts of a knowledge graph might

inherently possess different structural properties (e.g., some parts are hierarchical, others are more grid-like, some are cyclic) led to the idea of adaptive or multi-curvature embedding spaces. The limitation of a single, global geometry (Euclidean, hyperbolic, or spherical) is that it cannot optimally represent the heterogeneous nature of real-world KGs.

\* **Core Innovation:** These models aim to adapt the local curvature of the embedding space to the specific characteristics of entities and relations. \* **MADE (Multicurvature Adaptive Embedding for Temporal Knowledge Graph Completion) ?:** MADE introduces a mechanism to dynamically determine the optimal curvature for different entities and relations, or even different parts of the embedding space. This allows the model to use hyperbolic geometry for hierarchical components and Euclidean geometry for more grid-like structures, or spherical for cyclic ones, all within a unified framework. It applies this to temporal KGs, further increasing complexity. \* **IME (Integrating Multi-curvature Shared and Specific Embedding for Temporal Knowledge Graph Completion) ?:** IME builds on the multi-curvature concept by integrating both shared (global) and specific (local) embeddings across different curvatures. This allows for a more nuanced representation where some aspects of an entity might benefit from one curvature, while others from another. It also applies this to temporal KGs. \* **Conditions for Success:** Multi-curvature models are particularly powerful for large, heterogeneous KGs that exhibit diverse structural patterns. They offer the potential for higher accuracy by reducing distortion across the entire graph. \* **Theoretical Limitations:** The theoretical justification for dynamically assigning or learning optimal local curvatures is complex. The interaction between different curvatures and how transformations behave across these varying geometries is an active research area. The interpretability of embeddings in such mixed-geometry spaces can be significantly reduced. \* **Practical Limitations:** These models are significantly more complex computationally. Learning and optimizing parameters for multiple curvatures, potentially dynamically, adds substantial overhead. This directly challenges the parameter efficiency goals, as highlighted by ?, by introducing more intricate mathematical structures and optimization problems.

**Comparative Framework and Evolution:** The journey from Euclidean to non-

Euclidean and then to multi-curvature spaces represents the most advanced frontier in geometric KGE, driven by the need to match the embedding space’s intrinsic properties to the KG’s structural characteristics.

- **Euclidean Models (Translational, Rotational, Compound)** (Section 2.2, 3.1, 3.2) implicitly assume a flat, uniform space. They are computationally simpler but suffer from distortion when embedding non-Euclidean structures like hierarchies.
- **Hyperbolic KGEs** ?? explicitly address the challenge of embedding hierarchical data by leveraging negative curvature, offering superior performance for such structures. This is a direct response to the geometric limitations of Euclidean space for specific KG topologies.
- **Hyperspherical KGEs** ? provide an alternative for cyclic or clustered patterns, demonstrating that different non-Euclidean geometries serve different structural biases.
- **Multi-Curvature Models (MADE ?, IME ?)** represent the cutting edge, attempting to overcome the "one-size-fits-all" limitation of a single global geometry. They aim to adapt the embedding space’s curvature locally, offering the most flexible and structure-aware representations. This approach directly challenges the implicit assumption that a single, global geometric space is sufficient for all KGs, recognizing their inherent heterogeneity.

This evolution highlights a fundamental tension in KGE research: the trade-off between geometric simplicity and computational efficiency versus the fidelity of representation. While non-Euclidean and multi-curvature embeddings offer unparalleled expressiveness for complex KG structures, they introduce significant mathematical and computational challenges. The field is actively exploring how to balance these factors, with a clear trend towards more adaptive and structure-aware embedding spaces to achieve a more faithful representation of diverse knowledge graph characteristics. The unresolved debate lies in determining the optimal strategy for learning and integrating these diverse geometries efficiently and interpretably.

## 4. Deep Learning Architectures and Automated Design for KGE

The landscape of Knowledge Graph Embedding (KGE) has undergone a profound transformation with the integration of advanced deep learning architectures, marking a significant paradigm shift from the purely geometric and algebraic models discussed in Sections 2 and 3. While translational, rotational, and compound geometric models offered increasing expressiveness for specific relational patterns ???, they often relied on pre-defined scoring functions and lacked the inherent capacity for data-driven feature extraction and complex, non-linear interactions crucial for capturing the multifaceted semantics of real-world knowledge graphs (KGs) ?. This section delves into how deep learning, particularly Convolutional Neural Networks (CNNs), Graph Neural Networks (GNNs), and Transformer-based models, has revolutionized KGE by enabling sophisticated local and global feature learning, message passing, and contextualized representations. Furthermore, it explores the cutting-edge area of automated design and meta-learning, which aims to alleviate the burden of manual architecture engineering by discovering optimal KGE components adaptively.

The transition to deep learning architectures was motivated by several key limitations of earlier KGE models. Firstly, geometric models, despite their elegance, often struggled with the sheer diversity and complexity of relational patterns, especially those involving intricate logical rules or multi-hop dependencies ?. Their fixed scoring functions, whether based on distances or rotations, could not adapt flexibly to the varying semantic contexts of different entities and relations. Secondly, the expressiveness of these models often came at the cost of increased parameter counts, as highlighted by ?, which becomes a significant practical limitation for large-scale KGs. Deep learning offers mechanisms like parameter sharing and compositional representations that can mitigate this to some extent while enhancing model capacity. Thirdly, the interpretability of embeddings, while sometimes clearer in simple geometric models, became opaque as models grew more complex. Deep learning, particularly with attention mechanisms, offers new avenues for explainability ??.



Convolutional Neural Networks (CNNs) were among the first deep learning architectures to be applied to KGE, primarily for their ability to extract local features and detect interaction patterns within the embedding space. By treating embeddings as "images," CNNs could identify intricate patterns that might be missed by simpler dot products or distance functions. This marked an initial step towards feature learning beyond direct vector operations. Building on this, Graph Neural Networks (GNNs) emerged as a more natural fit for the graph-structured nature of KGs. GNNs, through their message-passing and aggregation mechanisms, inherently capture complex structural patterns by iteratively propagating and transforming information across neighboring entities and relations. This allows for the learning of context-aware representations that reflect the local graph topology.

The advent of Transformer-based models, initially successful in natural language processing, further pushed the boundaries of KGE. These models, exemplified by architectures like TGformer [10] and Contextualized KGE (CKGE) [11], excel at modeling long-range dependencies and contextualized features through self-attention mechanisms. By treating the KG structure or paths within it as sequences, Transformers can capture multi-structural and global contextual information, offering a powerful way to learn highly expressive and adaptive embeddings. This represents a move towards more holistic graph understanding, where the entire context of a triple or a path contributes to its embedding.

Finally, as KGE models grew in complexity, the manual design of optimal scoring functions, aggregation mechanisms, or network architectures became increasingly challenging and time-consuming. This led to the development of meta-learning and automated design approaches, such as AutoSF [12] and Message Function Search [13]. These methods leverage neural architecture search (NAS) principles to automatically discover efficient and effective KGE components, reducing human effort and enhancing adaptability to diverse KG characteristics. This frontier aims to democratize the development of high-performing KGE models, making them less reliant on expert knowledge and extensive hyperparameter tuning [14].

Collectively, these deep learning architectures and automated design paradigms repre-

sent a concerted effort to enhance the expressiveness, adaptability, and scalability of KGE. They move beyond fixed geometric assumptions to data-driven learning, enabling models to capture increasingly subtle and complex relational semantics, while simultaneously addressing the practical challenges of model design and deployment.

#### ### 4.1. Convolutional and Graph Neural Networks for KGE

The shift towards deep learning in KGE began with the adoption of Convolutional Neural Networks (CNNs) and subsequently evolved into the more graph-native Graph Neural Networks (GNNs). These architectures brought the power of data-driven feature extraction and complex interaction modeling to the task of learning knowledge graph embeddings, addressing limitations of earlier geometric models that relied on simpler, fixed scoring functions.

**Convolutional Neural Networks (CNNs) for KGE:**

- \* **Context and Problem Solved:** Early KGE models, particularly translational and bilinear models (as discussed in Section 2), often used simple scoring functions (e.g., dot product, L1/L2 distance) that struggled to capture intricate interaction patterns between entity and relation embeddings. CNNs were introduced to address this by treating the concatenated or outer-product representations of triples as multi-channel "images," allowing for the extraction of rich, local interaction features. This was a significant departure from models that relied on a single, global interaction.
- \* **Core Innovation:** CNN-based KGE models typically concatenate or reshape the embeddings of a head entity ( $\mathbf{h}$ ), relation ( $\mathbf{r}$ ), and tail entity ( $\mathbf{t}$ ) into a 2D matrix. Convolutional filters then slide over this matrix to detect local patterns and interactions. For example, ConvE concatenates  $\mathbf{h}$  and  $\mathbf{r}$  into a 2D matrix, applies 2D convolutions, and then projects the result to match the dimensionality of  $\mathbf{t}$  for scoring. Other variants, such as ConvKB, apply convolutions over the concatenated  $(\mathbf{h}, \mathbf{r}, \mathbf{t})$  vector. The core idea is that different parts of the embeddings interact in complex ways, and CNNs can learn these local interaction features.
- \* **Proposed Model:** proposed a KGE model with **Atrous Convolution and Residual Learning**, aiming to capture multi-scale features and alleviate vanishing gradients. This innovation allows for a larger receptive field without increasing parameters, enhancing the ability to capture broader

interaction patterns. \* ? introduced **ReInceptionE**, a relation-aware Inception Network that leverages joint local-global structural information. This model uses inception-like modules to extract features at different scales, combining local interaction patterns with broader structural context, thereby improving the expressiveness of CNN-based KGEs. \* ? recently proposed a **Convolutional Neural Network-Based Entity-Specific Common Feature Aggregation** method. This approach focuses on aggregating common features specific to entities using CNNs, suggesting a move towards more fine-grained, entity-centric feature learning within the CNN framework. \* **Conditions for Success:** CNN-based models generally perform well on datasets where local interactions between embedding dimensions are highly informative. They are particularly effective at capturing complex, non-linear relationships that are difficult for simpler models to model. Their ability to learn hierarchical features makes them robust to variations in embedding representations. \* **Theoretical Limitations:** While CNNs excel at local feature extraction, they are inherently designed for grid-like data (like images). Applying them to linear concatenations of embeddings might not fully leverage the graph structure of KGs. They often struggle to capture long-range dependencies or global structural information without very deep architectures or large receptive fields, which can be computationally expensive. The "locality" assumption of convolutions might not always align with the non-local nature of some KG relations. \* **Practical Limitations:** The parameter count can still be substantial due to the filters and subsequent fully connected layers. Training can be computationally intensive, especially with deeper CNNs. The choice of filter sizes, number of channels, and padding requires careful hyperparameter tuning, which ? highlights as a general challenge in KGE.

**Graph Neural Networks (GNNs) for KGE:** \* **Context and Problem Solved:** Recognizing that KGs are intrinsically graph-structured data, GNNs emerged as a more principled and powerful approach than CNNs. GNNs are designed to directly operate on graphs, allowing them to capture complex structural patterns, multi-hop relationships, and contextual information through iterative message passing and aggregation mechanisms ???. This directly addresses the limitation of CNNs that treat graph data as a

flattened sequence, losing structural information. \* **Core Innovation:** GNNs learn node representations by aggregating information from their neighbors. For KGE, this typically involves a message-passing scheme where each entity (node) gathers messages from its neighbors, transforms them based on the connecting relations, and then aggregates these messages to update its own embedding. \* **R-GCN (Relational Graph Convolutional Networks) ?:** One of the earliest and most influential GNNs for KGE. It extends standard GCNs to handle multi-relational graphs by using relation-specific weight matrices for message passing. This allows each relation to define its own transformation function, capturing diverse relational semantics. \* **CompGCN (Compositional Graph Convolutional Networks) ?:** This model combines the compositional properties of KGE models (like ComplEx or DistMult) with GCNs. It learns entity and relation embeddings simultaneously, where relation embeddings are used to compose messages during aggregation, leading to more expressive representations. \* ***DisenKGAT (Knowledge Graph Embedding with Disentangled Graph Attention Network) ?:** This work introduces disentangled representation learning into GNNs for KGE. It addresses the problem that entities often have multiple facets, and relations focus on distinct aspects. DisenKGAT uses a relation-aware aggregation mechanism for micro-disentanglement (identifying relevant neighbors for each component) and mutual information regularization for macro-disentanglement\* (ensuring independence between components). This allows for adaptive, robust, and interpretable entity embeddings by explicitly modeling the entanglement of latent factors, a significant advancement over static GNN embeddings.* \* **Logic Attention Based Neighborhood Aggregation ?:** This approach incorporates logical rules into GNN aggregation, allowing the model to prioritize messages based on their logical relevance, enhancing inductive capabilities. \* **Graph Attenuated Attention Networks ?:** This model uses attention mechanisms within GNNs to dynamically weigh the importance of neighboring entities and relations during message passing. This allows the model to focus on the most relevant contextual information, improving representation quality. \* **Multiview Feature Augmented Neural Network ?** and **Multisource Hierarchical Neural Network ?:** These works highlight the trend of integrating di-

verse features and hierarchical structures within GNNs, further enhancing their capacity to capture complex KG semantics. \* **Decoupled Semantic Graph Neural Network**?: This recent work proposes decoupling semantic information within GNNs for KGE, suggesting a continued effort to refine how GNNs process and represent complex relational semantics. \* **Conditions for Success**: GNNs are highly effective on KGs with rich structural information and where multi-hop reasoning is important. They excel at learning context-aware embeddings by propagating information across the graph. Models like DisenKGAT ? thrive when entities exhibit multi-faceted meanings that need to be disentangled. \* **Theoretical Limitations**: GNNs can suffer from over-smoothing, where representations of distant nodes become indistinguishable after many layers. They also face challenges with scalability to very large KGs due to the computational cost of message passing across all nodes. The assumption that local neighborhood information is sufficient might not hold for all global reasoning tasks. The effectiveness of disentanglement in DisenKGAT ? depends on the assumption that entities can be effectively decomposed into a predefined number of independent components, and the MI regularization needs to accurately enforce this independence. \* **Practical Limitations**: Training GNNs on large KGs can be memory-intensive, requiring techniques like mini-batching or sampling. The choice of aggregation function, number of layers, and attention mechanisms adds to hyperparameter complexity. The interpretability of the learned disentangled components in models like DisenKGAT, while improved, can still be challenging to fully understand.

**Comparative Framework and Evolution**: The evolution from CNNs to GNNs for KGE reflects a deeper understanding of the intrinsic nature of knowledge graphs.

- **CNNs (e.g., ConvE, ReInceptionE ?, Atrous Conv ?)** represent an initial foray into deep learning for KGE. They excel at extracting local interaction patterns from flattened embedding representations. Their strength lies in learning intricate feature maps, but they are inherently less "graph-aware" and struggle with explicitly modeling structural dependencies beyond local neighborhoods. They are a step beyond purely geometric models by introducing data-driven feature learning.
- **GNNs (e.g., R-GCN, CompGCN, DisenKGAT ?)** are a more natural fit

for KGs, directly leveraging the graph structure through message passing. They capture multi-hop dependencies and contextual information more effectively than CNNs. The development of GNNs for KGE signifies a paradigm shift towards truly graph-native deep learning, where the graph topology is central to representation learning.

A critical tension exists between the expressiveness of complex GNNs (like DisenKGAT, which offers disentangled representations) and their computational demands and interpretability. While DisenKGAT ? significantly advances the state-of-the-art by providing adaptive and interpretable embeddings, it introduces additional complexity in its aggregation and regularization mechanisms. This mirrors the recurring trade-off between model capacity and efficiency observed in earlier KGE models. The field continues to grapple with how to design GNNs that are both highly expressive and scalable to the ever-growing size and complexity of real-world knowledge graphs.

#### ### 4.2. Transformer-based KGE Models

Building upon the success of CNNs and GNNs in capturing local and structural features, the advent of Transformer-based models has introduced a new level of sophistication to Knowledge Graph Embedding (KGE). Originating from natural language processing, Transformers excel at modeling long-range dependencies and learning highly contextualized representations through self-attention mechanisms. Their application to KGE marks a significant step towards capturing global, multi-structural, and context-aware features, moving beyond the limitations of purely local or fixed-neighborhood aggregation.

**Context and Problem Solved:** While GNNs effectively aggregate information from local neighborhoods, they can struggle with capturing very long-range dependencies or global contextual information without resorting to many layers, which can lead to over-smoothing or increased computational cost. Furthermore, the sequential nature of attention in Transformers allows for a more flexible modeling of entity and relation interactions, moving beyond predefined graph traversal or fixed message-passing schemes. Transformer-based models aim to provide highly contextualized embeddings that can adapt to the specific roles of entities and relations within complex, multi-hop paths or subgraphs ?? . This

addresses the challenge of capturing the full "story" or context around a triple, which is often crucial for accurate link prediction and reasoning.

**Core Innovation:** Transformer-based KGE models adapt the self-attention mechanism to the graph structure. Instead of processing a linear sequence of words, they process sequences of entities, relations, or paths extracted from the KG. The attention mechanism allows each element in the sequence to weigh the importance of all other elements, thereby capturing global dependencies.

- \* **Contextualized Knowledge Graph Embedding (CoKE) ?**: CoKE was an early attempt to introduce contextualized embeddings, though not strictly a Transformer. It used a recurrent neural network (RNN) to encode paths, demonstrating the value of sequential context. This paved the way for more powerful attention-based models.
- \* **Position-Aware Relational Transformer (P-ART) ?**: P-ART explicitly adapts the Transformer architecture for KGE by incorporating position-aware mechanisms. It recognizes that the position of an entity or relation within a path or subgraph sequence is crucial for its meaning. By integrating relative positional encodings and relational information into the self-attention layers, P-ART can effectively model the order and type of interactions along a path, leading to more accurate contextualized embeddings. This is a direct extension of the Transformer's strength in sequence modeling to the relational paths in KGs.
- \* **TGformer (A Graph Transformer Framework for Knowledge Graph Embedding) ?**: TGformer represents a dedicated Graph Transformer framework for KGE. Its core innovation lies in its ability to model contextualized and multi-structural features by adapting the Transformer's self-attention to the graph topology. It likely employs sophisticated graph-aware positional encodings and attention masks to ensure that the attention mechanism respects the underlying graph structure, rather than treating the input as a purely linear sequence. This allows TGformer to capture both local and global dependencies, and potentially different types of structural patterns (e.g., paths, cycles) simultaneously.

**Contextualized Knowledge Graph Embedding for Explainable Talent Training Course Recommendation (CKGE) ?**: *This model applies a novel KG-based Transformer to a specific application domain. Its innovation includes constructing a meta-graph for each talent-course pair,*

*serializing entities and paths from this meta-graph into a sequential input for the Transformer. It incorporates specialized relational attention and structural encoding to model global dependencies within the KG. Crucially, it introduces a local path mask prediction\** mechanism to reveal the saliency of meta-paths, directly contributing to the explainability of recommendations. This highlights how Transformers can be adapted not just for accuracy but also for interpretability in complex reasoning tasks. \* **TracKGE (Transformer with Relation-pattern Adaptive Contrastive Learning for Knowledge Graph Embedding) ?**: TracKGE further refines Transformer-based KGE by integrating relation-pattern adaptive contrastive learning. This approach leverages the Transformer’s ability to capture complex relation patterns and enhances it with contrastive learning objectives, which push positive examples closer and negative examples further apart in the embedding space. The "relation-pattern adaptive" aspect suggests that the contrastive learning loss is tailored to the specific relational patterns identified by the Transformer, leading to more robust and discriminative embeddings. \* **SDFormer (A shallow-to-deep feature interaction for knowledge graph embedding) ?**: SDFormer focuses on capturing feature interactions across different depths (shallow to deep) using a Transformer-like mechanism. This suggests an architectural innovation that allows for a more comprehensive understanding of how features evolve and interact at various levels of abstraction within the embedding process, leveraging the Transformer’s capacity for complex interaction modeling.

**Conditions for Success:** Transformer-based models excel on KGs where rich contextual information and long-range dependencies are critical for accurate link prediction or reasoning. Their ability to dynamically weigh the importance of different parts of the input sequence makes them highly adaptive to diverse relational patterns. CKGE ? demonstrates their success in domains requiring explainability.

**Theoretical Limitations:** \* **Computational Complexity:** The self-attention mechanism in standard Transformers has a quadratic computational complexity with respect to the input sequence length ( $O(N^2)$ ). For very large KGs or long paths, this can be prohibitively expensive, making scalability a major concern. This exacerbates the



computational cost issues seen in deep GNNs. \* **Graph Structure Integration:** While models like TGformer [10] and P-ART [11] adapt Transformers for graphs, the fundamental architecture was designed for linear sequences. Integrating the full richness of graph topology (e.g., cycles, varying degrees, multiple paths between nodes) into a sequential attention mechanism remains a challenge. Simple serialization of paths, as in CKGE [12], might lose some inherent graph properties. \* **Interpretability vs. Complexity:** While attention weights can offer some interpretability (e.g., highlighting important paths in CKGE [12]), the overall complexity of deep Transformer models can still make end-to-end reasoning opaque, a recurring tension in advanced KGE models.

**Practical Limitations:** \* **Data Requirements:** Training large Transformer models typically requires substantial amounts of data to learn effective attention patterns. \* **Hyperparameter Tuning:** The numerous hyperparameters (number of layers, heads, hidden dimensions, learning rates) make tuning challenging, as noted by [13]. \* **Memory Footprint:** The large number of parameters and intermediate attention matrices can lead to high memory consumption, limiting batch sizes and scalability. The parameter efficiency concerns raised by [14] for even simpler KGEs are amplified here.

**Comparative Framework and Evolution:** Transformer-based KGE models represent the cutting edge in leveraging global context and flexible interaction modeling.

- **Geometric Models (Section 2, 3):** Offer limited expressiveness and fixed scoring functions, lacking data-driven feature learning.
- **CNNs (Section 4.1):** Introduce local feature extraction but are not inherently graph-aware.
- **GNNs (Section 4.1):** Directly operate on graph structures, capturing local and multi-hop dependencies through message passing. They are more graph-native than CNNs but can struggle with long-range dependencies and global context.
- **Transformer-based Models (e.g., P-ART [11], TGformer [10], CKGE [12]):** Represent the most advanced deep learning architectures for KGE. They excel at capturing global contextual information and long-range dependencies through self-attention,

offering highly adaptive and expressive embeddings. They move beyond fixed aggregation schemes to dynamic weighting of information.

The evolution from GNNs to Transformers for KGE highlights a shift from local structural aggregation to global contextual understanding. While GNNs focus on how information propagates through immediate neighbors, Transformers aim to understand how any two entities or relations, regardless of their direct proximity, might interact within a broader context. A critical tension exists between the unparalleled expressiveness and contextualization offered by Transformers and their significant computational demands, especially for very large KGs. Researchers are actively exploring efficient Transformer variants and better ways to integrate graph topology to mitigate these practical limitations while retaining their powerful representational capabilities.

#### ### 4.3. Automated Search and Meta-Learning for KGE Architectures

As Knowledge Graph Embedding (KGE) models have grown increasingly complex, incorporating sophisticated deep learning architectures like CNNs, GNNs, and Transformers, the manual design and tuning of these architectures and their scoring functions have become a significant bottleneck. This challenge, compounded by the diverse characteristics of different knowledge graphs, has spurred the emergence of automated design and meta-learning approaches. These methods aim to automatically discover optimal KGE components, reducing manual effort, enhancing adaptability, and pushing the boundaries of KGE expressiveness beyond human-engineered solutions.

**Context and Problem Solved:** The manual process of selecting or designing the "best" scoring function, aggregation mechanism, or overall KGE architecture for a given task and dataset is arduous. It typically involves extensive trial-and-error, expert intuition, and hyperparameter tuning, which is empirically demonstrated to be highly dataset-dependent. This problem is exacerbated by the proliferation of new KGE models and the increasing complexity of deep learning components. Automated search and meta-learning address this by framing the architecture design as an optimization problem, seeking to find high-performing configurations automatically.

**Core Innovation:** These approaches leverage principles from Neural Architecture

Search (NAS) and meta-learning to explore a search space of possible KGE components or entire architectures. Instead of a human designing a fixed model, an algorithm searches for the optimal design.

- \* **AutoSF (Searching Scoring Functions for Knowledge Graph Embedding) ?:**
- \* **Core Innovation:** AutoSF proposes an automated search framework specifically for discovering optimal scoring functions in KGE. Instead of manually designing a scoring function (e.g., TransE’s distance, RotatE’s rotation), AutoSF defines a search space of elementary operations (e.g., addition, multiplication, concatenation, projection, non-linear activations) and uses a **reinforcement learning (RL) based controller** to search for a sequence of these operations that forms an effective scoring function. The controller is trained to maximize the performance of the sampled scoring functions on a validation set.
- \* **Conditions for Success:** AutoSF excels when the optimal scoring function is not immediately obvious or when existing hand-crafted functions are suboptimal for a particular KG. It can discover novel and highly effective scoring functions that might not be intuitively designed by humans.
- \* **Theoretical Limitations:** The search space for scoring functions can be vast, making exhaustive search infeasible. RL-based search can be computationally expensive and prone to local optima. The generalizability of a discovered scoring function to unseen KGs or tasks is not guaranteed.
- \* **Practical Limitations:** The computational cost of AutoSF can be very high, as it involves training and evaluating many candidate scoring functions. This can be a significant barrier for large KGs or resource-constrained environments. The complexity of the discovered functions might also hinder interpretability.
- \* **Message Function Search for Knowledge Graph Embedding (MFS) ?:**
- \* **Core Innovation:** MFS focuses on automating the design of message functions within Graph Neural Networks (GNNs) for KGE. In GNNs, the message function defines how information is transformed and passed between nodes. MFS defines a search space for these message functions, comprising various operations (e.g., different transformations, aggregations, non-linearities). It then uses an efficient search strategy (e.g., evolutionary algorithms or differentiable NAS) to find the optimal message function that maximizes KGE performance. This is a more granular approach than AutoSF, focusing on a specific, critical component of GNN-based

KGEs. \* **Comparison with AutoSF:** While AutoSF searches for the overall scoring function, MFS specifically targets the message functions within GNNs. This highlights a trend towards automating more granular components of complex KGE architectures. ? also explores efficient relation-aware scoring function search, indicating a broader interest in automating scoring function design. \* **Conditions for Success:** MFS is particularly beneficial for GNN-based KGEs where the choice of message function significantly impacts performance. It can adapt the message passing mechanism to the specific structural and relational properties of a given KG, leading to more effective GNNs. \* **Theoretical Limitations:** Similar to AutoSF, the search space for message functions can be large. The computational cost of searching and evaluating GNNs with different message functions can be substantial. \* **Practical Limitations:** The practical deployment of MFS requires significant computational resources. The discovered message functions, while optimal, might be complex and difficult to interpret. \* **AutoETER (Automated Entity Type Representation with Relation-Aware Attention for Knowledge Graph Embedding) ?:** \* **Core Innovation:** AutoETER focuses on automating the learning of entity type representations and integrating them with relation-aware attention. While not a full architecture search, it automates a crucial aspect of KGE design: how to effectively leverage entity types and attention mechanisms. It uses an automated approach to learn the optimal way to combine type information with entity embeddings and how attention should be applied. This is an example of automating specific feature engineering or component integration. \* **Interstellar: Searching Recurrent Architecture for Knowledge Graph Embedding ?:** \* **Core Innovation:** Interstellar extends the idea of automated search to recurrent architectures for KGE. This suggests searching for optimal recurrent neural network (RNN) structures or recurrent components within a larger KGE framework. This is particularly relevant for models that process paths or sequences within KGs. \* **Learning Dynamic Knowledge Graph Embedding in Evolving Service Ecosystems via Meta-Learning ?:** \* **Core Innovation:** This work applies meta-learning to dynamic KGE, specifically in evolving service ecosystems. Meta-learning (or "learning to learn") aims to train models that can quickly adapt to

new tasks or evolving data distributions with minimal new training. For dynamic KGE, this means learning an initialization or an update rule that allows the model to efficiently incorporate new temporal facts or adapt to changes in the KG structure over time. This addresses the challenge of continual learning in dynamic KGs, which is a major practical concern. \* **Dynamic Graph Embedding via Meta-Learning ?**: This paper also explores meta-learning for dynamic graph embedding, further emphasizing the field’s move towards models that can adapt efficiently to changing graph structures.

**Comparative Framework and Evolution**: Automated search and meta-learning represent the pinnacle of the field’s journey towards adaptable and high-performing KGE models.

- **Manual Design (Geometric, CNN, GNN, Transformer)**: The earlier paradigms relied heavily on human intuition and extensive hyperparameter tuning. This approach is limited by human capacity and the sheer complexity of modern deep learning architectures.
- **Automated Component Search (AutoSF ?, MFS ?, Interstellar ?)**: These methods automate the design of specific KGE components (scoring functions, message functions, recurrent architectures). They leverage NAS techniques to explore a predefined search space, leading to potentially superior and novel designs. This addresses the "how to design" problem for specific parts of the KGE pipeline.
- **Meta-Learning for Adaptability (e.g., ??)**: This goes beyond finding a static optimal architecture to learning how to *adapt* KGE models efficiently to new data, tasks, or evolving KGs. This addresses the "how to learn to learn" problem, crucial for dynamic and real-world applications.

A critical tension exists between the potential for discovering highly optimized and novel KGE architectures through automation and the significant computational resources required for the search process. While AutoSF ? and MFS ? promise to reduce manual effort, they introduce a new layer of computational complexity. Furthermore, the interpretability of automatically discovered architectures can be even more challenging than

hand-crafted ones. The field is actively exploring more efficient NAS techniques (e.g., differentiable NAS, one-shot NAS) and meta-learning strategies to make these powerful automated design paradigms more practical and accessible, ultimately aiming for KGE models that are not only expressive but also highly adaptable and sustainable in dynamic, real-world knowledge environments.

### 3 Enriching KGE with Semantic Context, Rules, and Multi-modality

The foundational Knowledge Graph Embedding (KGE) models, as discussed in preceding sections, primarily focus on learning representations from the structural triples  $(h, r, t)$  within a knowledge graph (KG). While translational, rotational, and even advanced deep learning architectures (Sections 2, 3, and 4) have significantly enhanced the capacity to capture intricate structural patterns and relational semantics, they often operate under the implicit assumption that the triple structure alone provides sufficient information. However, real-world KGs are far richer than mere collections of triples; they are often accompanied by diverse auxiliary information, governed by logical rules, and can be enriched by multi-modal data sources such as textual descriptions and visual features. This section delves into advanced KGE paradigms that move beyond purely structural learning, incorporating these additional layers of semantic and contextual information to create more discriminative, meaningful, and robust embeddings.

The motivation for enriching KGE with external information stems from several critical limitations of structural-only approaches. Firstly, **data sparsity** is a pervasive issue in KGs, where many entities or relations have limited connections, making it difficult for models to learn robust representations solely from observed triples  $??$ . Auxiliary data can provide crucial supplementary signals. Secondly, structural embeddings often lack **semantic richness and interpretability**. While a vector might capture a relation, it rarely conveys the nuances of entity types, attributes, or temporal validity that are readily available in other forms. For instance, knowing that "Barack Obama" is a "Person" and

"President" is an "Occupation" adds significant semantic depth beyond just the triple (Barack Obama, hasOccupation, President). Thirdly, purely data-driven KGEs can suffer from **logical inconsistency** and struggle to incorporate **prior domain knowledge**. Logical rules, whether explicit or implicit, are fundamental to human reasoning and can enforce consistency or guide the embedding process, leading to more reliable inferences. Finally, the world is inherently **multi-modal**, and entities in KGs often correspond to textual descriptions, images, or other sensory data. Fusing these diverse modalities can provide a holistic understanding, mitigate sparsity, and enhance the semantic grounding of embeddings, especially in specialized domains like chemistry ? or healthcare ?.

This section explores three primary avenues for enriching KGE:

1. **Incorporating Auxiliary Information and Entity Types:** This paradigm leverages structured metadata associated with entities and relations, such as entity types, hierarchical classifications, numerical attributes, and temporal validity. By explicitly integrating these features, KGE models can learn more discriminative embeddings that respect the inherent characteristics and constraints of the knowledge elements.
2. **Rule-based and Constraint-driven Embedding:** This approach injects logical rules and constraints, either hard or soft, directly into the embedding learning process. These rules, often derived from domain expertise or automatically mined, enforce consistency, guide reasoning, and inject valuable prior knowledge, thereby improving the logical soundness and predictive accuracy of the embeddings.
3. **Multi-modal KGE: Integrating Textual and Other Modalities:** This cutting-edge area focuses on fusing information from heterogeneous data sources beyond the graph structure itself, such as textual descriptions, visual features, and even audio. By leveraging powerful pre-trained language models (PLMs) and advanced fusion techniques, multi-modal KGE aims to overcome data sparsity, enhance semantic understanding, and enable richer cross-modal reasoning.

Collectively, these enrichment strategies represent a significant evolution in KGE research,

moving towards models that are not only capable of capturing complex structural patterns but also deeply grounded in the broader semantic, logical, and contextual fabric of knowledge. This shift is crucial for developing KGE models that are more accurate, interpretable, and applicable to the nuanced demands of real-world AI systems.

### 3.1 Incorporating Auxiliary Information and Entity Types

Traditional Knowledge Graph Embedding (KGE) models, particularly those based on translational or rotational principles (as discussed in Sections 2 and 3), often treat entities and relations as atomic symbols, represented by vectors in a continuous space. This approach, while effective for capturing structural patterns, overlooks a wealth of explicit semantic information available in many Knowledge Graphs (KGs). Auxiliary information, such as entity types, hierarchical structures, numerical attributes, and temporal validity, provides crucial context that can significantly enrich entity and relation representations, leading to more discriminative and meaningful embeddings.

**Context and Problem Solved:** The primary problem addressed by incorporating auxiliary information is the limited expressiveness and discriminative power of embeddings learned solely from structural triples. When entities or relations are sparse or ambiguous, their structural context alone might be insufficient to learn robust representations. Auxiliary data offers additional, often explicit, semantic cues that can resolve ambiguities, enforce consistency, and guide the embedding process towards more semantically aligned representations. For instance, knowing that "Paris" is a "City" and "France" is a "Country" helps differentiate the relation "locatedIn" from "capitalOf," even if both appear in similar structural contexts. This also helps mitigate the data sparsity challenge, as entities with few direct links might still have rich auxiliary descriptions.

**Mechanism and Core Innovations:** Approaches to integrate auxiliary information vary, but generally involve either concatenating auxiliary features with structural embeddings, projecting them into a shared space, or using attention mechanisms to weigh their relevance.

**Entity Types and Hierarchies:** Entity types (e.g., Person, Organization, Location) and



their hierarchical relationships (e.g.,  $\text{City} \subset \text{Location}$ ) provide strong semantic signals. \*

? proposed a **type-augmented KGE framework** for knowledge graph completion, explicitly integrating entity type information to enhance embeddings. Their core innovation lies in designing a mechanism that allows type embeddings to influence entity embeddings, making them more discriminative based on their semantic categories. Similarly, ? introduced **TransET**, a KGE model that incorporates entity types, demonstrating improved performance by leveraging this categorical information. \*

Beyond simple types, some works consider **hierarchical structures**. ? explored KGE with hierarchical relation structure, arguing that relations often exist within a hierarchy, and incorporating this can refine their representations. More recently, ? presented a **Hierarchical Hyperbolic Neural Graph Embedding** model, which leverages hyperbolic geometry to naturally represent hierarchical structures, allowing for more accurate embeddings of entities and relations within a type hierarchy. This is further supported by ? with DensE, which uses an enhanced non-commutative representation for KGE with adaptive semantic hierarchy, and ? with a low-dimensional gated hierarchical hyperbolic embedding. \*

The concept of **ontologies**, which formally define types and their relationships, is also leveraged. ? proposed OntoEA for ontology-guided entity alignment, using ontology information to guide the joint KGE process. ? analyzed the compatibility between vector space representations and rules derived from ontologies, highlighting the challenge of integrating symbolic ontological knowledge. Recent work like ? focuses on generating ontologies via KGE query embedding, showcasing a bidirectional relationship. \*

**Differentiation of Concepts and Instances:** ? introduced a method for differentiating concepts and instances for KGE, recognizing that these distinct types of entities require different embedding strategies. This aligns with ?, which proposes KGE with concepts, further emphasizing the importance of type-level information. \*

**Automated Type Representation:** ? developed AutoETER, an automated entity type representation with relation-aware attention, which learns how to best represent and integrate entity type information without manual feature engineering. This represents an evolution towards more adaptive and less human-dependent integration of auxiliary data.

**Attributes (Numeric and Categorical):** Entities often possess descriptive attributes (e.g., population for a city, birth year for a person). \* ? specifically addressed **KGE with Numeric Attributes of Entities**, demonstrating that incorporating these features can significantly enhance entity representations. This is typically achieved by encoding numerical attributes into vectors and integrating them with the structural embeddings. \* ? proposed integrating entity attributes for error-aware KGE, suggesting that attributes can also help in identifying and mitigating errors in the KG. Similarly, ? and ? explored attribute-enhanced KGE for recommendation systems, emphasizing the role of semantic relevance attributed to entities. \* ? highlighted the importance of learning **high-order structural and attribute information** for enhancing KGE, often using attention networks to weigh attribute relevance.

**Temporal Information:** Facts in KGs are often time-sensitive. Ignoring time leads to static and potentially inaccurate representations. \* ? introduced **HyTE (Hyperplane-based Temporally aware KGE)**, a pioneering work that explicitly models time by associating each timestamp with a hyperplane. This allows for temporal validity to be geometrically represented, enabling temporally-guided inference and prediction of temporal scopes for facts. *Building on this, ? proposed **ATiSE (Additive Time Series Embedding)**, which models the evolution of entity and relation representations as multi-dimensional additive time series (trend, seasonal, random components). Crucially, ATiSE represents embeddings as Gaussian distributions\** to capture temporal uncertainty, a significant advancement over deterministic temporal models. This idea was further explored by ? with TeRo, a time-aware KGE via temporal rotation, and ? with ChronoR, which uses rotations for temporal embeddings. \* More recent works continue to refine temporal KGE. ? introduced MADE, a multicurvature adaptive embedding for temporal KGE, and ? proposed IME, integrating multi-curvature shared and specific embedding for temporal KGE. These models leverage complex geometric spaces to better capture temporal dynamics. Other notable contributions include ? using tensor decomposition, ? with TeAST (Archimedean Spiral Timeline), ? with fuzzy spatiotemporal embeddings, ? with a timespan-aware graph attention model, and ? with a cross-dimensional recurrent

graph network. \* The field is also exploring **dynamic KGE** that can adapt to evolving KGs, as seen in ? and ? (using meta-learning for evolving service ecosystems). ? proposed context-aware temporal KGE, and ? explored translating in time domain space. ? introduced THOR, a self-supervised temporal KGE model.

**Contextual/Path Information and Disentangled Representations:** Beyond direct auxiliary features, the broader structural context, such as paths between entities or disentangled aspects of entities, can be seen as enriched auxiliary information. ? introduced *CKGE*, a *Contextualized KGE for explainable recommendation*, which constructs a meta-graph\* for each talent-course pair, integrating contextualized neighbor semantics and high-order connections. This meta-graph acts as rich auxiliary input for a specialized Transformer. \* Path-based methods, like ? with path-specific KGE and ? with Path-RotatE, explicitly leverage multi-hop paths as contextual information to refine embeddings. ? further improved this by using locally and globally attentive relation paths. ? introduced **DisenKGAT**, a significant innovation that learns disentangled entity representations. *This addresses the problem that entities often have multiple facets, and relations focus on distinct aspects. DisenKGAT uses relation-aware aggregation for micro-disentanglement and mutual information regularization for macro-disentanglement\**, leading to adaptive, robust, and interpretable embeddings. This moves beyond a single, monolithic embedding per entity to a more nuanced, multi-component representation. \* Other approaches include ? with orthogonal relation transforms and graph context modeling, ? with context-dependent KGE, and ? with LightCAKE for lightweight context-aware KGE. ? uses contextual facts to guide generation for KGE completion.

### **Comparative Framework:**

**Critical Analysis:** The integration of auxiliary information into KGE models represents a significant step towards more semantically rich and robust representations. The core innovation across these methods is to move beyond the simplistic view of entities and relations as atomic symbols and instead leverage their rich metadata. For instance, while earlier models like TransE ? would embed "Paris" as a single vector, type-augmented

Table 1: Comparative Framework for Incorporating Auxiliary Information

l X X X X	
<b>Auxiliary Type &amp; Method Family &amp; Core Mechanism &amp; Advantages &amp; Limitations</b>	
<b>Entity Types/Hierarchies</b>	& Type-augmented KGEs (??????); Hierarchical KGEs (??????) & Concatenation, projection, attention, hyperbolic geometry & Enhance discriminability, enforce semantic consistency, improve reasoning over categories & Requires explicit type/hierarchy data, potential for type noise, complexity of hierarchical modeling
<b>Attributes (Numeric/Categorical)</b>	& Attribute-aware KGEs (????) & Feature encoding (e.g., MLP for numeric), concatenation, attention & Provide rich descriptive context, mitigate sparsity, improve entity understanding & Requires attribute data, heterogeneity of attribute types, feature engineering complexity, potential for noise
<b>Temporal Information</b>	& Temporal KGEs (????????????????) & Time-aware transformations (hyperplanes, rotations), time series decomposition, Gaussian distributions, dedicated temporal encoders & Capture dynamic validity, enable temporal reasoning, predict future facts & Requires explicit timestamps/intervals, increased model complexity, challenges in modeling uncertainty and long-term trends
<b>Contextual/Path Information</b>	& Path-based KGEs (???); Context-aware KGEs (????????????) & Path encoders (RNNs, Transformers), meta-graph construction, attention over neighbors & Capture multi-hop dependencies, provide richer context for ambiguous entities/relations & Increased computational cost for path enumeration/encoding, risk of noise from irrelevant paths, scalability issues for long paths
<b>Disentangled Representations</b>	& DisenKGAT (?) & Relation-aware aggregation for micro-disentanglement, MI regularization for macro-disentanglement & Capture multi-faceted nature of entities, improve interpretability, adaptive to relation context & Requires careful design of disentanglement mechanisms, choice of component number (hyperparameter), computational cost of MI regularization

models ?? would ensure this vector is influenced by its "City" type, making it more distinct from a "Person" or "River" named "Paris."

These approaches succeed under conditions where the auxiliary data is available, accurate, and relevant to the task. For example, temporal KGEs like HyTE ? and ATiSE ? achieve superior performance on datasets with explicit timestamps, demonstrating their ability to capture dynamic knowledge. ATiSE’s innovation of modeling temporal uncertainty with Gaussian distributions ? addresses a key theoretical limitation of deterministic temporal models, leading to more robust predictions. Similarly, DisenKGAT ? excels when entities exhibit polysemy or multiple distinct facets, as its disentangled representations can capture these nuances, offering improved interpretability compared to monolithic embeddings.

However, several theoretical and practical limitations persist. A major theoretical challenge is the **heterogeneity of auxiliary data**. Integrating diverse types of information (e.g., categorical types, numerical attributes, temporal intervals) into a unified embedding space is non-trivial. Different modalities of auxiliary data may require distinct encoding strategies and fusion mechanisms, increasing model complexity. For instance, while ? shows the benefit of numeric attributes, their integration method might not be optimal for categorical types. Another limitation is the **assumption of auxiliary data availability and quality**. Many real-world KGs lack complete or accurate auxiliary information, limiting the applicability of these methods. Furthermore, the increased model complexity due to auxiliary data integration can lead to higher computational costs and more challenging hyperparameter tuning, as highlighted by ? for general KGEs. For example, the mutual information regularization in DisenKGAT ?, while theoretically sound for disentanglement, adds computational overhead.

Compared to structural-only KGEs, these methods generally offer higher accuracy and better reasoning capabilities, especially for tasks requiring fine-grained semantic understanding or temporal awareness. However, this comes at the cost of increased data requirements and model complexity. The field is moving towards more adaptive and automated ways of integrating auxiliary information, as seen in AutoETER ?, to reduce the burden of manual feature engineering and make these powerful techniques more accessible. The tension lies in balancing the desire for rich, context-aware embeddings with the practical constraints of data availability, model complexity, and computational efficiency.

### 3.2 Rule-based and Constraint-driven Embedding

While incorporating auxiliary semantic information enriches the content of embeddings, it does not inherently guarantee logical consistency or allow for the direct injection of prior domain knowledge. Rule-based and constraint-driven embedding approaches address this by integrating logical rules, either hard or soft, directly into the Knowledge Graph Embedding (KGE) learning process. This paradigm aims to bridge the gap between symbolic reasoning and sub-symbolic representations, enforcing consistency, improving

reasoning, and leveraging expert knowledge.

**Context and Problem Solved:** The primary problem that rule-based and constraint-driven KGE seeks to solve is the inherent lack of logical consistency and the inability to inject prior knowledge in purely data-driven KGE models. Traditional KGEs learn statistical patterns from observed triples, but they often fail to capture implicit logical relationships (e.g., transitivity: if A is a `parentOf` B and B is a `parentOf` C, then A is a `grandparentOf` C). Without explicit guidance, embeddings might represent logically contradictory facts or fail to generalize to unseen but logically inferable triples. This limits their utility in applications requiring robust and explainable reasoning.

**Mechanism and Core Innovations:** The integration of rules and constraints typically involves modifying the KGE training objective function.

**Soft Rules and Regularization:** Most approaches integrate rules as soft constraints, meaning they are encouraged but not strictly enforced. This is usually done through regularization terms added to the loss function. \* ? proposed guiding KGE with **iterative guidance from soft rules**. Their approach iteratively mines rules from the KG and then uses these rules to generate additional "pseudo-positive" or "pseudo-negative" triples, or as a regularization term, to guide the embedding learning. This iterative refinement helps the model learn representations that are more consistent with the rules. \* Building on this, ? introduced a method for KGE **preserving soft logical regularity**. This work focuses on ensuring that the embeddings adhere to logical regularities (e.g., symmetry, transitivity, inversion) by incorporating specific regularization terms that penalize deviations from these properties in the embedding space. This allows for a more principled way to inject logical knowledge. \* ? demonstrated the effectiveness of **improving KGE using simple constraints**, showing that even basic logical constraints can significantly enhance embedding quality. These constraints might include properties like "functional" or "inverse functional" relations. \* ? further explored **enhancing KGE with relational constraints**, highlighting how explicit constraints on relations can improve the discriminative power and logical consistency of embeddings. \* More recently, ? proposed SimRE, which uses **simple contrastive learning with soft logical rules** for KGE.

This approach combines the power of contrastive learning (pushing positives closer, negatives further) with the guidance of soft logical rules, leading to more robust and logically consistent embeddings.

**Hard Rules and Rule-Guided Learning:** Some methods attempt to enforce rules more strictly or use rules to directly influence the sampling or generation of training data.

\* ? presented **Logic Rules Powered KGE**, which integrates logical rules more directly into the embedding process, often by modifying the scoring function or the sampling strategy to ensure rule adherence. \* ? proposed **structure-augmented KGE for sparse data with rule learning**, where rules are not just constraints but also help in augmenting sparse data, thereby improving embedding quality in data-scarce scenarios.

\* ? introduced **rule-based data augmentation for KGE**, a practical approach where logical rules are used to generate synthetic training triples, effectively expanding the training data in a logically consistent manner. This is particularly useful for improving generalization and handling sparse relations. \* ? addressed **rule-enhanced noisy KGE via low-quality error detection**, demonstrating how rules can be used not only to enforce consistency but also to identify and mitigate noise in the KG, leading to more reliable embeddings. \* ? developed RuleE, a KGE model that performs **reasoning with rule embedding**. This approach embeds rules themselves into the same space as entities and relations, allowing for direct computation of rule satisfaction and inference within the embedding space. This is a powerful way to bridge symbolic and sub-symbolic reasoning.

\* The work by ? on a translation-based KGE preserving logical properties of relations is an early example of embedding models designed with explicit logical considerations. Similarly, ? uses a **logical-default attention graph convolution neural network** for link prediction, showing how logical principles can be integrated into deep learning architectures.

### **Comparative Framework:**

**Critical Analysis:** Rule-based and constraint-driven KGE models represent a crucial step towards making KGE more robust, logically consistent, and aligned with human reasoning. The core innovation is to explicitly inject symbolic knowledge into the sub-

Table 2: Comparative Framework for Rule-based and Constraint-driven KGE

Approach Family & Key Papers		Rule Integration Mechanism & Advantages & Limitations
<b>Soft Rule Regularization</b> & [1][X][X][X][X]		Add regularization terms to loss function, iterative guidance & Flexible, allows for some rule violations (robust to noisy rules), improves logical consistency & Weaker enforcement, requires careful tuning of regularization weights, rules must be formulated for embedding space
<b>Rule-Guided Data Augmentation/Sampling</b> & [2][3]		Generate synthetic triples, guide negative sampling, detect noisy triples & Addresses data sparsity, improves generalization, can enhance data quality & Relies on correctness of rules, potential for error propagation if rules are flawed, computational cost of rule application
<b>Rule Embedding/Direct Integration</b> & [4][5]		Embed rules directly, modify scoring functions, integrate into GNNs & Stronger logical enforcement, enables direct rule-based reasoning in embedding space & Increased model complexity, challenges in representing complex rules, potential for over-constraining embeddings

symbolic embedding space, addressing the fundamental limitation of purely statistical models that might learn spurious correlations or overlook crucial logical dependencies. For example, while a standard KGE might learn that (Germany, hasCapital, Berlin) is true, it might not implicitly understand the inverse (Berlin, isCapitalOf, Germany) without explicit training or a rule-based mechanism. Models like ? and ? directly address this by incorporating logical regularities or embedding rules themselves.

These approaches succeed when high-quality, relevant logical rules are available, either from domain experts or through automated rule mining. They are particularly effective in domains where logical consistency is paramount (e.g., medical KGs ?) or where data is sparse, and rules can provide valuable inductive biases. The iterative guidance from soft rules ? demonstrates how rules can progressively refine embeddings.

However, significant theoretical and practical limitations exist. A major theoretical challenge is the "**rule acquisition bottleneck**": obtaining a comprehensive set of accurate logical rules for large-scale KGs is often difficult and labor-intensive. Automated rule mining can be noisy and error-prone, potentially introducing incorrect biases into the embeddings, as implicitly acknowledged by ?'s focus on error detection. Another challenge is the **integration of discrete symbolic logic into continuous vector spaces**. While regularization terms offer a soft way to bridge this, they can be difficult to tune, and the degree to which rules are truly "enforced" can vary. Over-constraining the embedding



space with too many or too strict rules can limit the model’s flexibility and ability to discover novel patterns, leading to underfitting. ? explicitly discusses the compatibility challenges between vector space representations and rules.

Practically, rule-based methods often incur **increased computational costs**. Mining rules, applying them during training (especially for data augmentation ?), and computing complex regularization terms can significantly slow down the learning process. The interpretability of how rules influence the final embeddings can also be challenging, as the interaction between the structural learning objective and the rule-based constraints is complex.

Compared to auxiliary information methods (Section 5.1), rule-based approaches focus on enforcing *relationships* and *consistency* rather than merely enriching entity/relation content. While auxiliary data provides semantic context, rules provide logical structure. The tension lies in finding the right balance between allowing the model to learn from data and imposing prior logical knowledge, ensuring that rules enhance rather than hinder the discovery of new, valid patterns. The field is moving towards more sophisticated ways of learning and integrating rules, potentially through differentiable rule learning, to overcome these limitations.

### 3.3 Multi-modal KGE: Integrating Textual and Other Modalities

The previous sections highlighted how structured auxiliary information and logical rules can enrich KGE. However, knowledge is not confined to structured triples or explicit metadata. Real-world entities and relations are often described, depicted, or otherwise represented across diverse modalities such as text, images, and even audio. Multi-modal KGE aims to fuse information from these heterogeneous sources with structural embeddings, providing a holistic understanding that can overcome data sparsity, enhance semantic understanding, and enable richer reasoning, especially in specialized domains.

**Context and Problem Solved:** The core problem addressed by multi-modal KGE is the inherent incompleteness and data sparsity of KGs, coupled with the limited semantic depth achievable from structural data alone. Many entities in KGs have few direct links,

making it challenging for purely structural KGE models to learn robust representations. However, these entities might have rich textual descriptions (e.g., Wikipedia articles), associated images, or other sensory data. Multi-modal KGE leverages these complementary data sources to: (1) **Alleviate data sparsity** by providing alternative signals for entities with few structural connections; (2) **Enhance semantic understanding** by grounding embeddings in real-world perceptual data; and (3) **Enable cross-modal reasoning**, allowing inference that spans different types of information. This is particularly crucial for specialized domains where knowledge is often distributed across various formats, such as chemistry ? or healthcare ?.

**Mechanism and Core Innovations:** Multi-modal KGE involves encoding information from different modalities into a common embedding space and then fusing these representations with the structural KGE.

**Textual Information Integration:** Textual descriptions are the most common auxiliary modality, often providing rich semantic context for entities and relations. \* ? proposed **SSP (Semantic Space Projection)**, an early work that integrates textual descriptions of entities into KGE by projecting text embeddings into the same semantic space as structural embeddings. This allows the text to inform and refine the entity representations. \* ? explored KGE via **reasoning over entities, relations, and text**, demonstrating how textual information can be jointly learned with structural embeddings to improve reasoning capabilities. \* ? introduced **Joint Language Semantic and Structure Embedding** for knowledge graph completion, emphasizing the importance of jointly learning from both modalities. \* **Leveraging Pre-trained Language Models (PLMs):** The advent of powerful PLMs like BERT has revolutionized textual information integration. \* ? presented "Marie and BERT," a KGE-based Question Answering system for chemistry, which leverages BERT to encode textual questions and entity descriptions, then integrates these with KGE for robust QA. This highlights the power of PLMs in capturing nuanced semantic meanings from text. \* ? advanced drug-target interaction prediction by integrating KGE and ProtBert pretraining, demonstrating how specialized PLMs can extract domain-specific textual features. \* ? explored embedding Chain-of-Thought into

LLMs for KGE construction, suggesting a role for LLMs not just in encoding but also in reasoning and knowledge generation. Similarly, ? proposed a joint KG and Large Language Model for fault diagnosis, showcasing the synergy between structured knowledge and generative text models. \* ? used an LLM-enhanced embedding approach for KG accuracy evaluation, further illustrating the broad utility of PLMs. \* Other applications include using financial news for stock price prediction ?, combining text embeddings for academic search engines ?, and zero-shot text classification via KGE ?. ? provides a comprehensive survey on utilizing textual information in KGE.

**Visual Information Integration:** Images associated with entities can provide rich perceptual features. \* ? explored **multimodal reasoning based on KGE for specific diseases**, where visual features (e.g., from medical images) are fused with KGE to enhance diagnostic accuracy. \* ? introduced **Modality-Aware Negative Sampling for Multimodal KGE**, addressing the challenge of generating effective negative samples when dealing with heterogeneous modalities, a crucial aspect for robust training. \* ? proposed DFMKE, a **dual fusion multi-modal KGE framework for entity alignment**, which effectively combines visual and structural information to align entities across different KGs. \* ? presented MMGK, a **Multimodality Multiview Graph Representations and Knowledge Embedding** for Mild Cognitive Impairment diagnosis, showcasing the fusion of various visual and graph-based features.

**Other Modalities and Fusion Strategies:** The concept extends to any data type that can be embedded. \* **Spatio-temporal data:** ? mined urban flow patterns by fusing multi-source heterogeneous data (e.g., traffic, weather) with KGE. Similarly, ? focused on urban multi-source spatio-temporal data analysis. \* **Multi-source fusion:** ? used attention-based learning for predicting drug-drug interactions based on multisource fusion information, demonstrating the power of combining diverse data streams. \* **Fusion Architectures:** The key challenge is how to effectively combine these diverse modalities. Early fusion involves concatenating raw features, while late fusion combines predictions from separate modal encoders. Intermediate fusion, often using attention mechanisms or cross-modal transformers, is common, allowing modalities to interact at various levels. For

example, ? proposes integrating LLMs and Möbius Group Transformations for temporal KGE, a complex multi-modal, multi-geometric fusion.

### Comparative Framework:

Table 3: Comparative Framework for Multi-modal KGE

Modality Type & Method Family & Core Mechanism & Advantages & Limitations	
<b>Textual Information</b> & SSP (?); Joint Semantic-Structural (??); PLM-enhanced (?????) & Text encoders (CNN, RNN, Transformer), projection, concatenation, cross-modal attention & Overcomes sparsity, provides rich semantic context, enables text-based reasoning & Requires high-quality text, modality gap (aligning text/structural embeddings), computational cost of PLMs	
<b>Visual Information</b> & Multimodal Reasoning (?); Modality-Aware Sampling (?); Dual Fusion (?); Multiview Graph Reps (?) & Image encoders (CNN), projection, attention, specialized fusion networks & Grounds embeddings in perceptual data, useful for visual entities, enhances entity alignment & Requires high-quality images, large datasets, significant computational resources, modality gap	
<b>Other Modalities (Spatio-temporal, etc.)</b> & Multi-source Data Fusion (???) & Specialized encoders for each modality, concatenation, attention, graph neural networks & Comprehensive understanding, leverages diverse data streams for specific applications & High heterogeneity challenges, complex fusion architectures, data synchronization issues, scalability	

**Critical Analysis:** Multi-modal KGE represents the most comprehensive approach to enriching knowledge representations by leveraging the full spectrum of available data. The core innovation is the ability to synthesize information from disparate sources, creating embeddings that are deeply grounded in both symbolic structure and perceptual reality. This directly addresses the fundamental problem of data sparsity in KGs, as entities with few structural links can still be richly described by their associated text or images. For instance, an obscure historical figure might have limited triples but a detailed biography, which PLM-enhanced KGEs ?? can effectively leverage.

These approaches succeed when there is a meaningful correlation between different modalities and when high-quality multi-modal data is available. They are particularly effective in specialized domains like biomedicine ?? or e-commerce, where entities often have rich textual descriptions and visual representations. The use of pre-trained language models (PLMs) has been a game-changer, as they provide powerful, context-aware textual embeddings that significantly reduce the "modality gap" between text and structural

representations.

However, multi-modal KGE faces substantial theoretical and practical limitations. The primary theoretical challenge is the **modality gap**: effectively aligning and fusing representations from fundamentally different data types (e.g., symbolic triples, continuous text, pixel arrays) into a coherent embedding space. While various fusion strategies exist, determining the optimal interaction between modalities remains an open research question. Simply concatenating embeddings might lead to high-dimensional, noisy representations, whereas complex attention mechanisms can be difficult to interpret. Another limitation is the **heterogeneity and potential noise of multi-modal data**. Textual descriptions can be ambiguous, images might be irrelevant, and aligning these to specific KG entities is non-trivial. ?’s work on modality-aware negative sampling highlights the challenges of handling noise and imbalance in multi-modal contexts.

Practically, multi-modal KGE models are often **computationally expensive** due to the need for multiple encoders (one for each modality), complex fusion networks, and potentially large pre-trained models. This leads to increased training time, memory footprint, and deployment challenges, especially for large-scale KGs. Data acquisition and preprocessing for multiple modalities can also be a significant bottleneck.

Compared to auxiliary information (Section 5.1) and rule-based methods (Section 5.2), multi-modal KGE offers the most holistic understanding but also introduces the highest complexity and data requirements. While auxiliary information focuses on structured metadata and rules on logical consistency, multi-modal KGE aims for a comprehensive semantic grounding across all available data forms. The tension lies in effectively integrating these diverse, often unstructured, data sources without overwhelming the model or introducing excessive noise, while also ensuring scalability and interpretability. The field is actively exploring more efficient fusion architectures, self-supervised pre-training across modalities, and adaptive mechanisms to make multi-modal KGE more robust and practical.

## 6. Dynamic, Efficient, and Robust KGE

The preceding sections have meticulously explored the foundational principles of Knowledge Graph Embedding (KGE), ranging from basic translational and rotational models to those enriched with semantic context, logical rules, and multi-modal information. While these advancements have significantly enhanced the expressiveness and predictive power of KGE, they often operate under idealized assumptions that rarely hold in real-world scenarios. Knowledge Graphs (KGs) are not static entities; they evolve constantly, acquiring new facts, entities, and relations over time. Furthermore, their sheer scale demands efficient and scalable learning algorithms, while their inherent imperfections (noise, incompleteness) necessitate robust models. Finally, the increasing demand for privacy-preserving machine learning introduces new challenges for distributed KGE. This section addresses these critical, practical challenges, focusing on the dynamic nature, efficient deployment, and reliability of KGE models.

The motivation for this shift towards dynamic, efficient, and robust KGE is multifaceted. Firstly, the **temporal dimension** is intrinsic to real-world knowledge. Facts have validity periods, entities change properties, and relations evolve. Static embeddings fail to capture this dynamism, leading to outdated or inaccurate inferences. Secondly, the **open-world nature** of KGs means new entities and relations are constantly added, rendering traditional transductive KGE models (which require full retraining for new data) impractical. This necessitates inductive and continual learning capabilities. Thirdly, the **massive scale** of modern KGs (often billions of triples and millions of entities) poses severe computational and memory constraints, making efficiency and compression paramount for practical deployment. Finally, KGs are rarely pristine; they contain **noise, errors, and biases** due to automatic construction or human curation, demanding robust training strategies and effective negative sampling. The emerging landscape of **distributed and privacy-sensitive data** further introduces the need for federated learning paradigms.

This section is structured around three interconnected themes that collectively define the frontier of practical KGE:

1. **Temporal, Spatiotemporal, and Fuzzy Knowledge Graph Embedding:**

This area focuses on explicitly modeling the evolution of facts over time, incorporating spatial dimensions, and handling the inherent uncertainty in knowledge. It moves KGE from a static snapshot to a dynamic, evolving representation.

2. **Inductive, Continual, and Federated Learning for KGE:** This theme addresses the adaptability and scalability of KGE models in evolving and distributed environments. It explores how models can generalize to unseen entities (inductive), incrementally update with new information (continual), and learn collaboratively while preserving data privacy (federated).

3. **Efficiency, Compression, and Robustness in Training:** This final theme tackles the practical deployment challenges of KGE, including reducing computational cost and memory footprint, optimizing training strategies (e.g., negative sampling), and enhancing resilience to noisy or adversarial data.

Together, these areas represent a concerted effort to develop KGE models that are not only expressive and semantically rich but also capable of operating effectively in the complex, dynamic, and resource-constrained environments of real-world applications. This marks a crucial transition from theoretical exploration to practical realization, ensuring KGE’s continued relevance and impact across diverse intelligent systems.

## 6.1. Temporal, Spatiotemporal, and Fuzzy Knowledge Graph Embedding

Traditional Knowledge Graph Embedding (KGE) models, as discussed in earlier sections, primarily focus on learning static representations of entities and relations from a snapshot of a knowledge graph (KG). This approach fundamentally overlooks the dynamic nature of real-world knowledge, where facts have limited validity periods, entities change properties, and relations evolve over time. The explicit integration of temporal information, and in some cases spatial and fuzzy dimensions, is critical for enabling KGE models to perform

accurate reasoning, predict future events, and handle the inherent uncertainty of evolving knowledge.

**Context and Problem Solved:** The core problem addressed by Temporal Knowledge Graph Embedding (TKGE) is the inability of static KGEs to capture the time-varying nature of facts. A triple like (Barack Obama, presidentOf, USA) is only true for a specific time interval, and ignoring this temporal context leads to incomplete or even incorrect inferences. TKGE models aim to represent (head, relation, tail, timestamp) quadruples, allowing for temporally-guided link prediction, fact validation, and event forecasting. Spatiotemporal KGE extends this to geographic and location-aware contexts, while fuzzy KGE addresses the uncertainty or vagueness often associated with temporal or relational facts.

**Mechanism and Core Innovations:** Approaches to TKGE generally involve augmenting entity and relation embeddings with temporal information, often through dedicated temporal components or transformations.

**Temporal Knowledge Graph Embedding (TKGE):**

- \* **Early Approaches:** Initial TKGE models often treated time as an additional embedding vector or a simple linear transformation. However, these methods struggled to capture complex temporal patterns or the duration of facts.
- \* **Hyperplane-based Temporal Embeddings:** A significant innovation was ?’s **HyTE (Hyperplane-based Temporally aware KGE)**. HyTE’s core mechanism is to associate each timestamp with a corresponding hyperplane in the embedding space. This allows the model to represent the temporal validity of facts geometrically, where a fact is valid at a given time if its embeddings lie close to the hyperplane associated with that time. This approach enables temporally-guided inference and the prediction of temporal scopes for facts with missing time annotations. While a notable advancement, HyTE still primarily models time deterministically.
- \* **Additive Time Series Decomposition and Uncertainty:** *Building on the need for more nuanced temporal modeling, ? introduced **ATiSE (Additive Time Series Embedding)**. This model innovatively represents the evolution of each entity and relation as a multi-dimensional additive time series, decomposing it into trend, seasonal, and random components. Cru-*



cially, *ATiSE* models entity and relation representations as multi-dimensional Gaussian distributions at each time step, with the mean evolving according to the time series components and the covariance matrix explicitly capturing temporal uncertainty\*. This explicit modeling of uncertainty is a significant theoretical leap beyond deterministic temporal models, leading to more robust predictions. \* **Rotation-based Temporal Models:** Following the success of rotational models like RotatE ?, several TKGEs adopted similar geometric transformations. ? proposed **TeRo (Time-aware KGE via Temporal Rotation)**, which models temporal relations as rotations in a complex space. Similarly, ? introduced **ChronoR**, another rotation-based temporal KGE model. These models leverage the expressive power of rotations to capture complex temporal dynamics. \* **Advanced Geometric and Neural Architectures:** More recent works have explored even more sophisticated ways to model temporal dynamics. ? introduced **MADE (Multicurvature Adaptive Embedding)** and ? proposed **IME (Integrating Multi-curvature Shared and Specific Embedding)**, both leveraging complex geometric spaces to better capture the non-Euclidean nature of temporal evolution. Other notable contributions include ? using tensor decomposition for temporal KGE, ? with **TeAST (Temporal KGE via Archimedean Spiral Timeline)**, which embeds time on an Archimedean spiral, and ? with a Timespan-aware Graph Attention-based Embedding Model. ? introduces a Cross-Dimensional Recurrent Graph Network with neural Stochastic Differential Equation for temporal KGE, pushing the boundaries of dynamic modeling. \* **Dynamic and Continual KGE:** The field is also moving towards models that can adapt to continuously evolving KGs. ? explored dynamic KGE via local embedding reconstructions, and ? used meta-learning for dynamic KGE in evolving service ecosystems. These approaches are closely related to continual learning, discussed in Section 6.2.

**Spatiotemporal and Fuzzy Extensions:** \* **Spatiotemporal KGE:** Beyond pure temporal dynamics, some KGs require modeling spatial information. ? introduced **FSTRE (Fuzzy Spatiotemporal RDF KGE)**, which uses uncertain dynamic vector projection and rotation to handle both spatial, temporal, and fuzzy aspects simultaneously. This is a highly integrated approach for complex real-world data. Earlier work by ? and ? focused

on location-aware KGE for geographic question answering. More broadly, ? explored urban multi-source spatio-temporal data analysis, and ? incorporated environmental knowledge embedding and spatio-temporal graph attention networks for traffic flow prediction. ? proposed SSTKG for interpretable spatio-temporal dynamic information embedding.

\* **Fuzzy KGE:** This area addresses the inherent vagueness or uncertainty in knowledge, often related to temporal or relational facts. ? (FSTRE) explicitly integrates fuzzy logic. ? proposed GTransE for uncertain KGE, and ? focused on approximate inferring with confidence prediction based on uncertain KGE. ? explored uncertain KGE by combining multi-relation and multi-path information. These models move beyond binary truth values to represent degrees of belief or membership.

### Comparative Framework:

Table 4: Comparative Framework for Dynamic and Uncertain KGE

X X X X	
Approach Family & Key Papers	Core Mechanism & Advantages & Limitations
<b>Deterministic Temporal KGE</b> & HyTE ?; TeRo ?; ChronoR ?; MADE ?; IME ? & Time-aware transformations (hyperplanes, rotations), complex geometries & Captures temporal validity, enables temporal reasoning, often efficient & Assumes deterministic temporal evolution, struggles with uncertainty, complex non-linear patterns	
<b>Uncertain Temporal KGE</b> & ATISE ?; FSTRE ? & Gaussian distributions for uncertainty, additive time series decomposition & Explicitly models temporal uncertainty, more robust predictions & Increased model complexity, assumptions about distribution (e.g., Gaussian, diagonal covariance), computational cost	
<b>Spatiotemporal KGE</b> & FSTRE ?; ????? & Integration of spatial features, dedicated spatio-temporal encoders, fuzzy logic & Handles location-aware knowledge, richer context for geographic entities & High data heterogeneity, complex feature engineering, increased dimensionality, scalability for large spatial data	
<b>Fuzzy KGE</b> & FSTRE ?; GTransE ?; ?? & Fuzzy logic integration, confidence scores, multi-relation/path fusion & Models vagueness and degrees of belief, robust to imprecise data & Defining fuzzy membership functions, interpretability of fuzzy embeddings, increased complexity	
<b>Dynamic/Continual KGE</b> & ????? & Local embedding reconstructions, meta-learning, context-aware mechanisms & Adapts to evolving KGs, reduces retraining cost, handles new facts incrementally & Catastrophic forgetting, balancing old vs. new knowledge, computational overhead of continuous updates	

**Critical Analysis:** The evolution of KGE from static to dynamic and uncertainty-aware models is a crucial step towards their real-world applicability. The core innovation across these methods is the recognition that time, space, and uncertainty are not mere metadata but integral components of knowledge that must be explicitly mod-

eled. For instance, while a static KGE might struggle to differentiate between (Trump, presidentOf, USA) in 2016 and (Biden, presidentOf, USA) in 2020, TKGEs like HyTE ? and ATiSE ? can accurately capture these temporal distinctions.

A clear pattern emerges in TKGE research: an initial focus on deterministic temporal modeling, followed by the integration of uncertainty. **Deterministic temporal KGEs** like HyTE ? and TeRo ? primarily capture the validity period of facts or the temporal ordering of events. However, they implicitly assume a precise, unambiguous temporal evolution. In contrast, **Uncertain Temporal KGEs** such as ATiSE ? explicitly model temporal uncertainty using probabilistic distributions, typically Gaussian. This allows ATiSE to not only predict *what* will happen but also *how certain* that prediction is, a capability that deterministic models fundamentally lack because their scoring functions yield a single, fixed value for a given timestamp. This theoretical leap provides a more realistic representation of evolving knowledge, particularly in domains where information is inherently noisy or incomplete.

Furthermore, while rotation-based models like TeRo ? and ChronoR ? leverage the expressive power of rotations to model temporal relations, they often operate in a fixed geometric space. More recent works like MADE ? and IME ? explore **multi-curvature adaptive embeddings**, suggesting a progression towards more flexible and expressive geometric spaces that can better capture the complex, non-Euclidean nature of temporal dynamics. This indicates a convergence on the principle that temporal dynamics are often not linear or simple, but diverge on the specific geometric transformations best suited to model them.

These dynamic approaches succeed under conditions where temporal or spatial information is explicitly available and relatively dense. They are particularly effective in domains like news analysis, historical KGs, or urban computing, where events unfold over time and space. The integration of fuzzy logic, as seen in FSTRE ?, further enhances robustness by allowing models to handle imprecise or vague knowledge, which is common in human-curated KGs. Unlike traditional KGEs that enforce binary truth values, fuzzy KGEs acknowledge degrees of belief, making them more suitable for real-world scenarios

where facts are not always black and white.

However, several theoretical and practical limitations persist. A major theoretical challenge for TKGE is the **complexity of temporal patterns**. While linear trends and seasonal components (as in ATiSE ?) are a good start, real-world temporal dynamics can be highly non-linear, irregular, and involve complex causal dependencies that are difficult to capture with current models. The assumption of constant or simple covariance matrices for uncertainty (e.g., diagonal in ATiSE) might oversimplify the true nature of temporal uncertainty, as it fails to capture complex inter-dimensional dependencies in the embedding space. This limitation arises because modeling full covariance matrices for high-dimensional embeddings is computationally prohibitive, forcing a trade-off between expressiveness and tractability. For spatiotemporal KGE, the **heterogeneity of spatial data** (e.g., points, lines, polygons) and the computational cost of integrating complex spatial relationships remain significant hurdles. This is because spatial data requires specialized indexing and query mechanisms that are not natively supported by standard KGE architectures, leading to increased complexity and potential performance bottlenecks. Furthermore, defining and integrating fuzzy membership functions for fuzzy KGE can be subjective and increase model complexity, as there is no universal method for quantifying vagueness, often requiring domain expertise or heuristic approaches.

Practically, dynamic KGE models often incur **significantly higher computational costs** and require more complex data preprocessing. Storing and processing time-indexed embeddings, especially for large KGs, demands substantial memory and computational resources. For instance, if a static KGE requires  $O(N \cdot D)$  parameters for  $N$  entities and  $D$  dimensions, a TKGE that models entity evolution over  $T$  time steps might require  $O(N \cdot D \cdot T)$  parameters or more complex temporal components, leading to a substantial increase in memory footprint. The hyperparameter tuning for temporal components (e.g., trend, seasonality, rotation parameters) can also be more challenging, as highlighted by ? in a general KGE context, due to the increased number of interacting parameters. The evaluation of these models is also more complex, requiring metrics that account for temporal accuracy and uncertainty.

Compared to static KGEs, dynamic models offer superior accuracy and reasoning capabilities for time-sensitive tasks, but at the cost of increased complexity and data requirements. The field is actively exploring more efficient architectures (e.g., attention mechanisms in ?), more expressive geometric spaces (?), and better ways to quantify and propagate uncertainty to make these powerful dynamic models more practical and robust. The tension lies in balancing the desire for high expressiveness and realism with the need for computational tractability and data availability.

## 6.2. Inductive, Continual, and Federated Learning for KGE

The static, transductive nature of many traditional Knowledge Graph Embedding (KGE) models presents a significant bottleneck for their deployment in dynamic, real-world environments. Knowledge Graphs (KGs) are constantly evolving, with new entities, relations, and facts emerging daily. Furthermore, the increasing emphasis on data privacy necessitates distributed learning paradigms. This section delves into advanced KGE paradigms that address these challenges: Inductive KGE (generalizing to unseen entities), Continual KGE (adapting to evolving knowledge), and Federated KGE (privacy-preserving distributed learning).

**Context and Problem Solved:** The primary problem is the **transductive limitation** of most KGE models, which learn embeddings only for entities and relations observed during training. When new entities or facts appear, these models require full retraining, which is computationally prohibitive for large KGs. This leads to the "cold-start" problem for new entities. **Inductive KGE** aims to learn a function that can generate embeddings for unseen entities or relations based on their local context or attributes, without retraining. **Continual KGE** addresses the challenge of incrementally updating KGE models as new facts arrive, without suffering from "catastrophic forgetting" of previously learned knowledge. Finally, **Federated KGE (FKGE)** tackles the problem of learning KGE models from decentralized KGs distributed across multiple clients, where raw data cannot be shared due to privacy concerns.

**Mechanism and Core Innovations:**

**Inductive KGE:** \* **Contextual Aggregation:** A common approach for inductive KGE is to learn an aggregation function that combines information from an entity’s neighbors to form its embedding. This allows for generating embeddings for new entities by aggregating their local graph structure. ? proposed **Logic Attention Based Neighborhood Aggregation** for inductive KGE, using attention mechanisms to weigh the contributions of neighbors, enabling generalization to unseen entities. \* **Meta-Learning for Adaptation:** **Meta-learning** has emerged as a powerful paradigm for inductive KGE, allowing models to "learn to learn" how to generate embeddings for new entities or relations with only a few examples. ? introduced **Meta-Knowledge Transfer for Inductive KGE**, which leverages meta-learning to quickly adapt to new entities and relations, significantly reducing the data requirements for induction. This approach focuses on transferring meta-knowledge (how to learn embeddings) rather than specific embeddings. \* **Relation Graph-based Induction:** ? proposed **InGram (Inductive KGE via Relation Graphs)**, which learns entity representations by leveraging the structure of relation graphs. This allows for robust induction by focusing on how entities interact through relations, making it less dependent on direct entity features. \* **Parameter-Efficient Entity-Agnostic Learning:** A particularly innovative approach is **Entity-Agnostic Representation Learning (EARL)** by ?. Instead of learning a unique embedding for every entity (which scales linearly with KG size), EARL learns embeddings for a small, fixed set of "reserved entities." For all other entities, it uses universal, entity-agnostic encoders that transform their "distinguishable information" (e.g., connected relations, k-nearest reserved entities, multi-hop neighbors) into embeddings. This design makes EARL inherently inductive and highly parameter-efficient, as its parameter count does not grow with the number of entities. This is a paradigm shift from traditional KGEs. \* **Context-Enhanced Induction:** ? introduced **ConeE (Global and local context-enhanced embedding)** for inductive KGC, further emphasizing the role of rich contextual information in generalizing to new entities.

**Continual KGE:** \* **Incremental Updates:** Continual KGE focuses on updating existing embeddings efficiently as new facts arrive without catastrophic forgetting. ? pro-

posed **Incremental Update of KGE by Rotating on Hyperplanes**, adapting existing embeddings with new data using geometric transformations (inspired by TransH?). This method aims to preserve previously learned knowledge while integrating new information.

- \* **Parameter-Efficient Adaptation (LoRA)**: Recent advancements in parameter-efficient fine-tuning (PEFT) from large language models have been adapted for KGE. ? introduced **FastKGE** which incorporates an **Incremental LoRA (Low-Rank Adaptation)** mechanism. Unlike full-parameter fine-tuning methods that update all model parameters, IncLoRA adds small, trainable low-rank matrices to the existing model, allowing for efficient incremental updates with minimal additional parameters. This significantly reduces the computational cost of adaptation, achieving up to 68% training time reduction on new datasets while maintaining or improving accuracy. FastKGE also employs a graph layering strategy to sort and divide new entities based on their influence on the old KG, and an adaptive rank allocation for LoRA matrices based on entity importance.
- \* **Knowledge Distillation**: Another strategy is **incremental distillation**, where knowledge from an "old" model (trained on previous data) is transferred to a "new" model that incorporates new facts. ? explored **Continual KGE via Incremental Distillation**, using distillation to prevent catastrophic forgetting and ensure efficient knowledge transfer. Building on this, the IncDE method ? further refines incremental distillation by explicitly leveraging the explicit graph structure. It introduces a hierarchical strategy for learning new triples layer-by-layer and devises a novel incremental distillation mechanism for seamless transfer of entity representations, promoting old knowledge preservation. This approach demonstrates improvements of 0.2%-6.5% in MRR over baselines, highlighting the benefit of structure-aware distillation.
- \* **Meta-Learning for Evolving KGs**: Similar to inductive learning, meta-learning can also facilitate continual adaptation. ? applied **meta-learning for dynamic KGE in evolving service ecosystems**, enabling models to quickly learn and adapt to new patterns and entities in a streaming fashion.

**Federated KGE (FKGE)**: \* **Privacy-Preserving Collaborative Learning**: FKGE enables multiple clients (e.g., hospitals, companies) to collaboratively train a global KGE

model without sharing their raw, sensitive KG data. \* **Embedding-Contrastive Learning:** ? proposed **Federated KGC via Embedding-Contrastive Learning**. Their innovation lies in using contrastive learning to align the embedding spaces learned by different clients, ensuring consistency while keeping local data private. Clients only share embedding updates or gradients, not raw triples. \* **Communication Efficiency:** A major challenge in federated learning is communication overhead. ? addressed this with **Communication-Efficient Federated KGE with Entity-Wise Top-K Sparsification**. Clients only share the top-K most important entity embedding updates, significantly reducing bandwidth requirements. \* **Personalized FKGE:** Recognizing that clients might have unique knowledge or preferences, ? introduced **Personalized Federated KGE with Client-Wise Relation Graph**. This allows clients to learn personalized relation embeddings while collaboratively learning shared entity embeddings, balancing global consistency with local specificity. \* **Heterogeneity and Unlearning:** The FedLU framework ? specifically addresses data heterogeneity and knowledge forgetting in FKGE. It proposes mutual knowledge distillation to transfer local knowledge to global and absorb global knowledge back, coping with the drift caused by diverse client data. Furthermore, FedLU introduces an unlearning method based on cognitive neuroscience, combining retroactive interference and passive decay to erase specific knowledge from local clients and propagate it to the global model, a crucial feature for privacy compliance. \* **Robustness and Attacks:** The federated setting introduces new security vulnerabilities. ? investigated **Poisoning Attack on Federated KGE**, highlighting the need for robust defense mechanisms in FKGE. Similarly, ? focused on quantifying and defending against privacy threats.

#### **Comparative Framework:**

**Critical Analysis:** The development of inductive, continual, and federated KGE models marks a pivotal shift towards making KGE practical and scalable for real-world, dynamic, and privacy-sensitive applications. The core innovation is moving beyond the static, closed-world assumption to embrace the inherent dynamism and distributed nature of knowledge. For example, traditional KGEs would be paralyzed by the continuous influx



Table 5: Comparative Framework for Inductive, Continual, and Federated KGE

Approach Family & Key Papers & Core Mechanism & Advantages & Limitations	
<b>Inductive KGE</b> & [1][X][X][X][X]	& Contextual aggregation, meta-learning, entity-agnostic encoding & Generalizes to unseen entities/relations, reduces retraining cost, handles cold-start & Relies on local context (may struggle with truly novel entities), computational cost of aggregation/meta-learning, selection of reserved entities (EARL)
<b>Continual KGE</b> & [X][X][X][X][X]	& Incremental updates (geometric, LoRA), knowledge distillation (structure-aware), meta-learning & Adapts to evolving KGs, prevents catastrophic forgetting, efficient updates & Risk of catastrophic forgetting, balancing old/new knowledge, computational overhead of continuous adaptation
<b>Federated KGE</b> & [X][X][X][X][X]	& Collaborative training without raw data sharing, contrastive learning, sparsification, personalization, mutual distillation, unlearning & Preserves data privacy, leverages distributed data, enables collaborative intelligence & Communication overhead, heterogeneity of client data, privacy-utility trade-off, vulnerability to poisoning attacks

of new drugs and interactions in a biomedical KG, whereas an inductive model like EARL ? or a continual model like FastKGE ? could adapt efficiently.

A key pattern across these paradigms is the focus on **parameter efficiency and adaptive learning**. Inductive KGEs like EARL ? achieve parameter efficiency by decoupling the number of parameters from the number of entities, relying on universal encoders. Similarly, continual KGEs like FastKGE ? leverage LoRA to introduce minimal new parameters for incremental updates. This contrasts sharply with earlier transductive models that required a unique embedding for every entity, leading to parameter explosion. This suggests a fundamental constraint in scaling KGEs: direct parameterization of every entity/relation is unsustainable for open-world KGs.

While **Inductive KGE** methods like contextual aggregation ? or relation graph-based induction ? excel at generalizing to unseen entities, they often struggle with truly novel entities that lack sufficient local context or attributes. This limitation arises because these methods fundamentally rely on the existing graph structure or entity features to infer new embeddings; if an entity is entirely isolated or has no discernible features, the aggregation function has little information to work with. Meta-learning-based inductive KGEs ? offer a more robust solution for few-shot scenarios, demonstrating rapid adaptation by learning *how* to generate embeddings, but their effectiveness can still be limited by the diversity

and quality of meta-training tasks.

For **Continual KGE**, catastrophic forgetting remains a persistent problem; balancing the retention of old knowledge with the integration of new information is a delicate trade-off. While techniques like LoRA in FastKGE [10] offer parameter efficiency and significant training time reductions (e.g., 51%-68% on new datasets), they don't fundamentally solve the forgetting problem, but rather mitigate it by isolating new knowledge. In contrast, knowledge distillation methods like those in [11] and IncDE [12] directly address forgetting by transferring knowledge from older models. IncDE's innovation of leveraging explicit graph structure for hierarchical learning and incremental distillation further highlights that the *order* and *method* of knowledge transfer are crucial for effective continual learning, moving beyond simple replay or regularization.

**Federated KGE** offers a compelling solution for privacy-sensitive domains like healthcare, where data cannot be centralized. Approaches like embedding-contrastive learning [13] and communication-efficient sparsification [14] demonstrate the feasibility of collaborative KGE training. However, the **communication overhead** is a practical bottleneck, although sparsification techniques [14] mitigate this by sharing only top-K updates. More critically, the **privacy-utility trade-off** is an ongoing debate: stronger privacy guarantees often come at the cost of reduced model accuracy. For instance, while FedLU [15] addresses data heterogeneity through mutual knowledge distillation, it implicitly makes assumptions about the acceptable level of privacy leakage during knowledge transfer. Furthermore, FKGEs are susceptible to various **adversarial attacks**, including data poisoning [16] and inference attacks, necessitating robust defense mechanisms. The heterogeneity of data across clients also poses challenges for model convergence and global generalization, as local optima might diverge significantly.

Compared to transductive KGEs, these advanced paradigms offer unparalleled adaptability and privacy, but they introduce increased model complexity, computational overhead (especially for meta-learning and communication in federated settings), and new challenges in ensuring robustness and fairness. The field is actively exploring more robust meta-learning algorithms, more efficient and provably private federated learning protocols,

and unified frameworks that can seamlessly integrate inductive, continual, and federated capabilities to create truly "living" and secure KGE systems.

### 6.3. Efficiency, Compression, and Robustness in Training

The practical deployment of Knowledge Graph Embedding (KGE) models in real-world applications is often hindered by their substantial computational demands, memory footprint, and sensitivity to noisy or incomplete data. As KGs continue to grow in scale and complexity, the need for efficient, compressed, and robust training strategies becomes paramount. This section explores methods designed to address these critical challenges, focusing on optimizing resource utilization, enhancing model resilience, and refining the training process.

**Context and Problem Solved:** The core problems addressed here are the **scalability and reliability** of KGE models. Training KGE models on massive KGs (billions of triples) is computationally expensive and memory-intensive, making deployment on resource-constrained environments challenging. This necessitates methods for **efficiency and compression**. Furthermore, real-world KGs are inherently noisy, incomplete, and prone to errors, which can significantly degrade the quality of learned embeddings. This calls for **robustness techniques** that can mitigate the impact of noise, and **optimized negative sampling strategies** that are both effective and efficient, as negative sampling is a critical component of most KGE training.

**Mechanism and Core Innovations:**

**Efficiency and Compression:** \* **Parameter Efficiency:** A major source of inefficiency is the linear scaling of embedding parameters with the number of entities and relations. \* ?'s **Entity-Agnostic Representation Learning (EARL)**, discussed in Section 6.2, is a prime example. Its core innovation is to significantly reduce the parameter count by not learning individual embeddings for all entities. Instead, it relies on a small set of "reserved entities" and universal encoders that generate embeddings for other entities from their local context. This results in a stable, efficient parameter count independent of KG size, making it suitable for deployment on edge devices or in

federated learning. \* **Knowledge Graph Embedding Compression** techniques, as surveyed by ?, include methods like quantization (reducing precision of embeddings) and pruning (removing less important parameters). These techniques aim to reduce model size and memory footprint without significant performance degradation. \* ? proposed a **Lightweight KGE Framework for Efficient Inference and Storage**, focusing on architectures that minimize computational and storage requirements. \* **Training Efficiency**: Beyond parameter count, the training process itself can be optimized. \* **Parallel Training**: ? provided a comprehensive comparison of techniques for **parallel training of KGE models**, highlighting strategies for distributed computation across multiple GPUs or machines to accelerate learning on large KGs. \* **System-Level Optimizations**: ? introduced **GE2 (General and Efficient KGE Learning System)**, which focuses on system-level optimizations (e.g., efficient data loading, memory management, optimized computation kernels) to achieve faster training. \* **Communication-Efficient Training**: In distributed settings, communication overhead is critical. ? proposed **HET-KG (Communication-Efficient KGE Training via Hotness-Aware Cache)**, which uses caching strategies for frequently accessed embeddings to reduce communication bandwidth during distributed training. \* **Knowledge Distillation**: ? developed **DualDE (Dually Distilling KGE)** for faster and cheaper reasoning. Knowledge distillation involves training a smaller, "student" model to mimic the behavior of a larger, more complex "teacher" model, thereby achieving efficiency gains for inference. \* **Non-Sampling Methods**: Most KGE models rely on negative sampling, which can be computationally expensive. ? explored **Efficient Non-Sampling Knowledge Graph Embedding**, aiming to bypass the overhead of negative sampling altogether. \* **Efficient Query Serving**: For real-time applications, efficient query serving is crucial. ? presented **Atom (An Efficient Query Serving System for Embedding-based KGE Reasoning)**, which uses operator-level batching to speed up inference.

**Robustness to Noisy Data and Negative Sampling**: \* **Negative Sampling (NS) Strategies**: NS is fundamental for KGE training, as it generates "false" triples to teach the model what is incorrect. However, inefficient or poor NS can degrade model qual-

ity and slow training. \* ? provides a comprehensive review of **Negative Sampling in KGE Representation Learning**. \* **Adaptive and Informative NS**: ? introduced a **Confidence-Aware Negative Sampling Method for Noisy KGE**. Their innovation is to assign confidence scores to triples and use these scores to guide negative sampling, making the process more robust to noise. ? proposed **NSCaching (Simple and Efficient Negative Sampling)**, which caches negative samples to speed up training. ? provided a deeper **Understanding Negative Sampling in KGE**, highlighting its impact on model performance. The work by ? further emphasizes that the choice of negative sampling strategies, alongside loss functions and hyperparameters, has a substantial impact on model efficiency and accuracy, an area often neglected by prior works focusing solely on scoring functions. \* **Automated and Modality-Aware NS**: ? explored **Simple and Automated Negative Sampling**, reducing the need for manual tuning. For multi-modal KGEs, ? developed **Modality-Aware Negative Sampling**, addressing the unique challenges of generating effective negative samples when dealing with heterogeneous data types. \* ? proposed **correlation embedding learning with dynamic semantic enhanced sampling** for KGC, showing how semantic information can guide better negative sample selection. \* **Robustness to Noise and Attacks**: Real-world KGs contain inherent noise, and models can be vulnerable to adversarial attacks. \* **Confidence-based Learning**: ?’s confidence-aware NS implicitly contributes to robustness. Similarly, ? proposed **PASSLEAF (Pool-bAsed Semi-Supervised LEArning Framework for Uncertain KGE)**, which handles uncertainty in triples. ? introduced **Weighted KGE**, where triples are assigned weights (e.g., based on confidence) to reduce the impact of noisy data. \* **Adversarial Training and Reinforcement Learning**: ? moved **Towards Robust KGE via Multi-Task Reinforcement Learning**, using RL to learn embeddings that are more resilient to noise and adversarial perturbations. \* **Error Detection and Mitigation**: ? addressed **Rule-enhanced Noisy KGE via Low-quality Error Detection**, leveraging logical rules to identify and mitigate noise, connecting to rule-based methods (Section 5.2). ? proposed **Integrating Entity Attributes for Error-Aware KGE**, using attributes to detect and handle errors. \* **Un-**

**Understanding Vulnerabilities:** Research into **Data Poisoning Attack** against KGE by ?? is crucial for understanding the types of robustness needed. This work demonstrates how malicious data injection can degrade KGE performance, highlighting the need for robust defense mechanisms. \* **Bias and Fairness:** ? questioned whether KGE models are biased or if it’s the data they are trained on, emphasizing the need to analyze data quality and potential biases. ? specifically investigated **Degree Bias in Embedding-Based KGC**, showing how entity popularity can affect embedding quality.

### Comparative Framework:

Table 6: Comparative Framework for KGE Efficiency, Compression, and Robustness

X X X X	
Approach Family & Key Papers & Core Mechanism & Advantages & Limitations	
<b>Parameter Efficiency/Compression</b> & EARL ?; ?? & Entity-agnostic encoding, quantization, pruning & Reduces memory footprint, enables deployment on resource-constrained devices, faster inference & Trade-off between compression ratio and accuracy, potential for information loss, increased complexity of encoding (EARL)	
<b>Training Efficiency</b> & ?????? & Parallel training, system optimizations, caching, knowledge distillation, non-sampling & Accelerates training, improves scalability, reduces communication overhead & Complexity of distributed systems, overhead of distillation, non-sampling methods may have theoretical limitations	
<b>Negative Sampling Optimization</b> & ???????? & Confidence-aware sampling, caching, automated methods, modality-aware sampling, impact of training choices & Improves training quality, enhances robustness to noise, can speed up training & Still heuristic, optimal strategy is dataset-dependent, computational cost can remain high	
<b>Robustness to Noise/Attacks</b> & ????????? & Confidence weighting, RL for robust learning, error detection, adversarial training analysis & Mitigates impact of noisy data, defends against adversarial attacks, improves reliability & Defining/detecting noise is hard, defense mechanisms can be complex, trade-off with model expressiveness, ongoing arms race with attackers	

**Critical Analysis:** The focus on efficiency, compression, and robustness is paramount for moving KGE from academic benchmarks to real-world applications. The core innovation is to engineer KGE models and their training processes to be resilient to practical constraints and imperfections. For instance, EARL ? directly tackles the parameter explosion problem, a fundamental practical limitation for large KGs, by decoupling parameter count from entity count. This is a significant paradigm shift, enabling KGE deployment in scenarios previously deemed impossible due to resource constraints. Similarly, confidence-aware negative sampling ? and weighted KGE ? directly address the pervasive problem

of noisy data in real-world KGs, leading to more reliable embeddings.

A clear pattern in this area is the recognition that **training components beyond the scoring function are critical**. While early KGE research primarily focused on developing novel scoring functions, later works, as highlighted by ?, systematically investigate the impact of loss functions, hyperparameters, and negative sampling strategies. This suggests a shift from purely model-centric innovation to a more holistic optimization of the entire KGE pipeline. For example, while a complex scoring function might offer high expressiveness, its practical utility is limited if it cannot be trained efficiently or robustly on noisy, large-scale data.

These approaches succeed when the chosen techniques align with the specific constraints of the deployment environment (e.g., memory limits, computational budget) and the characteristics of the data (e.g., noise level, sparsity). Parallel training techniques ? are essential for leveraging modern hardware, offering significant speedups by distributing computation. In contrast, knowledge distillation ? offers a pathway to efficient inference by transferring knowledge to smaller models, trading off a potentially minor accuracy drop for substantial gains in deployment speed and memory. The systematic study of negative sampling ?? has been crucial for understanding and improving a core component of KGE training, moving from simple random sampling to more informed, adaptive strategies like confidence-aware sampling ?.

However, several theoretical and practical limitations persist. For **efficiency and compression**, there is an inherent trade-off between model size/speed and accuracy. Aggressive compression (e.g., very low-bit quantization) can lead to significant information loss, as reducing precision inevitably discards some learned nuances in the embedding space. EARL’s effectiveness ? depends on the quality of its entity-agnostic encoders and the selection of reserved entities, which can be complex and dataset-dependent. This means that while EARL offers parameter efficiency, its inductive capability is not universally guaranteed across all KG structures. For **training efficiency**, parallelization introduces communication overhead, where the benefits of distributed computation can be offset by the time spent transferring embedding updates between nodes. System-

level optimizations ? are often platform-specific, limiting their generalizability. Non-sampling methods ? may have theoretical limitations in terms of expressiveness or convergence guarantees compared to well-designed sampling, as negative samples provide crucial counter-examples for learning decision boundaries.

For **robustness**, a major theoretical challenge is the precise definition and detection of "noise" or "adversarial examples" in KGs, which is more complex than in domains like images or text due to the discrete and relational nature of the data. The effectiveness of negative sampling strategies is highly dependent on hyperparameters and the specific characteristics of the KG, making universal optimal strategies elusive. This is a methodological critique: many papers propose new negative sampling methods without providing clear guidelines or theoretical justifications for their general applicability. Furthermore, the arms race against **data poisoning attacks** ?? is ongoing; defenses can be bypassed by more sophisticated attacks, indicating that robustness is a dynamic rather than a static problem. Addressing biases, such as degree bias ?, requires careful consideration of the underlying data distribution and model design, as simply learning embeddings from biased data will inevitably propagate those biases.

Compared to earlier KGE research that prioritized expressiveness and accuracy on clean, static benchmarks, this area explicitly tackles the messy realities of real-world data and systems. The tension lies in achieving high performance under severe constraints and imperfections. The field is actively researching more adaptive compression techniques, self-supervised learning for robust representations, and advanced adversarial training methods to build KGE models that are not only powerful but also truly practical, scalable, and trustworthy.

## 7. Applications and Explainability of Knowledge Graph Embedding

Knowledge Graph Embedding (KGE) has evolved from a theoretical endeavor to a cornerstone technology underpinning a wide array of Artificial Intelligence applications. While



the preceding sections have meticulously detailed the diverse models and architectural innovations in KGE, this section shifts focus to their practical utility and profound impact across various downstream AI tasks. The true measure of KGE’s success lies not merely in its ability to represent knowledge effectively but in its capacity to empower intelligent systems to perform complex reasoning, make informed decisions, and offer transparent insights.

The journey of KGE applications typically begins with foundational tasks such as link prediction and knowledge graph completion, which serve as primary benchmarks for evaluating the intrinsic quality and expressiveness of learned embeddings. These tasks, while seemingly straightforward, are critical for the continuous growth and refinement of knowledge bases, directly impacting the reliability of any system built upon them. However, the utility of KGE extends far beyond these core evaluations. Its ability to project discrete entities and relations into a continuous, semantically rich vector space provides a powerful mechanism for bridging disparate information modalities and facilitating complex inference.

This section will demonstrate KGE’s effectiveness in a spectrum of sophisticated applications. We will first revisit the foundational importance of link prediction and knowledge graph completion, acknowledging their role as the bedrock of KGE development. Subsequently, the discussion will expand to showcase KGE’s transformative role in complex AI tasks such as Question Answering over Knowledge Graphs, where it adeptly bridges natural language understanding with structured knowledge, and Entity Alignment, a crucial process for integrating and harmonizing heterogeneous knowledge bases. Furthermore, we will explore KGE’s significant contributions to enhancing recommendation systems, including the development of cross-domain and explainable recommendation paradigms, which are vital for personalized user experiences. The practical impact of KGE is also evident in critical domain-specific applications, ranging from accelerating drug repurposing in healthcare to enabling nuanced knowledge proximity measurements in patent analysis.

A recurring and increasingly critical theme across these applications is the growing emphasis on **explainability**. As KGE-driven systems are deployed in high-stakes envi-

ronments, the demand for transparent and actionable insights from their predictions has become paramount. Users, domain experts, and regulatory bodies require not just accurate answers but also a clear understanding of *why* a particular prediction was made. This necessitates the development of KGE models that can articulate their reasoning, identify influential factors, and provide interpretable justifications, moving beyond opaque black-box operations. This section will therefore also explore the emerging research directions aimed at fostering greater transparency and interpretability in KGE-driven predictions, ensuring that these powerful AI tools are not only effective but also trustworthy and accountable. This trajectory reflects a broader paradigm shift in AI, where interpretability is no longer an afterthought but an integral design principle, especially for systems leveraging complex knowledge representations.

### ### Core Tasks: Link Prediction and Knowledge Graph Completion

The foundational utility of Knowledge Graph Embedding (KGE) models is most directly assessed through their performance on link prediction and knowledge graph completion (KGC) tasks. These tasks are not merely academic benchmarks but represent the intrinsic capability of an embedding model to capture and infer missing relationships within a knowledge graph (KG). They are crucial for the continuous growth, validation, and maintenance of KGs, directly impacting the reliability of any downstream application.

**Context and Problem Solved:** The core problem addressed by link prediction and KGC is the inherent incompleteness of real-world KGs. Even the largest KGs are sparse, with many potential facts missing. Link prediction aims to predict the missing head, relation, or tail in an incomplete triple  $(h, r, ?)$ ,  $(?, r, t)$ , or  $(h, ?, t)$ , respectively. KGC is a broader term encompassing these tasks, focusing on enriching the KG by inferring new, valid triples. These tasks are critical because they allow KGs to be automatically expanded and corrected, reducing the manual effort required for knowledge base curation.

**Mechanism and Core Innovations:** KGE models address these tasks by learning low-dimensional vector representations (embeddings) for entities and relations such that the plausibility of a triple  $(h, r, t)$  can be scored by a function  $f(h, r, t)$ . A higher score indicates a more likely true triple. The evolution of KGE models for these tasks has been

characterized by increasingly sophisticated scoring functions and training strategies:

**Scoring Functions:**

- \* **Translational Models:** Early seminal works like TransE ? (and its variants like TransH ?, TransR ?, TransD ?) represent entities as points and relations as translations in a vector space. For instance, TransE aims to satisfy  $h + r \approx t$  for true triples. These models are computationally efficient but struggle with complex relation patterns (e.g., one-to-many, symmetric relations).
- \* **Bilinear Models:** DistMult and ComplEx ? use bilinear forms to score triples, often operating in complex vector spaces. ComplEx, in particular, can handle symmetric, antisymmetric, and inverse relations effectively due to its use of complex numbers.
- \* **Rotational Models:** RotatE ? introduced the idea of representing relations as rotations from head to tail entities in a complex vector space, satisfying  $h \circ r \approx t$  (where  $\circ$  denotes element-wise product). This geometrically elegant approach is highly expressive and performs well on various relation types. Subsequent models like QuatE ? extended this to quaternion space, and HousE ? used Householder parameterization for enhanced expressiveness. More recent works like Fully Hyperbolic Rotation ? and SpherE ? explore non-Euclidean geometries.
- \* **Neural Network-based Models:** Approaches like ConvE ? and R-GCN ? leverage Convolutional Neural Networks (CNNs) or Graph Neural Networks (GNNs) to capture richer interactions between entity and relation embeddings. For example, ConvE reshapes head and relation embeddings into 2D matrices and applies convolutions. DisenKGAT ? further enhances this with disentangled graph attention networks, learning adaptive and robust entity representations by separating latent factors. This allows DisenKGAT to better capture complex relations by considering different aspects of an entity based on the context of the relation, leading to superior accuracy and robustness.
- \* **Geometric Transformations:** Beyond simple translations and rotations, models like GeomE ? and CompoundE ? explore 3D compound geometric transformations, while STaR ? combines scaling, translation, and rotation. These methods aim to increase the expressiveness by allowing more complex transformations in the embedding space.

**Training Strategies:**

- \* **Negative Sampling (NS):** A critical component, NS generates "false" triples to train the model to distinguish true from false facts. Early methods

used uniform random sampling, but more advanced techniques emerged. ? introduced a **Confidence-Aware Negative Sampling Method** for noisy KGE, which assigns confidence scores to triples and uses them to guide negative sampling, making the process more robust to noise. ? proposed **NSCaching** for efficient negative sampling. ? provided a deeper understanding of NS, and ? offered a comprehensive review. Recent works like ? explore automated NS, and ? developed modality-aware NS for multi-modal KGEs. The choice of negative sampling strategy, alongside loss functions and hyperparameters, has a substantial impact on model efficiency and accuracy, an area often neglected by prior works focusing solely on scoring functions ?. \* **Incorporating Context and Rules:** Many models enhance embeddings by integrating additional information. ? and ? used soft logical rules to guide KGE, improving completion accuracy by enforcing consistency. ? also explored logic rules. ? used text descriptions to enrich embeddings, while ? incorporated entity types. ? used logic attention-based neighborhood aggregation for inductive KGE, allowing models to generalize to unseen entities by aggregating contextual information.

### Comparative Framework:

Table 7: Comparative Framework for Core KGE Tasks: Link Prediction and KG Completion

Model Family & Key Papers & Core Mechanism & Advantages & Limitations
<b>Translational</b> & TransE ?; TransH ?; TransR ? & Entities as points, relations as translations & Simple, efficient, good for 1-to-1 relations & Struggles with complex relations (symmetric, N-N), limited expressiveness
<b>Bilinear</b> & DistMult ?; ComplEx ? & Bilinear product, complex vector spaces & Handles symmetric/antisymmetric/inverse relations, higher expressiveness & Less intuitive geometric interpretation, can be prone to overfitting
<b>Rotational</b> & RotatE ?; QuatE ?; House ? & Relations as rotations in complex/hypercomplex space & Highly expressive, captures diverse relation patterns, geometrically intuitive & Higher computational cost than translational, complex optimization
<b>Neural/GNN-based</b> & ConvE ?; R-GCN ?; DisenKGAT ? & CNNs, GNNs, attention mechanisms for rich interactions & Captures complex local/global graph structures, high accuracy & High computational cost, requires more parameters, potential for black-box behavior
<b>Context/Rule-enhanced</b> & ??; ?; ? & Integrates logical rules, textual features, entity types & Improves accuracy by leveraging external knowledge, enhances consistency & Relies on availability/quality of auxiliary data, rule extraction can be complex

**Critical Analysis:** The evolution of KGE models for link prediction and KGC demonstrates a clear progression from simple, efficient geometric models to complex, expressive neural architectures. Early work, exemplified by TransE [1], assumed a straightforward translational property, which, while computationally efficient, proved insufficient for the diversity of real-world relations. This limitation led to the development of more expressive models like RotatE [2], which leverages rotations in complex space to capture symmetric, antisymmetric, and inverse relations more effectively. The tension here lies in balancing model simplicity and efficiency against expressiveness and accuracy. While translational models are fast, they often fall short on complex datasets. Rotational and neural models offer superior performance but come with increased computational overhead and parameter counts.

A critical, often unstated, assumption in many KGE models is the **quality and completeness of the training data**. Most models assume that observed triples are true and unobserved ones are false (or at least less likely). However, real-world KGs are inherently noisy and incomplete. This assumption is systematically questioned by works like [3], which explicitly introduces confidence scores to mitigate the impact of noisy data during negative sampling. This highlights a methodological trend: moving beyond ideal data assumptions to develop more robust training strategies. The choice of negative sampling, as emphasized by [3], is not a secondary detail but a primary determinant of model performance and efficiency, directly impacting the quality of learned embeddings.

Furthermore, the field has increasingly recognized that structural information alone is often insufficient. Integrating **auxiliary information** such as textual descriptions [4] or logical rules [5] has become a common strategy to enhance embedding quality. This exemplifies a broader theme of knowledge fusion, where KGE acts as a unifying layer. For instance, while a purely structural model might struggle with sparse entities, incorporating their textual descriptions can provide crucial semantic context. However, this introduces practical limitations related to the availability and alignment of such auxiliary data, as well as the increased complexity of multi-modal model architectures.

The transition from transductive to **inductive KGE** (as discussed in Section 6.2)

is another significant evolutionary step. Most traditional KGE models are transductive, meaning they can only generate embeddings for entities seen during training. This makes them impractical for evolving KGs where new entities are constantly added. Inductive approaches, such as those using neighborhood aggregation [10] or entity-agnostic encoding [11], directly address this limitation, allowing KGE models to generalize to unseen entities. This shift is crucial for the scalability and real-world applicability of KGE.

Despite significant advancements, limitations persist. Many models still struggle with **highly sparse relations** or **n-ary relations** [12], where a fact involves more than two entities. The computational cost of training complex neural or rotational models on massive KGs can be prohibitive, necessitating efficient training strategies and compression techniques (as explored in Section 6.3). The theoretical limitations often stem from the geometric assumptions underlying the models; for example, translational models cannot easily capture hierarchical relations. Practical limitations include the difficulty of hyperparameter tuning, which, as shown by [13], can vary significantly across datasets.

In comparison to early KGEs, modern models offer significantly higher accuracy and expressiveness on link prediction and KGC. However, this often comes at the cost of increased complexity and computational demands. The field continues to seek a balance, exploring more efficient architectures [14], better negative sampling strategies [15], and methods to integrate diverse information sources, all while striving for greater robustness to real-world data imperfections. The trajectory indicates a move towards KGE models that are not only accurate but also adaptable, scalable, and resilient.

### ### KGE for Question Answering and Entity Alignment

Beyond the intrinsic evaluation tasks of link prediction and knowledge graph completion, Knowledge Graph Embedding (KGE) has proven instrumental in powering more complex, downstream AI applications that require bridging natural language with structured knowledge or integrating disparate knowledge sources. Question Answering (QA) over KGs and Entity Alignment (EA) are two such critical applications where KGE’s ability to represent semantic relationships in a continuous space offers significant advantages.

**Context and Problem Solved: Question Answering over KGs (KGQA)** ad-

addresses the challenge of enabling users to query knowledge graphs using natural language, rather than formal query languages like SPARQL. This requires mapping ambiguous natural language questions to precise structured queries or directly inferring answers from the KG’s semantic space. KGE helps by providing a rich, continuous representation of entities and relations, facilitating semantic matching between natural language components and KG elements. **Entity Alignment (EA)** tackles the problem of identifying equivalent entities across different, often heterogeneous, knowledge graphs. This is crucial for integrating and enriching KGs, as real-world knowledge is often distributed across multiple, independently developed sources. KGE provides a powerful mechanism for this by mapping entities from different KGs into a shared embedding space, where semantically equivalent entities should be close.

#### **Mechanism and Core Innovations:**

**KGE for Question Answering (KGQA):** \* **Semantic Matching:** KGE models learn embeddings for entities and relations, allowing natural language questions to be embedded into the same vector space. The answer can then be found by identifying entities or paths in the KG whose embeddings are most similar to the question’s embedding. ? explored general KGE-based QA, demonstrating how embeddings can facilitate direct answer retrieval. \* **Hybrid Approaches:** Many KGQA systems combine KGE with Natural Language Processing (NLP) techniques, particularly deep learning models. ? developed a BERT-based triple classification model using KGE for QA, where KGE provides a structured understanding that BERT can leverage for semantic matching. Similarly, ? presented "Marie and BERT," a KGE-based QA system specifically for chemistry, showcasing the power of combining contextualized language models with structured knowledge representations. This approach bridges the gap between the nuanced semantics of natural language and the precise structure of chemical knowledge graphs. \* **Geographic QA:** For specialized domains, KGE can be tailored. ? proposed SE-KGE, a location-aware KGE model for Geographic Question Answering, which integrates spatial semantics into embeddings to answer questions involving geographical entities and relations. This highlights KGE’s adaptability to domain-specific features. \* **Path-based Reasoning:** KGE

can also be used to identify relevant reasoning paths in the KG. ? introduced GETT-QA, a Graph Embedding based T2T Transformer for KGQA, which leverages graph embeddings to enhance transformer-based reasoning over KG paths, improving the handling of multi-hop questions. \* **Limitations:** KGQA systems still struggle with complex questions (e.g., those requiring multi-hop reasoning, temporal understanding, or numerical aggregation), ambiguity in natural language, and scalability for very large KGs. The explainability of the reasoning path, especially in deep learning-heavy hybrid models, remains a challenge, although some KGE models (as discussed in Section 7.4) are beginning to address this.

**KGE for Entity Alignment (EA):** \* **Shared Embedding Space:** The fundamental idea is to embed entities from two (or more) KGs into a common vector space. If two entities refer to the same real-world object, their embeddings should be close in this shared space. \* **Bootstrapping and Iterative Refinement:** A significant challenge in EA is the scarcity of initial labeled alignments (seed entities). ? proposed **BootEA (Bootstrapping Entity Alignment with KGE)**, a pioneering approach that iteratively labels likely entity alignments to expand the training data. BootEA’s core innovation is an **alignment editing method** combined with a **global optimal labeling strategy** (solving a max-weighted matching problem on bipartite graphs) to mitigate error accumulation during bootstrapping. This ensures that the iteratively labeled data is high-quality, leading to substantial improvements (e.g., 13

**Multi-view Embeddings:** To capture diverse aspects of entities, ? introduced **Multi-view KGE for Entity Alignment**. This approach integrates different types of information (e.g., structural, attribute, textual) into distinct embedding views, which are then combined to achieve more robust alignment. \* **Ontology-guided Alignment:** ? proposed **OntoEA (Ontology-guided Entity Alignment)**, which leverages ontological information (e.g., class hierarchies, property constraints) to guide the KGE process for EA. This helps in aligning entities even when structural overlap is minimal but semantic types are consistent. \* **Semi-supervised and Inductive EA:** ? explored **Semi-Supervised Entity Alignment via KGE with Awareness of Degree Difference**, addressing



the challenge of entities with varying connectivity. More recent work, like ?, provides a deeper understanding and improvements for KGE-based EA. The survey by ? offers a comprehensive review of graph embedding-based EA. \* **Limitations:** EA still faces challenges with scalability for extremely large KGs, handling highly heterogeneous KGs (e.g., vastly different schemas or languages), and dealing with complex alignment scenarios (e.g., one-to-many mappings, fuzzy matches). The quality of initial seed alignments remains critical for bootstrapping methods, as errors can propagate.

### Comparative Framework:

Table 8: Comparative Framework for KGE in Question Answering and Entity Alignment

X X X X	
Application Area & Key Papers & Core Mechanism & Advantages &	Limitations
<b>Question Answering (KGQA)</b> & ?; ?; ?; ?; ? & Semantic matching in embedding space, hybrid NLP+KGE models, path-based reasoning & Bridges NL to structured knowledge, handles semantic variations, supports complex queries & Struggles with ambiguity, multi-hop reasoning complexity, scalability, explainability of deep models	
<b>Entity Alignment (EA)</b> & BootEA ?; ?; ?; ?; ? & Shared embedding space, iterative bootstrapping, multi-view embeddings, ontology guidance & Integrates heterogeneous KGs, robust to schema differences, effective with limited seeds (BootEA) & Scalability for massive KGs, handling highly heterogeneous schemas, one-to-many alignments, error propagation in bootstrapping	

**Critical Analysis:** The application of KGE to KGQA and EA represents a significant leap from merely evaluating embedding quality to leveraging it for practical AI problems. The core innovation across both domains is the transformation of discrete, symbolic knowledge into a continuous, semantically rich space, which facilitates similarity-based reasoning and integration. For KGQA, this means moving beyond keyword matching to understanding the semantic intent of a natural language question. For EA, it means overcoming schema heterogeneity by finding semantic equivalence in a shared vector space.

In **KGQA**, the evolution shows a clear trend towards **hybrid models** that combine the strengths of KGE with advanced NLP techniques like Transformers (e.g., BERT in ??). While early KGE-only QA systems relied heavily on direct semantic similarity, they often struggled with the nuances of natural language and complex reasoning. Integrating KGE with BERT, for instance, allows for a more robust encoding of the question’s context and a more effective mapping to KG triples, leading to higher accuracy. However, this

fusion also introduces a practical limitation: the increased complexity of these hybrid models often makes their reasoning opaque, hindering explainability. The challenge of multi-hop reasoning remains, as accurately traversing and aggregating information from multiple KG links is computationally intensive and prone to error.

For **Entity Alignment**, the field has moved from simple embedding similarity to sophisticated iterative and multi-modal approaches. BootEA [10] stands out by addressing a fundamental practical limitation of earlier embedding-based EA methods: their heavy reliance on a large number of initial seed alignments. By introducing a robust bootstrapping process with global optimization and alignment editing, BootEA significantly improves precision, especially in low-resource scenarios, making KGE-based EA viable for more real-world KGs. This highlights a critical tension: while KGE offers a powerful way to find semantic matches, its effectiveness is often bottlenecked by the availability of labeled data, which BootEA systematically tries to overcome. The use of multi-view embeddings [11] and ontology guidance [12] further underscores the need to incorporate diverse information sources to handle the inherent heterogeneity of KGs, as relying solely on structural similarity can be insufficient.

A common theoretical limitation across both KGQA and EA is the **expressiveness of the embedding space**. While KGE models have become increasingly sophisticated, they still operate under certain geometric or algebraic assumptions that might not fully capture the intricate logical and contextual dependencies present in real-world knowledge. For instance, representing complex logical implications or temporal constraints within a fixed-dimensional vector space remains a challenge. Practically, the **scalability** of these applications to truly massive KGs (billions of entities and relations) is an ongoing concern. Training KGE models for such scale is computationally expensive (as discussed in Section 6.3), and performing real-time QA or EA inference over them requires highly optimized systems.

Compared to symbolic methods (e.g., rule-based QA, schema-matching EA), KGE-based approaches offer superior flexibility and robustness to noise and schema heterogeneity, as they operate on learned semantic representations rather than rigid rules. However,

symbolic methods often provide clearer, more interpretable reasoning paths. The current trend is towards hybrid systems that leverage the strengths of both, using KGE for semantic understanding and fuzzy matching, while incorporating symbolic reasoning or explicit path traversal for precision and explainability. The field continues to grapple with the trade-off between semantic richness and computational tractability, pushing towards more robust, scalable, and interpretable solutions for these complex knowledge-driven tasks.

### ### Recommendation Systems and Domain-Specific Applications

The ability of Knowledge Graph Embedding (KGE) to capture rich, multi-relational semantic information has made it an indispensable tool for enhancing recommendation systems and driving innovation in a multitude of domain-specific applications. KGE moves beyond simple item-user interactions, leveraging the vast interconnectedness of knowledge to provide more accurate, diverse, and often explainable recommendations, and to unlock insights in complex scientific and industrial contexts.

**Context and Problem Solved: Recommendation Systems** traditionally suffer from data sparsity (few user-item interactions) and the cold-start problem (difficulty recommending new items or to new users). KGE addresses these by enriching item and user representations with explicit knowledge from KGs, capturing implicit relationships and side information (e.g., item attributes, categories, user demographics). This allows for more personalized and context-aware recommendations. **Domain-Specific Applications** often involve highly specialized and complex knowledge that is difficult to model with generic AI techniques. KGE provides a structured yet flexible framework to represent domain knowledge (e.g., drug-target interactions in healthcare, patent relationships in intellectual property), enabling powerful analytical and predictive capabilities tailored to specific industry needs.

#### **Mechanism and Core Innovations:**

**KGE for Recommendation Systems:** \* **Enriching Item/User Representations:** KGE learns embeddings for items and users (if users are modeled as entities or linked to entities in the KG), effectively integrating their attributes, categories, and relationships with other entities. This helps alleviate sparsity by propagating information through the

KG. ? explored recurrent KGE for effective recommendation, capturing sequential user behavior within the KG context. \* **Path-based Reasoning for Recommendations:** KGE can identify multi-hop paths between users and items or between items, revealing implicit relationships that drive recommendations. For instance, if a user likes a movie, and that movie shares a director with another, the KGE can infer a strong connection. ? and ? used attentive KGE for personalized recommendation, where attention mechanisms highlight the most relevant paths or entities contributing to a recommendation. \* **Cross-Domain Recommendation:** KGE is particularly effective in cross-domain scenarios where knowledge from one domain can inform recommendations in another. ? proposed **Cross-Domain Knowledge Graph Chiasmal Embedding** for multi-domain item-item recommendation, enabling the transfer of preferences and knowledge across different item categories or platforms. This is crucial for users with diverse interests. ***Explainable Recommendation:** A significant advantage of KGE is its potential for explainability. By tracing the paths or relations in the KG that led to a recommendation, KGE can provide transparent justifications. ? introduced CKGE (Contextualized Knowledge Graph Embedding) for explainable talent training course recommendation. CKGE's core innovation lies in constructing motivation-aware meta-graphs for talent-course pairs and employing a novel KG-based Transformer with a local path mask prediction mechanism. This mechanism explicitly quantifies and highlights the saliency of different meta-paths, directly revealing why\* a course is recommended (e.g., "because it aligns with your career path and addresses skill gap X"). This moves beyond black-box predictions to actionable insights. Other works like ? (XTransE for e-commerce) and ? (enhanced by neural collaborative filtering) also focus on explainable recommendations. ? proposed knowledge-aware fine-grained attention networks with refined KGE for personalized recommendations, further improving explainability. \* **Limitations:** Scalability to extremely large user-item KGs, managing dynamic user preferences over time, and the cold-start problem for entirely new entities without any KG links remain challenges. Balancing exploration and exploitation in recommendations is also complex.*

**Domain-Specific Applications:** \* **Healthcare and Biomedicine:** KGE is a powerful

tool for drug discovery, disease diagnosis, and personalized medicine. \* **Drug Repurposing and Interaction Prediction:** KGE can identify novel drug-disease or drug-target associations by embedding biomedical entities (drugs, proteins, diseases) and their relationships. ? pioneered KGE for drug repurposing in rare diseases. ? further advanced this with **ensemble KGE for explainable drug repurposing for COVID-19**, providing transparent insights into drug mechanisms. ? used a comprehensive graph-based approach integrating KGE for drug-target interaction prediction. ? leveraged KGE with capsule networks for multi-label drug-drug interaction prediction. ?????? are other examples focusing on various aspects of drug interactions and adverse drug reactions. \* **Disease Diagnosis and Medicine Recommendation:** ? used multimodal reasoning based on KGE for specific disease diagnosis, integrating diverse data types. ? embedded dynamic graph attention mechanisms into clinical KGs for enhanced diagnostic accuracy. ? and ? applied KGE for safe medicine recommendation. \* **Patent Analysis:** KGE can analyze complex relationships in patent data. ? embedded patent metadata to measure **knowledge proximity**, which is crucial for identifying emerging technologies, competitive landscapes, and innovation trends. This application highlights KGE's ability to quantify abstract relationships. \* **Industrial and Manufacturing:** KGE is increasingly used in Industry 4.0 for defect diagnosis, knowledge management, and cognitive manufacturing. ? discussed leveraging KGE for Industry 4.0 use cases. ? proposed an industrial-KGE-enabled pathway for cognitive manufacturing. ? applied KGE for defect diagnosis in additive manufacturing. ? used graph embedding for explainable manufacturing knowledge recommendation in collaborative design. ? developed a multi-hierarchical aggregation-based GCN for industrial KGE towards cognitive intelligent manufacturing. \* **Other Diverse Applications:** KGE has been applied to academic search ?, API recommendation ?, urban flow pattern mining ?, anomaly detection in cloud computing ?, and even ecotoxicological effect prediction ?. The breadth of these applications underscores KGE's versatility.

#### **Comparative Framework:**

**Critical Analysis:** KGE's impact on recommendation systems and domain-specific

Table 9: Comparative Framework for KGE in Recommendation Systems and Domain-Specific Applications

Application Area & Key Papers & Core Mechanism & Advantages & Limitations	
<b>Recommendation Systems</b> & [1][X][X][X][X]	& KG-enriched item/user embeddings, path-based reasoning, cross-domain transfer, local path mask prediction & Addresses sparsity/cold-start, provides richer context, enables cross-domain, offers explainability (CKGE) & Scalability for massive KGs, dynamic preference modeling, cold-start for truly new entities
<b>Healthcare/Biomedicine</b> & [?][?][?][?][?]	& Embedding biomedical entities/relations, multi-modal fusion, ensemble methods & Accelerates drug discovery/repurposing, predicts interactions, aids diagnosis, provides explainability & Requires high-quality biomedical KGs, complex data integration, validation in clinical settings
<b>Patent Analysis</b> & [?]	& Embedding patent metadata (inventors, technologies, citations) & Quantifies knowledge proximity, identifies innovation trends, supports strategic decision-making & Relies on structured patent data, interpretability of abstract proximity measures
<b>Industrial/Manufacturing</b> & [?][?][?][?]	& Embedding industrial processes, components, defects, knowledge & Enables defect diagnosis, supports cognitive manufacturing, facilitates knowledge recommendation & Requires domain-specific KG construction, real-time data integration, handling sensor data

applications highlights its versatility as a semantic feature extractor and a reasoning engine. The core innovation across these areas is the ability to leverage the rich, explicit, and implicit relationships within KGs to augment traditional data sources, leading to more intelligent and context-aware outcomes. For recommendation systems, this means moving beyond simple co-occurrence to understanding *why* a user might like an item, based on its attributes, categories, and connections to other items. For healthcare, it means uncovering non-obvious connections between drugs and diseases that could accelerate discovery.

In **recommendation systems**, the evolution from basic KGE-enhanced collaborative filtering to sophisticated explainable and cross-domain models is evident. Early KGE models primarily helped alleviate data sparsity by providing richer item representations. However, the introduction of explicit path-based reasoning and attention mechanisms [?] marked a significant advancement, allowing models to not only make better predictions but also offer insights into *why* a recommendation was made. CKGE [?] exemplifies this by providing fine-grained, motivation-aware explanations for training course recommendations, a critical feature for building user trust and improving the efficacy of learning platforms. This addresses a major practical limitation of traditional black-box recommenders,

which often leave users questioning the rationale behind suggestions. The challenge here is the trade-off between the depth of explanation and its human comprehensibility; overly complex paths might not be useful.

For **domain-specific applications**, KGE’s success stems from its ability to model the intricate relationships inherent in specialized knowledge. In **healthcare**, KGE has become a vital tool for drug repurposing ?? and drug-drug interaction prediction ?. The core innovation here is the transformation of vast, unstructured biomedical literature and databases into a structured KG, which KGE can then process to uncover novel associations. This addresses a critical practical limitation of traditional methods, which struggle with the sheer volume and heterogeneity of biomedical data. However, a theoretical limitation is the reliance on the completeness and accuracy of the underlying biomedical KGs, which are often noisy and incomplete. Furthermore, validating KGE-derived hypotheses in clinical settings remains a significant hurdle.

In **patent analysis**, KGE’s ability to measure "knowledge proximity" ? offers a powerful new lens for innovation intelligence. By embedding patent metadata, KGE can quantify the relatedness of technologies, inventors, or companies, providing insights that are difficult to obtain through keyword-based searches. This addresses a practical limitation of traditional patent analysis tools, which often lack the semantic depth to uncover subtle connections. However, the interpretability of these abstract proximity measures can be a challenge; translating embedding distances back into actionable business intelligence requires careful domain expertise.

A recurring pattern across these applications is the need for **domain-specific KG construction and curation**. While KGE models are general, their effectiveness in specialized domains hinges on the quality and comprehensiveness of the knowledge graphs they operate on. This often involves significant effort in data extraction, schema mapping, and validation, which can be a practical bottleneck. Furthermore, the integration of KGE with other domain-specific AI techniques (e.g., molecular docking simulations in drug discovery, time-series analysis in industrial fault diagnosis) is becoming increasingly common, leading to more powerful hybrid systems.

Compared to traditional methods, KGE-based approaches offer superior semantic richness, flexibility, and often better performance due to their ability to leverage relational context. However, they introduce increased model complexity, data requirements (for KG construction), and computational costs. The field is actively exploring ways to make KGE more adaptable to evolving domain knowledge (continual KGE, Section 6.2) and more robust to noisy data (Section 6.3), ensuring that these powerful tools can be reliably deployed in high-stakes, real-world environments. The tension lies in balancing the desire for deep, contextualized insights with the practical constraints of data availability, computational resources, and the need for transparent, trustworthy outcomes.

### ### Towards Explainable Knowledge Graph Embedding

As Knowledge Graph Embedding (KGE) models become increasingly complex and are deployed in critical, high-stakes applications—from healthcare to finance and personalized recommendations—the demand for explainability has moved from a desirable feature to an absolute necessity. Users, domain experts, and regulatory bodies require not just accurate predictions but also transparent, actionable insights into *why* a particular prediction was made. This section explores the burgeoning research area dedicated to making KGE models and their predictions more interpretable.

**Context and Problem Solved:** The core problem is the **black-box nature** of many advanced KGE models. While they achieve high predictive accuracy, the underlying mechanisms that lead to a specific embedding or prediction are often opaque. This lack of transparency hinders trust, makes debugging difficult, impedes compliance with regulations (e.g., GDPR’s “right to explanation”), and prevents users from gaining actionable insights. Explainable KGE (XKGE) aims to address this by providing human-understandable justifications for KGE-driven predictions, revealing the most influential entities, relations, or reasoning paths.

#### **Mechanism and Core Innovations:**

**Path-based Explanations:** ***Explicit Path Tracing:** A natural way to explain KGE predictions is to identify and highlight the most relevant paths within the KG that support a particular inference. For recommendation systems, ?’s **CKGE (Contextualized KGE)***



for talent training course recommendation employs a **local path mask prediction** mechanism. This innovation explicitly quantifies the saliency of different meta-paths between a talent and a course, directly revealing why\* a course is recommended (e.g., "because it aligns with your skill development path via relation X and Y"). This provides a clear, interpretable justification rooted in the KG structure. \* **Reasoning Path Identification:** For link prediction, models like XTransE ? aim to provide explainable link predictions by identifying relevant "lifestyles" or contextual paths in e-commerce KGs. Similarly, ? proposed an explainable knowledge reasoning framework that leverages multiple KGE models to provide richer explanations.

**Attention Mechanisms and Feature Importance:** *Highlighting Influential Neighbors/Relations:* KGE models that incorporate attention mechanisms (e.g., Graph Attention Networks) can inherently provide a form of explainability by assigning weights to different neighbors or relations during embedding aggregation. These weights indicate their relative importance. **DisenKGAT ? (Disentangled Graph Attention Network)**, for instance, uses a "relation-aware aggregation" mechanism where attention scores can reveal which neighboring entities and relations are most influential for a given entity's representation in a specific context. This provides insights into the micro-level\* influences on an embedding. **Disentangled Representations:** DisenKGAT ? further enhances interpretability through **disentangled representation learning**. By decomposing an entity's embedding into several independent components (micro-disentanglement via relation-aware aggregation and macro-disentanglement via mutual information regularization), DisenKGAT allows for explanations that pinpoint which aspect\* of an entity (e.g., its "genre" aspect, its "actor" aspect) is relevant for a particular prediction. This provides a more structured and compositional explanation of the underlying semantics.

**Rule Extraction and Integration:** \* **Symbolic Explanations:** Logical rules are inherently interpretable. Some approaches aim to extract rules from KGE models or integrate rules into KGE training. Rule ? learns rule embeddings, allowing for reasoning with explicit rules. Similarly, works like ? (Logic Attention) and ?? (Iterative Guidance from Soft Rules) integrate logical rules to guide embedding learning, making the resulting

embeddings more consistent with human-understandable logic. These rules can then serve as direct explanations. \* **Compatibility with Ontologies:** ? analyzed the compatibility between vector space representations and rules, suggesting that KGEs can be designed to better align with symbolic knowledge, thereby facilitating rule-based explanations.

**Post-hoc Interpretability and Counterfactuals:** \* **Model-agnostic Techniques:** Standard interpretability techniques like LIME or SHAP can be applied post-hoc to KGE predictions to identify features (entities, relations, paths) that contribute most to an outcome. \* **Inference Reconciliation:** ? explored inference reconciliation for KGE in supporting robot actions, aiming to explain discrepancies between expected and actual outcomes. \* **Confidence-aware Explanations:** ? proposed SpherE, an expressive and interpretable KGE model for set retrieval, which inherently provides confidence scores that can be used to qualify explanations.

#### Comparative Framework:

Table 10: Comparative Framework for Explainable Knowledge Graph Embedding (XKGE)

l X X X X	
Approach Family & Key Papers	Core Mechanism & Advantages & Limitations
<b>Path-based Explanations</b> & CKGE ?; XTransE ?; ?	Identifies and highlights influential KG paths/meta-paths for predictions & Direct, intuitive, rooted in KG structure, human-understandable & Paths can be too long/complex, computationally intensive to find optimal paths, limited for non-path-based reasoning
<b>Attention/Disentanglement</b> & DisenKGAT ?	Attention weights for neighbor importance, disentangled latent factors & Reveals micro-level influences, provides compositional explanations (aspect-based) & Interpretability of latent factors can be abstract, still requires human interpretation of components
<b>Rule Extraction/Integration</b> & RuleE ?; ??; ?	Extracts logical rules from KGE or guides KGE with rules & Inherently interpretable (symbolic logic), ensures consistency with domain knowledge & Rule incompleteness, difficulty in extracting complex rules, potential trade-off with embedding flexibility
<b>Post-hoc/Counterfactual</b> & ?	Model-agnostic interpretability tools, perturbing inputs to observe changes & Applicable to any KGE model, identifies influential inputs & Explanations can be local (not global), fidelity to true model reasoning can be limited, computational cost

**Critical Analysis:** The drive towards explainable KGE represents a crucial maturation of the field, acknowledging that utility in real-world, high-stakes scenarios demands more than just predictive accuracy. The core innovation is the shift from treating KGE

models as opaque black boxes to designing them with inherent transparency or developing methods to peer inside their reasoning. This addresses a fundamental practical limitation: the lack of trust and accountability in complex AI systems.

A clear pattern in XKGE research is the convergence of **symbolic and sub-symbolic approaches**. While KGEs are inherently sub-symbolic, methods like rule extraction [10] or rule-guided training [11] attempt to bridge this gap by providing explanations rooted in human-understandable logic. This highlights a recurring tension in AI: the trade-off between the flexibility and expressiveness of continuous embeddings and the interpretability of discrete symbols. The success of these hybrid approaches suggests that combining the strengths of both paradigms is key to achieving both high performance and transparency.

**Path-based explanations**, exemplified by CKGE [12], are highly intuitive because they directly leverage the graph structure. By identifying salient meta-paths, CKGE provides explanations that are directly traceable to the knowledge graph, making them easily verifiable by domain experts. This addresses a major practical limitation of abstract embedding-space explanations. However, a theoretical limitation is that not all KGE models rely on explicit path traversal for their predictions; some learn global patterns that are not easily decomposable into discrete paths. Furthermore, for very dense KGs, the number of potential paths can be overwhelming, making it challenging to identify and present truly concise and informative explanations.

**Disentangled representations**, as pioneered by DisenKGAT [13], offer a more structured and compositional form of explainability. By separating latent factors, DisenKGAT allows for explanations like "the prediction is due to the 'chemical structure' aspect of drug A and its 'target protein' aspect." This moves beyond merely identifying *what* was important to *which aspect* of an entity was important. This is a significant theoretical advancement, as it aims to make the internal workings of the embedding more interpretable. However, a practical limitation is that the interpretability of these latent factors often still requires careful human analysis and mapping to domain concepts, and the optimal number of disentangled components ( $K$ ) is a hyperparameter that needs careful tuning.

A critical challenge for XKGE is the **evaluation of explainability itself**. Un-

like accuracy, which is objectively quantifiable, "good" explanations are subjective and context-dependent. Metrics for human comprehensibility, fidelity to the model's true reasoning, and actionability are still evolving. This methodological gap means that while many XKGE methods are proposed, their comparative effectiveness in providing *useful* explanations is often hard to rigorously assess. This is an unstated assumption in many papers: that the proposed explanation mechanism is inherently good, without robust user studies or domain expert validation.

Compared to traditional, black-box KGEs, explainable models offer a significant advantage in terms of user trust, debugging, and compliance. However, this often comes with increased model complexity, computational overhead (e.g., for path search or disentanglement regularization), and a potential trade-off with raw predictive performance. The field is actively exploring more robust evaluation methodologies, developing inherently interpretable KGE architectures, and integrating XKGE with other explainable AI techniques to ensure that future knowledge-driven AI systems are not only powerful but also transparent, trustworthy, and aligned with human values. This represents a crucial trajectory for the responsible development and deployment of KGE technology.

## 8. Conclusion and Future Directions

The journey of Knowledge Graph Embedding (KGE) research has been a testament to the field's intellectual dynamism, evolving from foundational geometric intuitions to sophisticated deep learning paradigms. This extensive review has chronicled the remarkable advancements in representing the intricate web of entities and relations within knowledge graphs (KGs) as continuous, low-dimensional vectors. The intellectual trajectory of KGE has been characterized by a persistent pursuit of greater expressiveness, efficiency, and adaptability, collectively addressing fundamental challenges in knowledge representation. Early models, rooted in simple geometric transformations, laid the groundwork by demonstrating the feasibility of embedding symbolic knowledge. However, their inherent limitations in capturing complex relational patterns spurred a rapid diversification into more powerful architectures, including bilinear models, rotational transformations in

complex spaces, and the integration of deep neural networks.

This evolution reflects a broader paradigm shift in artificial intelligence, where the ability to distill vast, discrete knowledge into semantically rich, continuous representations has become paramount for enabling intelligent systems. The field has systematically confronted challenges such as the inherent incompleteness and noise in real-world KGs, the static nature of traditional representations in dynamic environments, and the computational demands of scaling to massive knowledge bases. Breakthroughs in negative sampling strategies ??, the development of inductive KGE models ??, and the explicit modeling of temporal dynamics ?? underscore this progress. Furthermore, the increasing demand for transparency in AI has propelled KGE research towards explainable models ??, ensuring that these powerful tools are not only effective but also trustworthy.

The synthesis of insights from diverse methodological families—ranging from translation-based models like TransE ? to graph neural networks like DisenKGAT ? and parameter-efficient designs like EARL ?—illustrates a collective effort to overcome the limitations of individual approaches. This interconnectedness of research efforts has led to a cumulative knowledge building, where each advancement builds upon or critically refines its predecessors. For instance, the efficiency concerns of complex models are addressed by techniques like parameter-efficient learning, while their expressiveness is leveraged in applications requiring nuanced semantic understanding. This concluding section consolidates these core achievements, critically examines the remaining challenges, and outlines promising future directions, emphasizing the profound societal impact of KGE research.

## Summary of Key Advancements

The intellectual journey of Knowledge Graph Embedding (KGE) has been marked by a profound evolution in how symbolic knowledge is represented, moving from simplistic geometric models to highly sophisticated deep learning architectures. This trajectory reflects a continuous drive to enhance the expressiveness, efficiency, and adaptability of KGE models, collectively addressing the inherent complexities of real-world knowledge graphs (KGs).

The earliest foundational models, such as TransE [1], introduced the intuitive concept of representing entities as points and relations as translations in a low-dimensional vector space. This paradigm, while computationally efficient and effective for simple one-to-one relations, quickly revealed its limitations when confronted with more complex relational patterns like symmetric, antisymmetric, or many-to-many relationships. This critical gap spurred the development of variants like TransH [2], which projected entities onto relation-specific hyperplanes, and TransR [3], which introduced relation-specific projection matrices, thereby increasing expressiveness. The subsequent emergence of rotational models, exemplified by RotatE [4], marked a significant breakthrough. RotatE’s innovation of representing relations as rotations in a complex vector space allowed for a more elegant and powerful capture of diverse relational patterns, including symmetry and inversion, outperforming its translational predecessors on various benchmarks. Building on this, models like QuatE [5] and HouseE [6] extended relational rotations to hypercomplex spaces (quaternions) and Householder parameterization, respectively, further pushing the boundaries of geometric expressiveness. More recently, the exploration of non-Euclidean geometries, such as hyperbolic embeddings [7] and Lie groups [8], has shown promise in capturing hierarchical structures more naturally and efficiently, a limitation often observed in Euclidean space models.

Alongside these geometric advancements, the field witnessed a parallel surge in the adoption of neural architectures, particularly Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs). CNN-based models like ConvE [9] moved beyond simple scoring functions by applying convolutional filters to reshaped entity and relation embeddings, capturing richer local interaction features. This approach demonstrated superior performance on link prediction by leveraging the power of deep learning for feature extraction. GNNs, such as R-GCN [10], revolutionized KGE by explicitly modeling the graph structure through iterative message passing and neighborhood aggregation. This allowed GNNs to learn context-aware entity representations by integrating information from multi-hop neighbors, a capability largely absent in earlier models. A notable advancement in this family is DisenKGAT [11], which introduced a disentangled graph attention network.

DisenKGAT’s core innovation lies in learning adaptive, robust, and interpretable entity representations by separating latent factors through a relation-aware aggregation mechanism and mutual information regularization. This directly addresses the limitation of static entity representations in capturing the multi-faceted nature of entities in different relational contexts, leading to superior accuracy and robustness in Knowledge Graph Completion (KGC). Recent research has also seen the adaptation of Transformer architectures for KGE, such as TGformer [?] and Position-Aware Relational Transformer [?], leveraging their attention mechanisms for global dependency modeling.

The field has also made significant strides in addressing practical challenges related to data quality and scalability. The pervasive issue of noisy and incomplete KGs led to the development of more sophisticated negative sampling strategies. While early KGE models relied on uniform random negative sampling, which often generated uninformative negatives, works like [?] introduced confidence-aware negative sampling to mitigate the impact of noisy data. NSCaching [?] focused on efficiency, while [?] provided a comprehensive review of negative sampling, highlighting its critical role. The integration of logical rules [??] further enhanced embedding quality by enforcing consistency and leveraging external knowledge. For the challenge of temporal dynamics, HyTE [?] introduced a hyperplane-based approach to model time-aware relations, associating each timestamp with a hyperplane. Building on this, ATiSE [?] proposed an innovative additive time series decomposition for entity and relation representations, modeling their evolution as multi-dimensional Gaussian distributions to explicitly capture temporal uncertainty, a significant departure from deterministic temporal models.

Scalability and efficiency have remained central concerns. The parameter explosion problem, where embedding parameters grow linearly with the number of entities, was critically addressed by approaches like Entity-Agnostic Representation Learning (EARL) [?]. EARL’s core innovation is to learn embeddings for a small set of reserved entities and use universal, entity-agnostic encoders to transform distinguishable information (e.g., connected relations, k-nearest reserved entities) into representations for all other entities. This drastically reduces parameter count while maintaining competitive performance, making

KGE deployment feasible in resource-constrained environments. Further advancements include federated learning for KGE ?? and embedding compression techniques ?.

Finally, the utility of KGE has been demonstrated across a wide array of applications, from fundamental tasks like link prediction and entity alignment to more complex ones like question answering and recommendation systems. BootEA ? significantly advanced entity alignment by introducing a robust bootstrapping process with global optimal labeling and alignment editing, effectively addressing the scarcity of initial labeled alignments. For recommendation systems, KGE has moved beyond simple item-user interactions to provide explainable and cross-domain recommendations. CKGE ?, for instance, constructs motivation-aware meta-graphs and employs a novel KG-based Transformer with a local path mask prediction mechanism to provide explicit, path-based explanations for training course recommendations, directly addressing the demand for transparency. This evolution from opaque black-box models to systems capable of offering interpretable justifications marks a crucial milestone in KGE research, integrating insights from various methodological families to collectively advance the field.

## Open Challenges and Theoretical Gaps

Despite the remarkable advancements in Knowledge Graph Embedding (KGE) research, several fundamental open challenges and theoretical gaps persist, hindering the full realization of KGE’s potential in real-world, complex scenarios. These challenges often represent inherent trade-offs or areas where current models make simplifying assumptions that do not hold universally.

One of the most pervasive and unresolved debates revolves around the **expressiveness versus efficiency trade-off**. More complex models, such as rotational embeddings like RotatE ? or deep neural architectures like GNNs ? and Transformers ?, offer significantly higher expressiveness and accuracy in capturing intricate relational patterns. However, this often comes at the cost of increased computational complexity, higher memory footprint, and longer training times, as highlighted by empirical studies on hyperparameter effects ?. Conversely, simpler translational models like TransE ? are computationally



efficient but lack the expressive power to handle diverse relation types. The field continuously grapples with finding an optimal balance, or developing adaptive mechanisms that can dynamically adjust complexity based on the specific task and data characteristics. For instance, while Entity-Agnostic Representation Learning (EARL) [?] addresses parameter efficiency, its performance might still be constrained by the inherent expressiveness of its underlying scoring function compared to a fully parameterized, but less efficient, counterpart.

A significant theoretical gap lies in **handling complex relational patterns beyond triplets**. The vast majority of KGE models are inherently triplet-centric, designed to embed (head, relation, tail) structures. This design struggles with n-ary relations (facts involving more than two entities, e.g., "Person A bought Product B from Seller C at Location D") [?], hierarchical relations [?], and complex logical constructs. While some models attempt to linearize n-ary relations or use hyper-relational embeddings, a unified and principled approach that naturally integrates such complex structures into the embedding space without significant information loss or increased complexity remains elusive. This limitation stems from the fundamental assumption that knowledge can be adequately represented by binary relationships, which often breaks down in rich, real-world KGs.

The modeling of **temporal and dynamic KGs** presents another profound challenge. While models like HyTE [?] and ATiSE [?] have made strides in incorporating time, they often simplify complex temporal phenomena. Current models struggle with nuanced temporal reasoning, such as causality, periodicity, duration, and the precise ordering of events. ATiSE’s innovative use of Gaussian distributions for temporal uncertainty [?] is a step forward, but it still relies on simplified trend and seasonal components. Furthermore, the ability to perform **continual learning** on evolving KGs, where new facts, entities, and relations are constantly added, is critical. Most KGE models require retraining from scratch, which is computationally prohibitive. Emerging work on dynamic KGE [?] and incremental updates [?] is promising, but robust and efficient solutions for real-time adaptation remain an open problem.

**Inductive generalization and the cold-start problem** continue to be significant

hurdles. While GNN-based KGEs ? and meta-learning approaches ? offer some inductive capabilities by learning functions to generate embeddings for unseen entities based on their local graph structure, robust generalization to entirely new entities or relations with minimal or no observed connections remains challenging. EARL ? addresses this by learning entity-agnostic encoders, but its effectiveness for entities with truly novel structural patterns is still an area for further investigation. The assumption that new entities will always share sufficient structural or attribute similarity with known entities to derive meaningful embeddings is often unstated but critical.

**Robustness to noise, incompleteness, and inherent biases** in real-world KGs is a persistent concern. Despite advancements in negative sampling ??, KGE models can still be sensitive to data quality. The field often implicitly assumes that the training data, even if incomplete, is largely representative and unbiased. However, KGs can contain systematic biases ? or be vulnerable to adversarial attacks ?, leading to skewed embeddings and unreliable predictions. Developing KGE models that are inherently resilient to such imperfections, perhaps through more sophisticated uncertainty quantification or robust optimization techniques, is crucial.

Perhaps the most pressing challenge, particularly for high-stakes applications, is **interpretability and explainability**. Despite efforts to develop explainable KGE (XKGE) models ???, many deep KGE models remain largely black-box. The abstract nature of vector space operations makes it difficult to provide human-understandable justifications for predictions. While path-based explanations ? and disentangled representations ? offer promising avenues, the rigorous evaluation of "good" explanations is itself an open methodological problem. There is a lack of standardized metrics and user studies to assess the fidelity, comprehensibility, and actionability of explanations, leading to an unstated assumption that proposed explanation mechanisms are inherently useful without empirical validation from domain experts.

Finally, the effective integration of **multi-modal and heterogeneous KGs** remains a complex task. KGs often exist alongside textual descriptions, images, sensor data, and time series. While some models incorporate textual features ? or entity types ?, a co-

herent and scalable framework for fusing diverse modalities into a unified, semantically rich embedding space, while preserving the unique characteristics of each modality, is still under active development. This involves not just concatenating embeddings but learning deep, cross-modal interactions. These challenges underscore the need for continued innovation, pushing beyond current paradigms to develop KGE models that are more robust, adaptive, interpretable, and capable of handling the full complexity of real-world knowledge.

## Emerging Trends and Broader Societal Impact

The trajectory of Knowledge Graph Embedding (KGE) research is increasingly shaped by emerging technological trends and a growing awareness of its broader societal implications. As KGE models become more sophisticated and integrated into critical AI systems, their impact extends far beyond academic benchmarks, influencing various sectors and raising important ethical considerations.

One of the most significant emerging trends is the development of **hybrid models and Neuro-Symbolic AI**. The limitations of purely sub-symbolic KGEs (e.g., lack of inherent interpretability, difficulty with complex logical reasoning) are driving a convergence with symbolic AI. This involves combining KGE with explicit logic rules ??? to enforce consistency and provide interpretable reasoning paths. More recently, the advent of Large Language Models (LLMs) has opened new frontiers, with KGE being integrated to ground LLMs in structured knowledge, enhance their factual accuracy, and enable more complex reasoning ?. This hybrid approach aims to leverage the semantic richness and generalization capabilities of KGE/LLMs while retaining the precision and explainability of symbolic reasoning. For instance, KGE can provide a structured context for LLMs to generate more accurate and relevant answers in question-answering systems, mitigating the "hallucination" problem often associated with LLMs.

Another prominent trend is the continued exploration of **Geometric Deep Learning on KGs**. Moving beyond Euclidean spaces, researchers are increasingly investigating non-Euclidean geometries such as hyperbolic ??, spherical ?, and Lie group embeddings

?. These geometries offer unique advantages for representing specific types of knowledge: hyperbolic spaces are particularly adept at capturing hierarchical structures and tree-like relations more efficiently than Euclidean spaces, while spherical embeddings can model cyclic or periodic relationships. This trend is driven by the theoretical insight that the intrinsic geometry of real-world KGs is often non-Euclidean, and aligning the embedding space's geometry with the data's inherent structure can lead to more expressive and efficient representations. Compound geometric transformations ? further enhance this expressiveness.

The increasing concern for data privacy and distributed computing has spurred research into **Federated and Privacy-Preserving KGE**. As KGs often contain sensitive information and are distributed across multiple organizations, training KGE models in a centralized manner becomes impractical or undesirable. Federated KGE approaches ?? allow multiple clients to collaboratively train a global KGE model without sharing their raw data, addressing privacy concerns. However, this introduces new challenges such as communication efficiency ? and vulnerability to poisoning attacks ?, which are actively being researched. Quantifying and defending against privacy threats on federated KGE is a critical area ?.

Beyond prediction, KGE is moving towards enabling **causal inference and counterfactual reasoning**. Current KGE models primarily capture correlations and associations. However, for high-stakes applications like drug discovery or policy-making, understanding causal relationships ("does X cause Y?") and performing counterfactual analysis ("what if X had not happened?") is paramount. Richer KGEs, combined with causal inference frameworks, could unlock these capabilities, moving AI systems from merely predicting outcomes to understanding underlying mechanisms.

The development of **Human-in-the-Loop KGE** systems is also gaining traction. This involves designing interactive frameworks where human experts can provide feedback to refine KGE models, validate predictions, and interpret explanations. Such systems can address the limitations of fully automated KGE, particularly in domains requiring high accuracy and trust. This also ties into the broader push for **Responsible AI**, where KGE

models are designed with fairness ?, accountability, and transparency as core principles.

The broader societal impact of KGE is profound and multifaceted:

\* **Healthcare and Biomedicine:** KGE is revolutionizing drug discovery and repurposing ??, accelerating the identification of novel drug-target interactions ??, and enhancing disease diagnosis ? and personalized medicine. The ability to integrate vast, heterogeneous biomedical knowledge into an explainable framework ?? promises to significantly improve patient outcomes and accelerate scientific discovery.

\* **Education and Talent Management:** Explainable KGE models, such as CKGE for talent training course recommendation ?, are transforming personalized learning and professional development. By providing transparent, motivation-aware explanations, these systems can guide individuals towards optimal skill development paths, fostering greater engagement and efficacy in learning.

\* **Industrial AI and Smart Manufacturing:** KGE is a cornerstone of Industry 4.0, enabling advanced applications like defect diagnosis in additive manufacturing ?, explainable manufacturing knowledge recommendation ?, and cognitive intelligent manufacturing ?. By modeling complex industrial processes and components, KGE enhances predictive maintenance, optimizes supply chains, and facilitates intelligent decision-making, leading to increased efficiency and reduced downtime.

\* **Ethical AI and Trust:** The emphasis on explainability ?? and fairness ? in KGE research is crucial for building trustworthy AI systems. As KGE powers critical applications, ensuring that its decisions are transparent, unbiased, and justifiable is paramount for societal acceptance and regulatory compliance.

\* **Knowledge Democratization:** KGE facilitates the democratization of complex knowledge by enabling intuitive natural language querying ?? and personalized access to information. This makes vast knowledge bases accessible to non-technical users, fostering informed decision-making across various domains, from academic search ? to urban planning ?.

In conclusion, the field of KGE is at an exciting juncture, continually pushing the boundaries of knowledge representation. The shift towards hybrid models, advanced geometric embeddings, and privacy-preserving techniques, coupled with a strong focus on explainability and ethical considerations, ensures that KGE will remain a pivotal technol-

ogy. Its profound and expanding societal impact across healthcare, education, industry, and beyond underscores its role as a fundamental enabler of the next generation of intelligent, transparent, and responsible AI systems.

## References

## References

- Zequn Sun, Wei Hu, Qingheng Zhang, et al. (2018). *Bootstrapping Entity Alignment with Knowledge Graph Embedding*. International Joint Conference on Artificial Intelligence.
- 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.
- Mingyang Chen, Wen Zhang, Zhen Yao, et al. (2023). *Entity-Agnostic Representation Learning for Parameter-Efficient Knowledge Graph Embedding*. AAAI Conference on Artificial Intelligence.
- Yang Yang, Chubing Zhang, Xin Song, et al. (2023). *Contextualized Knowledge Graph Embedding for Explainable Talent Training Course Recommendation*. ACM Trans. Inf. Syst..
- Yantao Jia, Yuanzhuo Wang, Hailun Lin, et al. (2015). *Locally Adaptive Translation for Knowledge Graph Embedding*. AAAI Conference on Artificial Intelligence.
- Oliver Lloyd, Yi Liu, and T. Gaunt (2022). *Assessing the effects of hyperparameters on knowledge graph embedding quality*. Journal of Big Data.
- 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.
- Chengjin Xu, M. Nanyeri, Fouad Alkhoury, et al. (2019). *Temporal Knowledge Graph Embedding Model based on Additive Time Series Decomposition*. arXiv.org.
- Yingchun Shan, Chenyang Bu, Xiaojian Liu, et al. (2018). *Confidence-Aware Negative Sampling Method for Noisy Knowledge Graph Embedding*. International Conference on Big Knowledge.

- Zhuoxun Zheng, Baifan Zhou, Hui Yang, et al. (2024). *Knowledge graph embedding closed under composition*. Data mining and knowledge discovery.
- Peng He, Gang Zhou, Yao Yao, et al. (2023). *A type-augmented knowledge graph embedding framework for knowledge graph completion*. Scientific Reports.
- Han Xiao, Minlie Huang, and Xiaoyan Zhu (2015). *TransG : A Generative Model for Knowledge Graph Embedding*. Annual Meeting of the Association for Computational Linguistics.
- Shu Guo, Quan Wang, Lihong Wang, et al. (2017). *Knowledge Graph Embedding with Iterative Guidance from Soft Rules*. AAAI Conference on Artificial Intelligence.
- 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.
- 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.
- Xiaochi Zhou, Shaocong Zhang, Mehal Agarwal, et al. (2023). *Marie and BERT—A Knowledge Graph Embedding Based Question Answering System for Chemistry*. ACS Omega.
- Yuejia Xiang, Ziheng Zhang, Jiaoyan Chen, et al. (2021). *OntoEA: Ontology-guided Entity Alignment via Joint Knowledge Graph Embedding*. Findings.
- Jiahang Cao, Jinyuan Fang, Zaiqiao Meng, et al. (2022). *Knowledge Graph Embedding: A Survey from the Perspective of Representation Spaces*. ACM Computing Surveys.
- Peng Wang, Jing Zhou, Yuzhang Liu, et al. (2021). *TransET: Knowledge Graph Embedding with Entity Types*. Electronics.



- Shu Guo, Lin Li, Zhen Hui, et al. (2020). *Knowledge Graph Embedding Preserving Soft Logical Regularity*. International Conference on Information and Knowledge Management.
- Xiaoxiong Zhang, Zhiwei Zeng, Xin Zhou, et al. (2024). *Communication-Efficient Federated Knowledge Graph Embedding with Entity-Wise Top-K Sparsification*. Knowledge-Based Systems.
- Jianhao Shen, Chenguang Wang, Linyuan Gong, et al. (2022). *Joint Language Semantic and Structure Embedding for Knowledge Graph Completion*. International Conference on Computational Linguistics.
- 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.
- 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.
- Yongqi Zhang, Quanming Yao, Wenyan Dai, et al. (2019). *AutoSF: Searching Scoring Functions for Knowledge Graph Embedding*. IEEE International Conference on Data Engineering.
- 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.
- 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.
- Jiapu Wang, Boyue Wang, Junbin Gao, et al. (2024). *MADE: Multicurvature Adaptive Embedding for Temporal Knowledge Graph Completion*. IEEE Transactions on Cybernetics.

- Han Xiao, Yidong Chen, and X. Shi (2019). *Knowledge Graph Embedding Based on Multi-View Clustering Framework*. IEEE Transactions on Knowledge and Data Engineering.
- Mrinmaya Sachan (2020). *Knowledge Graph Embedding Compression*. Annual Meeting of the Association for Computational Linguistics.
- Tiroshan Madushanka, and R. Ichise (2024). *Negative Sampling in Knowledge Graph Representation Learning: A Review*. arXiv.org.
- Chaoyu Zhu, Zhihao Yang, Xiaoqiong Xia, et al. (2022). *Multimodal reasoning based on knowledge graph embedding for specific diseases*. Bioinform..
- Qiuyu Liang, Weihua Wang, F. Bao, et al. (2024). *Fully Hyperbolic Rotation for Knowledge Graph Embedding*. European Conference on Artificial Intelligence.
- 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.
- Takuma Ebisu, and R. Ichise (2017). *TorusE: Knowledge Graph Embedding on a Lie Group*. AAAI Conference on Artificial Intelligence.
- 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.
- Xiao Huang, Jingyuan Zhang, Dingcheng Li, et al. (2019). *Knowledge Graph Embedding Based Question Answering*. Web Search and Data Mining.
- 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.
- Zhu Sun, Jie Yang, Jie Zhang, et al. (2018). *Recurrent knowledge graph embedding for effective recommendation*. ACM Conference on Recommender Systems.

- Xiou Ge, Yun Cheng Wang, Bin Wang, et al. (2023). *Knowledge Graph Embedding: An Overview*. APSIPA Transactions on Signal and Information Processing.
- Rui Wang, Bicheng Li, Shengwei Hu, et al. (2020). *Knowledge Graph Embedding via Graph Attenuated Attention Networks*. IEEE Access.
- Rui Li, Jianan Zhao, Chaozhuo Li, et al. (2022). *HousE: Knowledge Graph Embedding with Householder Parameterization*. International Conference on Machine Learning.
- Qingheng Zhang, Zequn Sun, Wei Hu, et al. (2019). *Multi-view Knowledge Graph Embedding for Entity Alignment*. International Joint Conference on Artificial Intelligence.
- Xiaojuan Tang, Song-Chun Zhu, Yitao Liang, et al. (2022). *RuleE: Knowledge Graph Reasoning with Rule Embedding*. Annual Meeting of the Association for Computational Linguistics.
- 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.
- 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..
- Jing Qian, Gangmin Li, Katie Atkinson, et al. (2021). *Understanding Negative Sampling in Knowledge Graph Embedding*. International Journal of Artificial Intelligence Applications.
- Yuanfei Dai, Shiping Wang, N. Xiong, et al. (2020). *A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks*. Electronics.
- 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.

- Qi Yan, Jiaxin Fan, Mohan Li, et al. (2022). *A Survey on Knowledge Graph Embedding*. International Conference on Data Science in Cyberspace.
- 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.
- 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.
- 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.
- Quan Wang, Pingping Huang, Haifeng Wang, et al. (2019). *CoKE: Contextualized Knowledge Graph Embedding*. arXiv.org.
- Shimin Di, and Lei Chen (2023). *Message Function Search for Knowledge Graph Embedding*. The Web Conference.
- Yantao Jia, Yuanzhuo Wang, Xiaolong Jin, et al. (2017). *Knowledge Graph Embedding*. ACM Transactions on the Web.
- Shivani Choudhary, Tarun Luthra, Ashima Mittal, et al. (2021). *A Survey of Knowledge Graph Embedding and Their Applications*. arXiv.org.
- 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.
- 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.

- Zhen Wang, Jianwen Zhang, Jianlin Feng, et al. (2014). *Knowledge Graph Embedding by Translating on Hyperplanes*. AAAI Conference on Artificial Intelligence.
- 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.
- 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.
- Sameh K. Mohamed, A. Nounu, and V. Nováček (2020). *Biological applications of knowledge graph embedding models*. Briefings Bioinform..
- 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.
- 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.
- 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.
- Xiaoxiong Zhang, Zhiwei Zeng, Xin Zhou, et al. (2024). *Personalized Federated Knowledge Graph Embedding with Client-Wise Relation Graph*. Applied intelligence (Boston).
- Paolo Rosso, Dingqi Yang, and P. Cudré-Mauroux (2020). *Beyond Triplets: Hyper-Relational Knowledge Graph Embedding for Link Prediction*. The Web Conference.
- Enyuan Zhou, Song Guo, Zhixiu Ma, et al. (2024). *Poisoning Attack on Federated Knowledge Graph Embedding*. The Web Conference.

- 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.
- Tengwei Song, Jie Luo, and Lei Huang (2021). *Rot-Pro: Modeling Transitivity by Projection in Knowledge Graph Embedding*. Neural Information Processing Systems.
- 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.
- Xiou Ge, Yun Cheng Wang, Bin Wang, et al. (2022). *CompoundE: Knowledge Graph Embedding with Translation, Rotation and Scaling Compound Operations*. arXiv.org.
- Feiliang Ren, Jucheng Li, Huihui Zhang, et al. (2020). *Knowledge Graph Embedding with Atrous Convolution and Residual Learning*. International Conference on Computational Linguistics.
- Jun Yuan, Neng Gao, and Ji Xiang (2019). *TransGate: Knowledge Graph Embedding with Shared Gate Structure*. AAAI Conference on Artificial Intelligence.
- Han Xiao, Minlie Huang, Yu Hao, et al. (2015). *TransA: An Adaptive Approach for Knowledge Graph Embedding*. arXiv.org.
- 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.
- Guoliang Ji, Shizhu He, Liheng Xu, et al. (2015). *Knowledge Graph Embedding via Dynamic Mapping Matrix*. Annual Meeting of the Association for Computational Linguistics.
- Lifan Lin, and Kun She (2020). *Tensor Decomposition-Based Temporal Knowledge Graph Embedding*. IEEE International Conference on Tools with Artificial Intelligence.

- 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.
- 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.
- 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.
- N. Fanourakis, Vasilis Efthymiou, D. Kotzinos, et al. (2022). *Knowledge graph embedding methods for entity alignment: experimental review*. Data mining and knowledge discovery.
- Peifeng Wang, Jialong Han, Chenliang Li, et al. (2018). *Logic Attention Based Neighborhood Aggregation for Inductive Knowledge Graph Embedding*. AAAI Conference on Artificial Intelligence.
- Pedro Tabacof, and Luca Costabello (2019). *Probability Calibration for Knowledge Graph Embedding Models*. International Conference on Learning Representations.
- 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.
- 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.
- Zelong Li, Jianchao Ji, Zuohui Fu, et al. (2021). *Efficient Non-Sampling Knowledge Graph Embedding*. The Web Conference.

- Guangtong Li, L. Siddharth, and Jianxi Luo (2022). *Embedding knowledge graph of patent metadata to measure knowledge proximity*. J. Assoc. Inf. Sci. Technol..
- Boyang Ding, Quan Wang, Bin Wang, et al. (2018). *Improving Knowledge Graph Embedding Using Simple Constraints*. Annual Meeting of the Association for Computational Linguistics.
- 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.
- 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).
- 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.
- 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.
- 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.
- 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.
- Shu Guo, Quan Wang, Bin Wang, et al. (2015). *Semantically Smooth Knowledge Graph Embedding*. Annual Meeting of the Association for Computational Linguistics.



- Chengjin Xu, M. Nayyeri, Fouad Alkhoury, et al. (2020). *TeRo: A Time-aware Knowledge Graph Embedding via Temporal Rotation*. International Conference on Computational Linguistics.
- Chenguang Zheng, Guanxian Jiang, Xiao Yan, et al. (2024). *GE2: A General and Efficient Knowledge Graph Embedding Learning System*. Proc. ACM Manag. Data.
- Zhao Zhang, Fuzhen Zhuang, Meng Qu, et al. (2018). *Knowledge Graph Embedding with Hierarchical Relation Structure*. Conference on Empirical Methods in Natural Language Processing.
- Beibei Zhu, Ruolin Wang, Junyi Wang, et al. (2024). *A survey: knowledge graph entity alignment research based on graph embedding*. Artificial Intelligence Review.
- 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.
- 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.
- 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.
- A. Sadeghian, Mohammadreza Armandpour, Anthony Colas, et al. (2021). *ChronoR: Rotation Based Temporal Knowledge Graph Embedding*. AAAI Conference on Artificial Intelligence.
- Jiajun Liu, Wenjun Ke, Peng Wang, et al. (2024). *Fast and Continual Knowledge Graph Embedding via Incremental LoRA*. International Joint Conference on Artificial Intelligence.
- 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.
- Andrea Rossi, D. Firmani, Antonio Martinata, et al. (2020). *Knowledge Graph Embedding for Link Prediction*. ACM Transactions on Knowledge Discovery from Data.
- 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.
- Yanhui Peng, and Jing Zhang (2020). *LineaRE: Simple but Powerful Knowledge Graph Embedding for Link Prediction*. Industrial Conference on Data Mining.
- Hao Ji, Li Yan, and Z. Ma (2024). *Multihop Fuzzy Spatiotemporal RDF Knowledge Graph Query via Quaternion Embedding*. IEEE transactions on fuzzy systems.
- 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.
- Adrian Kochsiek (2021). *Parallel Training of Knowledge Graph Embedding Models: A Comparison of Techniques*. Proceedings of the VLDB Endowment.
- 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.
- 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.
- 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.

- Cosimo Gregucci, M. Nayyeri, D. Hern'andez, et al. (2023). *Link Prediction with Attention Applied on Multiple Knowledge Graph Embedding Models*. The Web Conference.
- Zhe Pan, and Peng Wang (2021). *Hyperbolic Hierarchy-Aware Knowledge Graph Embedding for Link Prediction*. Conference on Empirical Methods in Natural Language Processing.
- 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.
- Rui Li, Chaozhuo Li, Yanming Shen, et al. (2024). *Generalizing Knowledge Graph Embedding with Universal Orthogonal Parameterization*. International Conference on Machine Learning.
- Chenyan Xiong, Russell Power, and Jamie Callan (2017). *Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding*. The Web Conference.
- Fan Gong, Meng Wang, Haofen Wang, et al. (2020). *SMR: Medical Knowledge Graph Embedding for Safe Medicine Recommendation*. Big Data Research.
- 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.
- 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.
- Zhehui Zhou, Can Wang, Yan Feng, et al. (2022). *JointE: Jointly utilizing 1D and 2D convolution for knowledge graph embedding*. Knowledge-Based Systems.
- Da Xu, Chuanwei Ruan, Evren Körpeoglu, et al. (2019). *Product Knowledge Graph Embedding for E-commerce*. Web Search and Data Mining.

- Haithem Mezni, D. Benslimane, and Ladjel Bellatreche (2021). *Context-Aware Service Recommendation Based on Knowledge Graph Embedding*. IEEE Transactions on Knowledge and Data Engineering.
- 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).
- 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.
- 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.
- 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.
- Niannian Guan, Dandan Song, and L. Liao (2019). *Knowledge graph embedding with concepts*. Knowledge-Based Systems.
- 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.
- Hengtong Zhang, T. Zheng, Jing Gao, et al. (2019). *Data Poisoning Attack against Knowledge Graph Embedding*. International Joint Conference on Artificial Intelligence.
- 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.
- Xin Wang, Shengfei Lyu, Xiangyu Wang, et al. (2022). *Temporal knowledge graph embedding via sparse transfer matrix*. Information Sciences.

- Mingyang Chen, Wen Zhang, Zonggang Yuan, et al. (2022). *Federated knowledge graph completion via embedding-contrastive learning*. Knowledge-Based Systems.
- 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.
- Haichuan Fang, Youwei Wang, Zhen Tian, et al. (2022). *Learning knowledge graph embedding with a dual-attention embedding network*. Expert systems with applications.
- 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.
- Xiao Sha, Zhu Sun, and Jie Zhang (2019). *Hierarchical attentive knowledge graph embedding for personalized recommendation*. Electronic Commerce Research and Applications.
- Zhifei Li, Hai Liu, Zhaoli Zhang, et al. (2021). *Recalibration convolutional networks for learning interaction knowledge graph embedding*. Neurocomputing.
- Han Xiao, Minlie Huang, Yu Hao, et al. (2015). *TransG : A Generative Mixture Model for Knowledge Graph Embedding*. arXiv.org.
- 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.
- 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.
- 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.

- 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).
- 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.
- 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.
- 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.
- Jun Feng, Minlie Huang, Mingdong Wang, et al. (2016). *Knowledge Graph Embedding by Flexible Translation*. International Conference on Principles of Knowledge Representation and Reasoning.
- 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.
- Shengtian Sang, Zhihao Yang, Xiaoxia Liu, et al. (2019). *GrEDeL: A Knowledge Graph Embedding Based Method for Drug Discovery From Biomedical Literatures*. IEEE Access.
- M. Wang, Mengyue Liu, Jun Liu, et al. (2017). *Safe Medicine Recommendation via Medical Knowledge Graph Embedding*. arXiv.org.
- Dan Jiang, Ronggui Wang, Juan Yang, et al. (2021). *Kernel multi-attention neural network for knowledge graph embedding*. Knowledge-Based Systems.

- 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.
- Nasrullah Khan, Zongmin Ma, Aman Ullah, et al. (2022). *Similarity attributed knowledge graph embedding enhancement for item recommendation*. Information Sciences.
- Haithem Mezni (2021). *Temporal Knowledge Graph Embedding for Effective Service Recommendation*. IEEE Transactions on Services Computing.
- Qianjin Zhang, Ronggui Wang, Juan Yang, et al. (2021). *Structural context-based knowledge graph embedding for link prediction*. Neurocomputing.
- Xuqian Huang, Jiuyang Tang, Zhen Tan, et al. (2021). *Knowledge graph embedding by relational and entity rotation*. Knowledge-Based Systems.
- Aleksandar Pavlovic, and Emanuel Sallinger (2022). *ExpressivE: A Spatio-Functional Embedding For Knowledge Graph Completion*. International Conference on Learning Representations.
- Shensi Wang, Kun Fu, Xian Sun, et al. (2021). *Hierarchical-aware relation rotational knowledge graph embedding for link prediction*. Neurocomputing.
- Shuai Zhang, Yi Tay, Lina Yao, et al. (2019). *Quaternion Knowledge Graph Embedding*. arXiv.org.
- Gengchen Mai, Bo Yan, K. Janowicz, et al. (2019). *Relaxing Unanswerable Geographic Questions Using A Spatially Explicit Knowledge Graph Embedding Model*. Agile Conference.
- 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.
- Feiyang Wang, Zhongbao Zhang, Li Sun, et al. (2022). *DiriE: Knowledge Graph Embedding with Dirichlet Distribution*. The Web Conference.

- Ilaria Ferrari, Giacomo Frisoni, Paolo Italiani, et al. (2022). *Comprehensive Analysis of Knowledge Graph Embedding Techniques Benchmarked on Link Prediction*. Electronics.
- Guirong Fu, Zhao Meng, Zhen Han, et al. (2022). *TempCaps: A Capsule Network-based Embedding Model for Temporal Knowledge Graph Completion*. SPNLP.
- Yanrong Wu, and Zhichun Wang (2018). *Knowledge Graph Embedding with Numeric Attributes of Entities*. Rep4NLP@ACL.
- Qianjin Zhang, Ronggui Wang, Juan Yang, et al. (2021). *Knowledge graph embedding by reflection transformation*. Knowledge-Based Systems.
- Sameh K. Mohamed, V. Nováček, P. Vandenbussche, et al. (2019). *Loss Functions in Knowledge Graph Embedding Models*. DL4KG@ESWC.
- 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.
- Binling Nie, and Shouqian Sun (2019). *Knowledge graph embedding via reasoning over entities, relations, and text*. Future generations computer systems.
- 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.
- 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.
- Chen Li, Xutan Peng, Yuhang Niu, et al. (2021). *Learning graph attention-aware knowledge graph embedding*. Neurocomputing.
- S. Yu, Sujit Rokka Chhetri, A. Canedo, et al. (2019). *Pykg2vec: A Python Library for Knowledge Graph Embedding*. Journal of machine learning research.



- Bahare Fatemi, Siamak Ravanbakhsh, and D. Poole (2018). *Improved Knowledge Graph Embedding using Background Taxonomic Information*. AAAI Conference on Artificial Intelligence.
- 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.
- 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.
- Fengyuan Lu, Peijin Cong, and Xinli Huang (2020). *Utilizing Textual Information in Knowledge Graph Embedding: A Survey of Methods and Applications*. IEEE Access.
- Weidong Li, Rong Peng, and Zhi Li (2022). *Improving knowledge graph completion via increasing embedding interactions*. Applied intelligence (Boston).
- Yuanfei Luo, Quan Wang, Bin Wang, et al. (2015). *Context-Dependent Knowledge Graph Embedding*. Conference on Empirical Methods in Natural Language Processing.
- 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.
- Feng Zhao, Haoran Sun, Langjunqing Jin, et al. (2020). *Structure-augmented knowledge graph embedding for sparse data with rule learning*. Computer Communications.
- Yantao Jia, Yuanzhuo Wang, Xiaolong Jin, et al. (2018). *Path-specific knowledge graph embedding*. Knowledge-Based Systems.
- Gengchen Mai, K. Janowicz, and Bo Yan (2018). *Combining Text Embedding and Knowledge Graph Embedding Techniques for Academic Search Engines*. Semdeep/NLIWoD@ISWC.

- 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.
- Xiaoli Tang, Rui Yuan, Qianyu Li, et al. (2020). *Timespan-Aware Dynamic Knowledge Graph Embedding by Incorporating Temporal Evolution*. IEEE Access.
- Lingbing Guo, Qiang Zhang, Zequn Sun, et al. (2022). *Understanding and Improving Knowledge Graph Embedding for Entity Alignment*. International Conference on Machine Learning.
- Dan Jiang, Ronggui Wang, Lixia Xue, et al. (2022). *Multiview feature augmented neural network for knowledge graph embedding*. Knowledge-Based Systems.
- Yu Liu, Wen Hua, Kexuan Xin, et al. (2019). *Context-Aware Temporal Knowledge Graph Embedding*. WISE.
- Qianjin Zhang, Ronggui Wang, Juan Yang, et al. (2020). *Knowledge graph embedding by translating in time domain space for link prediction*. Knowledge-Based Systems.
- Liang Chang, Manli Zhu, T. Gu, et al. (2017). *Knowledge Graph Embedding by Dynamic Translation*. IEEE Access.
- 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.
- Yongqi Zhang, Zhanke Zhou, Quanming Yao, et al. (2022). *Efficient Hyper-parameter Search for Knowledge Graph Embedding*. Annual Meeting of the Association for Computational Linguistics.
- 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).

- 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.
- László Grad-Gyenge, A. Kiss, and P. Filzmoser (2017). *Graph Embedding Based Recommendation Techniques on the Knowledge Graph*. User Modeling, Adaptation, and Personalization.
- Xiaofei Zhou, Lingfeng Niu, Qiannan Zhu, et al. (2021). *Knowledge Graph Embedding by Double Limit Scoring Loss*. IEEE Transactions on Knowledge and Data Engineering.
- Yao Chen, Jiangang Liu, Zhe Zhang, et al. (2021). *MöbiusE: Knowledge Graph Embedding on Möbius Ring*. Knowledge-Based Systems.
- Yongqi Zhang, Quanming Yao, and Lei Chen (2020). *Interstellar: Searching Recurrent Architecture for Knowledge Graph Embedding*. Neural Information Processing Systems.
- Armand Boschini (2020). *TorchKGE: Knowledge Graph Embedding in Python and PyTorch*. arXiv.org.
- P. Wang, D. Dou, Fangzhao Wu, et al. (2019). *Logic Rules Powered Knowledge Graph Embedding*. arXiv.org.
- E. B. Myklebust, Ernesto Jiménez-Ruiz, Jiaoyan Chen, et al. (2019). *Knowledge Graph Embedding for Ecotoxicological Effect Prediction*. International Workshop on the Semantic Web.
- Miriyala Kartheek, and G. Sajeesh (2021). *Building Semantic Based Recommender System Using Knowledge Graph Embedding*. International Conference on Intelligent Information Processing.
- Xiao Sha, Zhu Sun, and Jie Zhang (2019). *Attentive Knowledge Graph Embedding for Personalized Recommendation*. arXiv.org.
- Haonan Lu, and Hailin Hu (2020). *DensE: An Enhanced Non-Abelian Group Representation for Knowledge Graph Embedding*. arXiv.org.

- Siheng Zhang, Zhengya Sun, and Wensheng Zhang (2020). *Improve the translational distance models for knowledge graph embedding*. Journal of Intelligence and Information Systems.
- Mingda Li, Zhengya Sun, Siheng Zhang, et al. (2020). *Enhancing Knowledge Graph Embedding with Relational Constraints*. 2020 IEEE International Conference on Knowledge Graph (ICKG).
- Jian Li, Zhuoming Xu, Yan Tang, et al. (2020). *Deep Hybrid Knowledge Graph Embedding for Top-N Recommendation*. Web Information System and Application Conference.
- Kuekyeng Kim, Yuna Hur, Gyeongmin Kim, et al. (2020). *GREG: A Global Level Relation Extraction with Knowledge Graph Embedding*. Applied Sciences.
- Jizhao Zhu, Yantao Jia, Jun Xu, et al. (2018). *Modeling the Correlations of Relations for Knowledge Graph Embedding*. Journal of Computational Science and Technology.
- Kien Do, T. Tran, and S. Venkatesh (2018). *Knowledge Graph Embedding with Multiple Relation Projections*. International Conference on Pattern Recognition.
- Yunpu Ma, Volker Tresp, Liming Zhao, et al. (2019). *Variational Quantum Circuit Model for Knowledge Graph Embedding*. Advanced Quantum Technologies.
- Yuhang Zhang, Jun Wang, and Jie Luo (2020). *Knowledge Graph Embedding Based Collaborative Filtering*. IEEE Access.
- 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.
- Zhihao Wang, and Xin Li (2019). *Hybrid-TE: Hybrid Translation-Based Temporal Knowledge Graph Embedding*. IEEE International Conference on Tools with Artificial Intelligence.
- Hung Nghiep Tran, and A. Takasu (2019). *Analyzing Knowledge Graph Embedding Methods from a Multi-Embedding Interaction Perspective*. EDBT/ICDT Workshops.

- Shengwu Xiong, Weitao Huang, and P. Duan (2018). *Knowledge Graph Embedding via Relation Paths and Dynamic Mapping Matrix*. ER Workshops.
- 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.
- Ling Zhao, Hanhan Deng, L. Qiu, et al. (2020). *Urban Multi-Source Spatio-Temporal Data Analysis Aware Knowledge Graph Embedding*. Symmetry.
- Maoyuan Zhang, Qi Wang, Wukui Xu, et al. (2018). *Discriminative Path-Based Knowledge Graph Embedding for Precise Link Prediction*. European Conference on Information Retrieval.
- Ningning Jia, Xiang Cheng, and Sen Su (2020). *Improving Knowledge Graph Embedding Using Locally and Globally Attentive Relation Paths*. European Conference on Information Retrieval.
- Qiannan Zhu, Xiaofei Zhou, P. Zhang, et al. (2019). *A neural translating general hyperplane for knowledge graph embedding*. Journal of Computer Science.
- 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).
- 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.
- Nasrullah Sheikh, Xiao Qin, B. Reinwald, et al. (2021). *Knowledge Graph Embedding using Graph Convolutional Networks with Relation-Aware Attention*. arXiv.org.
- 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.

- Chunhong Zhang, Miao Zhou, Xiao Han, et al. (2017). *Knowledge Graph Embedding for Hyper-Relational Data*. Unpublished manuscript.
- 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.
- Martina Garofalo, Maria Angela Pellegrino, Abdulrahman Altabba, et al. (2018). *Leveraging Knowledge Graph Embedding Techniques for Industry 4.0 Use Cases*. arXiv.org.
- Kai Wang, Yu Liu, Xiujuan Xu, et al. (2018). *Knowledge Graph Embedding with Entity Neighbors and Deep Memory Network*. arXiv.org.
- 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.
- 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.
- Jun Shi, Huan Gao, G. Qi, et al. (2017). *Knowledge Graph Embedding with Triple Context*. International Conference on Information and Knowledge Management.
- Wen Zhang (2017). *Knowledge Graph Embedding with Diversity of Structures*. The Web Conference.
- 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.
- Natthawut Kertkeidkachorn, Xin Liu, and R. Ichise (2019). *GTransE: Generalizing Translation-Based Model on Uncertain Knowledge Graph Embedding*. JSAI.
- Jia Zhu, Zetao Zheng, Min Yang, et al. (2019). *A semi-supervised model for knowledge graph embedding*. Data mining and knowledge discovery.

- Hengtong Zhang, T. Zheng, Jing Gao, et al. (2019). *Towards Data Poisoning Attack against Knowledge Graph Embedding*. arXiv.org.
- 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.
- Vibhor Kanojia, Hideyuki Maeda, Riku Togashi, et al. (2017). *Enhancing Knowledge Graph Embedding with Probabilistic Negative Sampling*. The Web Conference.
- Huan Gao, Jun Shi, G. Qi, et al. (2018). *Triple Context-Based Knowledge Graph Embedding*. IEEE Access.
- 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.
- Han Xiao, Minlie Huang, and Xiaoyan Zhu (2016). *Knowledge Semantic Representation: A Generative Model for Interpretable Knowledge Graph Embedding*. arXiv.org.
- 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.
- Jin-cheng Zhang, A. Zain, Kai Zhou, et al. (2024). *A review of recommender systems based on knowledge graph embedding*. Expert systems with applications.
- 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.
- Xiangrong Zhu, Guang-pu Li, and Wei Hu (2023). *Heterogeneous Federated Knowledge Graph Embedding Learning and Unlearning*. The Web Conference.
- Jiajun Liu, Wenjun Ke, Peng Wang, et al. (2024). *Towards Continual Knowledge Graph Embedding via Incremental Distillation*. AAAI Conference on Artificial Intelligence.

- 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.
- Duantengchuan Li, Tao Xia, Jing Wang, et al. (2024). *SDFormer: A shallow-to-deep feature interaction for knowledge graph embedding*. Knowledge-Based Systems.
- Jaejun Lee, Chanyoung Chung, and Joyce Jiyoung Whang (2023). *InGram: Inductive Knowledge Graph Embedding via Relation Graphs*. International Conference on Machine Learning.
- 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.
- 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.
- 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).
- Zengcan Xue, Zhao Zhang, Hai Liu, et al. (2023). *Learning knowledge graph embedding with multi-granularity relational augmentation network*. Expert systems with applications.
- 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..
- Zhen Chen, Dalin Zhang, Shanshan Feng, et al. (2024). *KGTS: Contrastive Trajectory Similarity Learning over Prompt Knowledge Graph Embedding*. AAAI Conference on Artificial Intelligence.



- Jia Zhu, Changqin Huang, and P. D. Meo (2022). *DFMKE: A dual fusion multi-modal knowledge graph embedding framework for entity alignment*. Information Fusion.
- 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.
- Harry Shomer, Wei Jin, Wentao Wang, et al. (2023). *Toward Degree Bias in Embedding-Based Knowledge Graph Completion*. The Web Conference.
- Mingjie Wang, Zijie Li, Jun Wang, et al. (2024). *TracKGE: Transformer with Relation-pattern Adaptive Contrastive Learning for Knowledge Graph Embedding*. Knowledge-Based Systems.
- Zhifei Li, Wei Huang, Xuchao Gong, et al. (2024). *Decoupled semantic graph neural network for knowledge graph embedding*. Neurocomputing.
- Mingqi Li, Wenming Ma, and Zihao Chu (2024). *KGIE: Knowledge graph convolutional network for recommender system with interactive embedding*. Knowledge-Based Systems.
- Yan Jia, Mengqi Lin, Yechen Wang, et al. (2023). *Extrapolation over temporal knowledge graph via hyperbolic embedding*. CAAI Transactions on Intelligence Technology.
- Wei Huang, Jia Liu, Tianrui Li, et al. (2023). *FedCKE: Cross-Domain Knowledge Graph Embedding in Federated Learning*. IEEE Transactions on Big Data.
- Ruoxin Wang, and C. F. Cheung (2023). *Knowledge graph embedding learning system for defect diagnosis in additive manufacturing*. Computers in industry (Print).
- 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.
- Dan Jiang, Ronggui Wang, Lixia Xue, et al. (2023). *Multisource hierarchical neural network for knowledge graph embedding*. Expert systems with applications.

- H. Lu, Hailin Hu, and Xiaodong Lin (2022). *DensE: An enhanced non-commutative representation for knowledge graph embedding with adaptive semantic hierarchy*. Neurocomputing.
- 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.
- Yuchao Zhang, Xiangjie Kong, Zhehui Shen, et al. (2024). *A survey on temporal knowledge graph embedding: Models and applications*. Knowledge-Based Systems.
- 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.
- Zhen Yao, Wen Zhang, Mingyang Chen, et al. (2023). *Analogical Inference Enhanced Knowledge Graph Embedding*. AAAI Conference on Artificial Intelligence.
- Zhipeng Li, Shanshan Feng, Jun Shi, et al. (2023). *Future Event Prediction Based on Temporal Knowledge Graph Embedding*. Computer systems science and engineering.
- Shihan Yang, Weiya Zhang, R. Tang, et al. (2022). *Approximate inferring with confidence predicting based on uncertain knowledge graph embedding*. Information Sciences.
- 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.
- Yuke Hu, Wei Liang, Ruofan Wu, et al. (2023). *Quantifying and Defending against Privacy Threats on Federated Knowledge Graph Embedding*. The Web Conference.
- Daiyi Li, Li Yan, Xiaowen Zhang, et al. (2023). *EventKGE: Event knowledge graph embedding with event causal transfer*. Knowledge-Based Systems.
- Xinkun Hao, Qingfeng Chen, Haiming Pan, et al. (2022). *Enhancing drug–drug interaction prediction by three-way decision and knowledge graph embedding*. Granular Computing.

- 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).
- Thanh-Binh Le, Ngoc Huynh, and Bac Le (2022). *Knowledge graph embedding by projection and rotation on hyperplanes for link prediction*. Applied intelligence (Boston).
- Shuang Liang (2023). *Knowledge Graph Embedding Based on Graph Neural Network*. IEEE International Conference on Data Engineering.
- 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.
- Peng He, Gang Zhou, Mengli Zhang, et al. (2022). *Improving temporal knowledge graph embedding using tensor factorization*. Applied intelligence (Boston).
- Linshan Shen, Rongbo He, and Shaobin Huang (2022). *Entity alignment with adaptive margin learning knowledge graph embedding*. Data Knowledge Engineering.
- 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.
- Guanglin Niu, Bo Li, Yongfei Zhang, et al. (2020). *AutoETER: Automated Entity Type Representation with Relation-Aware Attention for Knowledge Graph Embedding*. Findings.
- 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).
- Jiayi Li, and Yujiu Yang (2022). *STaR: Knowledge Graph Embedding by Scaling, Translation and Rotation*. Autonomous Infrastructure, Management and Security.

- 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.
- 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.
- 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.
- 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.
- 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.
- Franziska Krause (2022). *Dynamic Knowledge Graph Embeddings via Local Embedding Reconstructions*. Extended Semantic Web Conference.
- Zhao Zhang, Fuzhen Zhuang, Meng Qu, et al. (2021). *Knowledge graph embedding with shared latent semantic units*. Neural Networks.
- Guang-pu Li, Zequn Sun, Lei Qian, et al. (2021). *Rule-based data augmentation for knowledge graph embedding*. AI Open.
- Kai Wang, Yu Liu, Xiujuan Xu, et al. (2020). *Enhancing knowledge graph embedding by composite neighbors for link prediction*. Computing.
- 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).

- Yongqi Zhang, Quanming Yao, and Lei Chen (2021). *Simple and automated negative sampling for knowledge graph embedding*. The VLDB journal.
- Nasrullah Sheikh, Xiao Qin, B. Reinwald, et al. (2022). *Scaling knowledge graph embedding models for link prediction*. EuroMLSys@EuroSys.
- 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.
- Victoria Eyharabide, I. E. I. Bekkouch, and Nicolae Dragoş Constantin (2021). *Knowledge Graph Embedding-Based Domain Adaptation for Musical Instrument Recognition*. De Computis.
- 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).
- Yan Huang, Haili Sun, Xu Ke, et al. (2020). *CoRelatE: Learning the correlation in multi-fold relations for knowledge graph embedding*. Knowledge-Based Systems.
- M. Kurokawa (2021). *Explainable Knowledge Reasoning Framework Using Multiple Knowledge Graph Embedding*. IJCKG.
- Sameh K. Mohamed, Emir Muñoz, and V. Nováček (2021). *On Training Knowledge Graph Embedding Models*. Inf..
- 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.
- 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.

- 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.
- 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..
- Mingsheng He, Lin Zhu, and Luyi Bai (2024). *ConvTKG: A query-aware convolutional neural network-based embedding model for temporal knowledge graph completion*. Neurocomputing.
- Dong Zhang, Zhe Rong, Chengyuan Xue, et al. (2024). *SimRE: Simple contrastive learning with soft logical rule for knowledge graph embedding*. Information Sciences.
- 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.
- Yanzhen Jing, Guanghui Zhou, Chao Zhang, et al. (2024). *XMKR: Explainable manufacturing knowledge recommendation for collaborative design with graph embedding learning*. Advanced Engineering Informatics.
- Pengcheng Jiang, Lang Cao, Cao Xiao, et al. (2024). *KG-FIT: Knowledge Graph Fine-Tuning Upon Open-World Knowledge*. Neural Information Processing Systems.
- 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.
- Huafeng Quan, Yiting Li, Dashuai Liu, et al. (2024). *Protection of Guizhou Miao batik culture based on knowledge graph and deep learning*. Heritage Science.

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.

- 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.
- Ming Lu, Yancong Li, Jiangxiao Zhang, et al. (2024). *Deep hyperbolic convolutional model for knowledge graph embedding*. Knowledge-Based Systems.
- 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..
- Nasrullah Khan, Zongmin Ma, Ruizhe Ma, et al. (2024). *Continual knowledge graph embedding enhancement for joint interaction-based next click recommendation*. Knowledge-Based Systems.
- Zengcan Xue, Zhaoli Zhang, Hai Liu, et al. (2025). *MHRN: A multi-perspective hierarchical relation network for knowledge graph embedding*. Knowledge-Based Systems.
- Xiao Long, Liansheng Zhuang, Aodi Li, et al. (2024). *Fact Embedding through Diffusion Model for Knowledge Graph Completion*. The Web Conference.
- Chen Huang, Fei Yu, Zhiguo Wan, et al. (2024). *Knowledge graph confidence-aware embedding for recommendation*. Neural Networks.
- Yuzhuo Wang, Hongzhi Wang, Xianglong Liu, et al. (2024). *GFedKG: GNN-based federated embedding model for knowledge graph completion*. Knowledge-Based Systems.
- Xinyan Wang, Kuo Yang, Ting Jia, et al. (2024). *KDGene: knowledge graph completion for disease gene prediction using interactional tensor decomposition*. Briefings Bioinform..
- Yuhan Liu, Zelin Cao, Xing Gao, et al. (2024). *Bridging the Space Gap: Unifying Geometry Knowledge Graph Embedding with Optimal Transport*. The Web Conference.
- Yongfang Li, and Chunhua Zhu (2024). *TransE-MTP: A New Representation Learning Method for Knowledge Graph Embedding with Multi-Translation Principles and TransE*. Electronics.



- Qianjin Zhang, and Yandan Xu (2024). *Knowledge graph embedding with inverse function representation for link prediction*. Engineering applications of artificial intelligence.
- Hao Wang, Dandan Song, Zhijing Wu, et al. (2024). *A collaborative learning framework for knowledge graph embedding and reasoning*. Knowledge-Based Systems.
- Shengkun Ni, Xiangtai Kong, Yingying Zhang, et al. (2024). *Identifying compound-protein interactions with knowledge graph embedding of perturbation transcriptomics*. Cell Genomics.
- Jixuan Nie, Xia Hou, Wenfeng Song, et al. (2024). *Knowledge Graph Efficient Construction: Embedding Chain-of-Thought into LLMs*. VLDB Workshops.
- 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.
- Yuren Mao, Yu Hao, Xin Cao, et al. (2024). *Dynamic Graph Embedding via Meta-Learning*. IEEE Transactions on Knowledge and Data Engineering.
- 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..
- Yalin Wang, Yubin Peng, and Jingyu Guo (2024). *Enhancing knowledge graph embedding with structure and semantic features*. Applied intelligence (Boston).
- 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.
- Yadan Han, Guangquan Lu, Shichao Zhang, et al. (2024). *A Temporal Knowledge Graph Embedding Model Based on Variable Translation*. Tsinghua Science and Technology.

- 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.
- Yunjie He, Daniel Hernández, M. Nayyeri, et al. (2024). *Generating SROI<sup>-</sup> Ontologies via Knowledge Graph Query Embedding Learning*. Unpublished manuscript.
- Yan Fang, Xiaodong Liu, Wei Lu, et al. (2024). *Knowledge graph completion with low-dimensional gated hierarchical hyperbolic embedding*. Knowledge-Based Systems.
- 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.
- 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.
- 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.
- Tao Wang, Bo Shen, Jinglin Zhang, et al. (2024). *Knowledge Graph Embedding via Triplet Component Interactions*. Neural Processing Letters.
- Pengfei Zhang, Xiaoxue Zhang, Yang Fang, et al. (2024). *Knowledge Graph Embedding for Hierarchical Entities Based on Auto-Embedding Size*. Mathematics.
- K. Liang, Yue Liu, Hao Li, et al. (2024). *Clustering then Propagation: Select Better Anchors for Knowledge Graph Embedding*. Neural Information Processing Systems.
- 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.

- 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.
- 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.
- Nan Li, Zhihao Yang, Jian Wang, et al. (2024). *Drug-target interaction prediction using knowledge graph embedding*. iScience.
- 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.
- Guoshun Xu, Guozheng Rao, Li Zhang, et al. (2024). *Entity-relation aggregation mechanism graph neural network for knowledge graph embedding*. Applied intelligence (Boston).
- Qiuyu Liang, Weihua Wang, Jie Yu, et al. (2024). *Effective Knowledge Graph Embedding with Quaternion Convolutional Networks*. Natural Language Processing and Chinese Computing.
- Jie Liu, Lizheng Zu, Yunbin Yan, et al. (2024). *Multi-Filter soft shrinkage network for knowledge graph embedding*. Expert systems with applications.
- 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.
- 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.

- Xinyue Liu, Jianan Zhang, Chi Ma, et al. (2024). *Temporal Knowledge Graph Reasoning with Dynamic Hypergraph Embedding*. International Conference on Language Resources and Evaluation.
- Ruiyi Yang, Flora D. Salim, and Hao Xue (2024). *SSTKG: Simple Spatio-Temporal Knowledge Graph for Interpretable and Versatile Dynamic Information Embedding*. The Web Conference.
- Bo Li, Haowei Quan, Jiawei Wang, et al. (2024). *Neural Library Recommendation by Embedding Project-Library Knowledge Graph*. IEEE Transactions on Software Engineering.
- 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.
- 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 and Intelligent Systems.
- 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.
- 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.
- 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.