PAAL: Pattern-Anchor Alignment for Continual Knowledge Graph Embedding Under Structural Distribution Shift

Yue Jian[1] , **Lin Li**[1*] , **Kaize Shi**[2] , **Junwei Zhou**[1] , **Yu Yang**[3]
[1]Wuhan University of Technology, Wuhan, China
[2]University of Southern Queensland, Toowoomba, Australia
[3]The Education University of Hong Kong, Hong Kong, China
jianyue@whut.edu.cn, cathylilin@whut.edu.cn, kaize.shi@unisq.edu.au, junweizhou@whut.edu.cn, yangyy@eduhk.hk

Abstract

Continual knowledge graph embedding (CKGE) has gained popularity for managing dynamic knowledge graphs. Unlike general graph continual-learning approaches, CKGE focuses on retaining triple-level knowledge, thereby overcoming the inability of static models to accommodate continuously arriving facts. However, as new fact triples come in, real-world knowledge graphs often experience **structural distribution shifts**: the underlying topology and the way facts interconnect evolve over time. This presents a challenge that the structural distribution of historical data may diverge from that of emerging data, creating a distributional mismatch that complicates the adaptation to new trends. To address this, we propose Pattern-Anchor Alignment (PAAL), which introduces relation-level structural anchors to explicitly model these structural shifts. By quantifying the alignment between the structural distribution of emerging triples and historical anchors, PAAL implements an alignment-modulated optimization strategy. It utilizes a calibrated gating mechanism to precisely regulate the intensity of historical constraints based on structural stability. Experimental results indicate the effectiveness of PAAL, showing average H@1 improvements of 11.77% and 9.29% on pattern-shift and standard CKGE benchmarks, respectively. Our code is available at https://github.com/cangjie553/PAAL.

1 Introduction

In the era of dynamic information, modern artificial intelligence systems are increasingly required to operate as continual learners, capable of adapting to non-stationary data streams throughout their lifecycle. This continuous adaptation is critical for applications such as conversational agents [Dinan *et al.*, 2019], real-time question answering [Lan *et al.*, 2021], and personalized recommendation systems [Guo *et al.*, 2022; Wang *et al.*, 2024], where the reasoning module must evolve in tandem with emerging data [Zhu *et al.*, 2023]. Consequently, the underlying Knowledge Graphs (KGs) supporting these systems cannot remain static; they must evolve dynamically, continuously integrating new entities, relations, and facts to reflect the changing world [Lin *et al.*, 2022]. For instance, DBpedia, a widely used KG, incorporated over 1 million entities and 20 million triples between 2016 and 2018 alone [Lehmann *et al.*, 2015]. Such rapid and unceasing expansion highlights the necessity for Knowledge Graph Embedding (KGE) frameworks to possess dynamic adaptability, ensuring that reasoning knowledge remains aligned with the latest real-world knowledge [Jian *et al.*, 2025a; Song *et al.*, 2026; Jian *et al.*, 2025b].

Static retraining necessitates rebuilding the entire model from scratch whenever new data arrives. To address this limitation, several graph continual learning works have focused on general graph tasks such as node classification [Liu *et al.*, 2021; Zhang *et al.*, 2023]. In contrast, Continual Knowledge Graph Embedding (CKGE) has emerged as a framework specifically focused on the retention of factual triples in knowledge graphs. CKGE aims to incrementally update embeddings by integrating newly added knowledge while preserving previously learned representations [Daruna *et al.*, 2021; Cui *et al.*, 2023; Pons *et al.*, 2025]. The core challenge in this paradigm lies in effectively acquiring emerging information without disrupting the integrity of historical knowledge. Existing CKGE methods fall into four lines: *dynamic-structure* (PNN [Rusu *et al.*, 2016], CWR [Lomonaco and Maltoni, 2017]), *memory-replay* (GEM [Lopez-Paz and Ranzato, 2017], EMR [Wang *et al.*, 2019], DiCGRL [Kou *et al.*, 2020]), *regularization* (SI [Zenke *et al.*, 2017], EWC [Kirkpatrick *et al.*, 2017]), and *CKGE-specific* approaches (LKGE [Cui *et al.*, 2023], FastKGE [Liu *et al.*, 2024b], IncDE [Liu *et al.*, 2024a], SAGE [Li *et al.*, 2025b]). These methods have achieved substantial progress in handling incremental data.

However, in real-world scenarios, the continuous evolution of KGs involves more than a simple increase in fact triple counts. In addition to the growth of factual knowledge, the graph topology often undergoes structural distribution shifts, where connectivity distribution patterns evolve over time [Zhao *et al.*, 2025; Li *et al.*, 2025a]. For instance, as illustrated in Figure 1, consider the evolution of a social interaction KG. In the initial phase, the graph typically exhibits a centralized hub structure, where connections concentrate around a few popular entities, and reasoning re-

*Corresponding author.

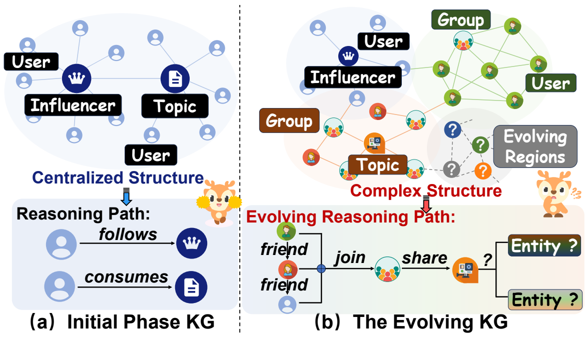

Figure 1: An illustration of structural connectivity evolution in dynamic knowledge graphs.

lies on simple paths like `User-Follows-Influencer`. As the network matures, it evolves into a dense, interconnected community driven by user interactions, shifting the global topology from a sparse star-like shape to a complex mesh. This introduces intricate multi-hop patterns such as `User-Friend-Group-Topic-Group`. Consequently, embeddings learned under the historical hub-centric prior are difficult to capture the mutual dependencies of emerging community structures, leading to challenging reasoning. Existing CKGE methods primarily focus on data volume growth rather than the evolution of topological structure distributions. Consequently, these approaches face challenges in handling topological changes within evolving knowledge graphs, underscoring the need for a framework explicitly designed for structural distribution shifts.

To this end, we propose **P**attern-**A**nchor **Al**ignment (PAAL), a continual knowledge graph embedding framework explicitly designed to model structural distribution shifts. To account for the topological evolution of KGs, PAAL tracks the changing distribution of each relation via relation-level structural anchors. These anchors, constructed by aggregating structural representations at each snapshot, serve as pattern anchors that define the dominant connective distribution patterns. For every new triple, PAAL computes an alignment coefficient relative to its corresponding anchor, further refined by a calibration assessment to quantify structural consistency. This alignment signal functions as a dynamic modulation mechanism: for triples exhibiting high consistency, PAAL intensifies historical constraints to preserve valid inductive biases. Conversely, for instance, diverging from the anchors attenuates these constraints to prevent the rigid retention of obsolete structural assumptions. Integrated with consistency-aware optimization techniques, this design enables PAAL to maintain embedding stability during evolution and effectively adapt to evolving structural distributions. In summary, while traditional methods primarily focus on the balance between new and old knowledge, PAAL advances a consistency-aware optimization approach. By utilizing pattern structural anchors, PAAL maintains stability under structural distribution shifts during evolution.

The primary contributions of this work are as follows:

- We observe that challenges in CKGE arise not only from data growth but also from structural distribution shifts in evolving connectivity patterns.
- We propose PAAL, which utilizes structural anchors to adaptively regulate historical constraints via pattern-anchor alignment to maintain stability under structural distribution shifts during evolution.
- Extensive experiments on pattern-shift and standard benchmarks demonstrate PAAL achieves state-of-the-art performance against baselines, validating its effectiveness in evolving knowledge graphs.

2 Related Work

2.1 Continual Knowledge Graph Embedding

General Continual Graph Learning methods predominantly target node or graph classification tasks [Liu *et al.*, 2021; Zhang *et al.*, 2023; Zhou and Cao, 2021]. Derived from generic continual learning paradigms, these approaches are categorized into three types: dynamic architecture methods that extend model structures [Rusu *et al.*, 2016; Lomonaco and Maltoni, 2017; Zhang *et al.*, 2023]; memory-replay methods that alleviate forgetting by revisiting historical subgraphs [Lopez-Paz and Ranzato, 2017; Zhou and Cao, 2021]; and regularization methods [Kirkpatrick *et al.*, 2017; Zenke *et al.*, 2017] that constrain topology-aware weights [Liu *et al.*, 2021] or stabilize relational embeddings [Kou *et al.*, 2020; Hamaguchi *et al.*, 2017].

Continual Knowledge Graph Embedding (CKGE) has emerged as a practical framework specifically focused on the retention of factual triples in dynamic knowledge graphs, diverging from general graph continual learning works that primarily target general graph tasks. Unlike static KGE approaches [Bordes *et al.*, 2013; Sun *et al.*, 2019] that require retraining, CKGE aims to incrementally integrate newly arriving knowledge while maintaining previously learned knowledge. To overcome these challenges, recent CKGE-specific studies have addressed handling unseen elements [Cui *et al.*, 2023], optimizing learning order [Liu *et al.*, 2024a], improving parameter efficiency [Liu *et al.*, 2024b; Zhu *et al.*, 2025b], adapting to update scales [Li *et al.*, 2025b; Li *et al.*, 2026], modeling relational asymmetry [Daruna *et al.*, 2021], and mitigating spurious forgetting [Zhu *et al.*, 2025a]. Despite these advancements, the adaptation to evolving structural distributions has not been explicitly explored.

2.2 Structural Distribution Shift

Studies on Graph Out-of-Distribution (OOD) generalization [Wu *et al.*, 2024; Li *et al.*, 2024; Li *et al.*, 2025a] confirm that structural distribution shifts impair performance, yet they are strictly limited to static train-test discrepancies. Consequently, these static paradigms are incapable of addressing the dynamic evolutionary scenarios inherent to CKGE. In this evolving setting, while a recent CKGE benchmark [Zhao *et al.*, 2025] characterizes pattern shifts, it primarily serves as an evaluation tool, failing to address the reduction in expressivity caused by the accumulation of connectivity changes during embedding. **Existing CKGE methods** thus lack a direct mechanism to explicitly track structural distribution shifts over time, motivating **our work** to model evolving relational patterns and adaptively regulate historical constraints.

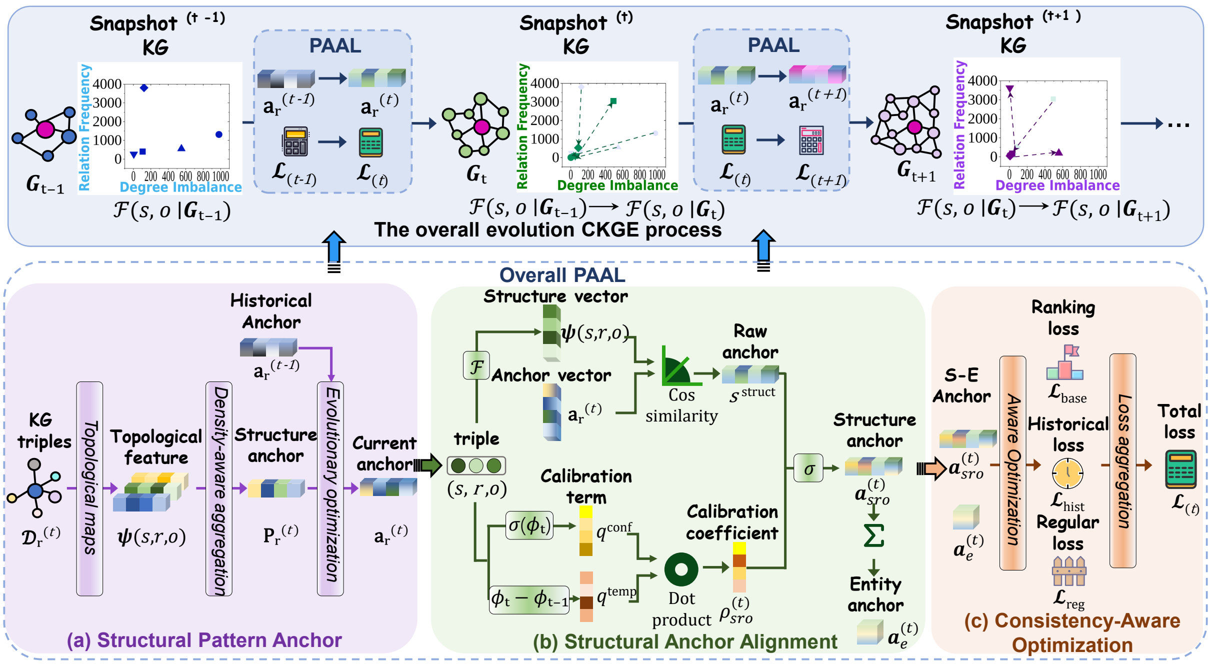

Figure 2: An overview of our proposed PAAL framework.

3 Preliminaries

Definition 1. ***Evolving Knowledge Graph:*** *A knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ consists of a set of entities $\mathcal{E}$, relations $\mathcal{R}$, and triples $\mathcal{T}$, where each triple $(s, r, o) \in \mathcal{T}$ represents a factual statement with subject entity s, relation r, and object entity o. An evolving knowledge graph represents the temporal evolution of a KG as a sequence of snapshots $\mathcal{G}_t = (\mathcal{E}_t, \mathcal{R}_t, \mathcal{T}_t)$ at discrete time steps t. To characterize the dynamic growth between consecutive snapshots, we define $\Delta\mathcal{T}_t = \mathcal{T}_t \setminus \mathcal{T}_{t-1}$, $\Delta\mathcal{E}_t = \mathcal{E}_t \setminus \mathcal{E}_{t-1}$, and $\Delta\mathcal{R}_t = \mathcal{R}_t \setminus \mathcal{R}_{t-1}$ as the sets of newly added triples, entities, and relations at time t, respectively.*

Definition 2. ***Continual Knowledge Graph Embedding (CKGE):*** *CKGE focuses on learning and maintaining latent vector representations for entities and relations within an evolving knowledge graph. Given a sequence of KG snapshots $\{\mathcal{G}_t\}$, CKGE incrementally integrates newly added triples, entities, and relations while updating existing embeddings to align with the current snapshot, preserving previously acquired knowledge.*

4 Methodology

This section details the Pattern-Anchor Alignment (PAAL) framework, as illustrated in Figure 2. The PAAL method comprises three primary components: (1) **Structural Anchor Construction**: constructs structural evolutionary anchors; (2) **Structural Anchor Alignment**: performs calibrated anchor alignment; (3) **Consistency-Aware Optimization**: optimizes embeddings using consistency constraints.

4.1 Structural Anchor Construction

To explicitly characterize structural evolution, we visualize the temporal trajectories of five relations randomly selected from the PS-CKGE$_{3:1:1}$ dataset at the top of Figure 2. We map these relations into a topological space defined by Degree Imbalance and Relation Frequency to quantify their structural distribution states. As illustrated, the structural distribution of relations exhibits notable temporal variations across consecutive snapshots. For instance, a relation may evolve from a structurally balanced state to an imbalanced one (e.g., shifting from 1-to-1 to 1-to-N connectivity) or experience fluctuations in frequency accompanied by structural adjustments. These trajectories highlight the presence of a structural distribution shift, where the distribution of emerging data diverges from that of historical data. This analysis suggests that simple static aggregations may face limitations in capturing such dynamic regularities. Motivated by this, we propose the structural pattern anchor, a mechanism designed to track these evolving structural trends and provide an adaptive reference for alignment.

Structural Anchor Evolution. To effectively adapt to these evolving dynamics, we define the structural pattern anchor $\mathbf{a}_r^{(t)}$ as a temporally consistent reference that balances fitness to the current snapshot with historical stability. We formulate the anchor evolution by minimizing the following objective function $\mathcal{J}$:

$$\mathcal{J} = \|\mathbf{a}_r^{(t)} - \mathbf{p}_r^{(t)}\|_2^2 + \lambda_a \|\mathbf{a}_r^{(t)} - \mathbf{a}_r^{(t-1)}\|_2^2, \tag{1}$$

where $\mathbf{p}_r^{(t)}$ represents the structural anchor observed at the current snapshot t, and λ_a is a regularization hyperparameter that controls the trade-off between fitting the current observation and preserving historical constraint. Minimizing $\mathcal{J}$

yields a recursive closed-form update rule:

$$\mathbf{a}_r^{(t)} = (1-\gamma)\mathbf{p}_r^{(t)} + \gamma \mathbf{a}_r^{(t-1)}, \tag{2}$$

where $\gamma = \frac{\lambda_a}{1+\lambda_a} \in [0, 1)$ is a momentum factor. This formulation enables the anchor to smoothly track the continuous structural drift observed in Figure 2.

Density-Aware Anchor Aggregation. To ensure the current observation $\mathbf{p}_r^{(t)}$ is representative despite noise, we employ a density-aware aggregation strategy. We first extract a structural feature vector $\psi(s, r, o)$ that characterizes the local structural context of the triple. Rather than relying on raw entity indices, this vector encodes intrinsic structural distribution via a structural feature extraction function $\mathcal{F}$:

$$\boldsymbol{\psi}(s, r, o) = \mathcal{F}(s, o \mid \mathcal{G}_t) \in \mathcal{R}^k, \tag{3}$$

specifically, the extraction function $\mathcal{F}$ maps the local context of each triple to a 16-dimensional feature vector ψ that aggregates statistics from three complementary views: (i) *head entity statistics*, including the average, standard deviation, minimum/maximum, and total count of head-entity degrees; (ii) *tail entity statistics*, which compute the analogous quantities on the tail side; and (iii) *relational topology*, capturing relation frequency, head–tail degree imbalance $|deg_h - deg_t|$, unique entity ratio, reciprocity score, average degree of all involved entities, and degree correlation. Together, these metrics jointly characterize the centrality, distributional spread, and connectivity patterns of relations within the evolving graph. To filter outliers, we assign a reliability weight $\omega_{sro}^{(t)}$ based on the proximity to the local density center $\boldsymbol{\mu}_r^{(t)}$ via a Gaussian kernel:

$$\omega_{sro}^{(t)} = \frac{\exp\left(-\|\boldsymbol{\psi}(s, r, o) - \boldsymbol{\mu}_r^{(t)}\|_2^2 \,/\, 2\sigma^2\right)}{\sum_{(s',r,o')\in\mathcal{D}_r^{(t)}} \exp\left(-\|\boldsymbol{\psi}(s', r, o') - \boldsymbol{\mu}_r^{(t)}\|_2^2 \,/\, 2\sigma^2\right)}, \tag{4}$$

the robust structural anchor is then computed as the weighted sum:

$$\mathbf{p}_r^{(t)} = \sum_{(s,r,o)\in\mathcal{D}_r^{(t)}} \omega_{sro}^{(t)} \cdot \boldsymbol{\psi}(s, r, o). \tag{5}$$

Overall, this mechanism establishes a persistent yet adaptive structural anchor. By dynamically modeling structural distribution shifts through time-aware accumulation and density-based filtering, it serves as the foundational reference for the subsequent alignment.

4.2 Structural Anchor Alignment

While relation anchors capture snapshot-specific structure patterns, effective adaptation requires a stable reference. Therefore, we utilize the learned structural anchor $\mathbf{a}_r^{(t)}$ as the robust reference target. Not all newly observed triples should exert equal importance; deviations may arise from noise or transient observations. Consequently, PAAL introduces a calibrated structural anchor alignment mechanism that jointly accounts for structural temporal consistency and triple plausibility.

Raw Structural Alignment. Given a triple (s, r, o) and its structural feature vector $\boldsymbol{\psi}(s, r, o)$, we compute a raw structural alignment coefficient against the corresponding structural pattern anchor $\mathbf{a}_r^{(t)}$:

$$s^{\text{struct}} = \frac{\boldsymbol{\psi}(s, r, o)^\top \mathbf{a}_r^{(t)}}{\|\boldsymbol{\psi}(s, r, o)\|_2 \cdot \|\mathbf{a}_r^{(t)}\|_2}, \tag{6}$$

which utilizes cosine similarity to measure the geometric consistency between the current triple's structural characteristics and the persistent structural anchor of relation r.

Calibration Modulated Alignment. To calibrate this alignment signal, we further model from two complementary perspectives: plausibility and temporal stability, defined as follows:

$$\begin{cases} q^{\text{conf}} = \tau - \sigma(\phi_t), \\ q^{\text{temp}} = \exp\big(-|\phi_t - \phi_{t-1}|\big), \end{cases} \tag{7}$$

here, $\sigma(\cdot)$ denotes the sigmoid function, ϕ_t denotes the plausibility score, $\tau > 0$ serves as a tolerance threshold. Adopting the standard translation-based scoring function [Bordes *et al.*, 2013], we define $\phi_t = -\|\mathbf{s} + \mathbf{r} - \mathbf{o}\|_2$, where $\mathbf{s}, \mathbf{r}, \mathbf{o}$ are the embeddings of the subject entity, relation, and object entity. The first term q^{conf} utilizes model confidence to filter implausible triples. The second term q^{temp} rewards temporal stability, decaying with score fluctuations. These factors are synthesized into a joint calibration coefficient $\rho_{sro}^{(t)}$, which serves as a gating mechanism to validate the structural signal:

$$\rho_{sro}^{(t)} = q^{\text{conf}} \cdot q^{\text{temp}}, \tag{8}$$

finally, we derive the pattern anchor alignment coefficient $\alpha_{sro}^{(t)}$ by modulating the raw structural alignment score with this calibration gate:

$$\alpha_{sro}^{(t)} = \sigma\Big(\beta \cdot s^{\text{struct}} \cdot \rho_{sro}^{(t)}\Big), \tag{9}$$

where β is a scaling hyperparameter that controls the sensitivity of the modulation. This formulation explicitly positions the structural anchor alignment s^{struct} as the primary driving force, while $\rho_{sro}^{(t)}$ ensures that historical constraints are enforced only when the observation is deemed calibrated from semantic and temporal.

Entity Anchor Alignment. To propagate alignment signals to entity representations, we further define an entity-level stability coefficient $\bar{\alpha}_e^{(t)}$ for any entity e:

$$\bar{\alpha}_e^{(t)} = \frac{1}{|N_t(e)|} \sum_{(s,r,o)\in N_t(e)} \alpha_{sro}^{(t)}, \tag{10}$$

where $N_t(e)$ denotes the set of triples incident to entity e at time t. This alignment ensures that entities involved in reliable structures are prioritized for stability preservation.

Overall, by explicitly quantifying structural shifts via pattern-anchor alignment, this mechanism translates topological signals into dynamic calibration gates. While traditional CKGE methods primarily focus on the balance between new and old knowledge, PAAL advances a consistency-aware approach that focuses on structural distribution shifts by quantitatively aligning structural anchors in evolving topologies to adaptively regulate historical constraints.

Dataset	Metric		Dynamic Structure		Memory Replay			Regularization		CKGE-Specific				
		Fine-tune	PNN (Arxiv'16)	CWR (CoRL'17)	GEM (NIPS'17)	EMR (NAACL'19)	DiCGRL (EMNLP'20)	SI (ICML'17)	EWC (PNAS'17)	LKGE (AAAI'23)	FastKGE (IJCAI'24)	IncDE (AAAI'24)	SAGE (KDD'25)	**PAAL (Ours)**
ENTITY	MRR	0.165	0.229	0.088	0.165	0.171	0.107	0.154	0.229	0.234	0.239	0.253	0.280	**0.287**
	H@1	0.085	0.130	0.028	0.085	0.090	0.057	0.072	0.130	0.136	0.146	0.151	0.176	**0.184**
	H@10	0.321	0.425	0.202	0.321	0.330	0.211	0.311	0.423	0.425	0.427	0.448	0.477	**0.485**
RELATION	MRR	0.093	0.167	0.021	0.093	0.111	0.133	0.113	0.165	0.192	0.185	0.199	0.217	**0.235**
	H@1	0.039	0.096	0.010	0.040	0.052	0.079	0.055	0.093	0.106	0.107	0.111	0.122	**0.134**
	H@10	0.195	0.305	0.043	0.196	0.225	0.241	0.224	0.306	0.366	0.359	0.370	0.397	**0.421**
FACT	MRR	0.172	0.157	0.083	0.175	0.171	0.162	0.172	0.201	0.210	0.203	0.216	0.222	**0.237**
	H@1	0.090	0.084	0.030	0.092	0.090	0.084	0.088	0.113	0.122	0.117	0.128	0.134	**0.150**
	H@10	0.339	0.290	0.192	0.345	0.337	0.320	0.343	0.382	0.387	0.384	0.391	0.394	**0.415**
HYBRID	MRR	0.135	0.185	0.037	0.136	0.141	0.149	0.111	0.186	0.207	0.211	0.224	0.224	**0.238**
	H@1	0.069	0.101	0.015	0.070	0.073	0.083	0.049	0.102	0.121	0.128	0.131	0.131	**0.142**
	H@10	0.262	0.349	0.077	0.263	0.267	0.277	0.229	0.350	0.379	0.382	0.401	0.400	**0.422**
GraphEqual	MRR	0.183	0.212	0.122	0.189	0.185	0.104	0.179	0.207	0.214	-	0.234	0.247	**0.262**
	H@1	0.096	0.118	0.041	0.099	0.099	0.040	0.092	0.113	0.118	-	0.134	0.147	**0.159**
	H@10	0.358	0.405	0.277	0.372	0.359	0.226	0.353	0.400	0.407	-	0.432	0.446	**0.467**
GraphHigher	MRR	0.198	0.186	0.189	0.197	0.202	0.116	0.190	0.198	0.207	-	0.227	0.239	**0.254**
	H@1	0.108	0.097	0.096	0.109	0.113	0.041	0.099	0.106	0.120	-	0.132	0.143	**0.156**
	H@10	0.375	0.364	0.374	0.372	0.379	0.242	0.371	0.385	0.382	-	0.412	0.426	**0.447**
GraphLower	MRR	0.185	0.213	0.032	0.170	0.188	0.102	0.186	0.210	0.210	-	0.228	0.237	**0.259**
	H@1	0.098	0.119	0.005	0.084	0.101	0.039	0.099	0.116	0.116	-	0.129	0.138	**0.156**
	H@10	0.363	0.407	0.080	0.346	0.362	0.222	0.366	0.405	0.403	-	0.426	0.432	**0.467**

Table 1: Performance comparison on standard CKGE datasets representing general evolutionary scenarios: ENTITY, RELATION, FACT, HYBRID, GraphEqual, GraphHigher, and GraphLower. Bold values denote the best results, while underlined values indicate the second-best.

4.3 Consistency-Aware Optimization

Under structural distribution shifts, uniformly enforcing historical constraints may hinder adaptation and lead to negative transfer. Therefore, historical knowledge should be selectively preserved. Guided by pattern-anchor alignment signals, PAAL introduces a consistency-aware optimization strategy that maintains embedding stability in evolving.

Margin Ranking Loss. Using the plausibility score ϕ_t defined previously, the margin-based ranking loss is expressed as:

$$\mathcal{L}_{\text{base}} = \sum_{(s,r,o)\in\mathcal{T}_t} \max(0, \gamma_m - \phi_t(s,r,o) + \phi_t(s',r,o')), \tag{11}$$

where $\phi_t(s',r,o')$ denotes the score of a negative sample and γ_m is the margin hyperparameter.

Historical Consistency Loss. To preserve stable historical knowledge while allowing adaptation, we introduce an alignment-guided historical consistency loss:

$$\mathcal{L}_{\text{hist}} = \sum_{(s,r,o)\in\mathcal{T}_t} \alpha_{sro}^{(t)}(\phi_t - \phi_{t-1})^2 + \sum_{e\in\mathcal{E}_t} \bar{\alpha}_e^{(t)} \|\mathbf{e}^{(t)} - \mathbf{e}^{(t-1)}\|_2^2. \tag{12}$$

Adaptive Regularization Loss. Furthermore, to suppress noise fitting, we incorporate an adaptive parameter regularization term:

$$\mathcal{L}_{\text{reg}} = \sum_{(s,r,o)\in\mathcal{T}_t} (1 - \alpha_{sro}^{(t)}) \left(\|\mathbf{s}\|_2^2 + \|\mathbf{r}\|_2^2 + \|\mathbf{o}\|_2^2\right). \tag{13}$$

The final optimization objective of PAAL is:

$$\mathcal{L} = \mathcal{L}_{\text{base}} + \lambda_h \mathcal{L}_{\text{hist}} + \lambda_r \mathcal{L}_{\text{reg}}, \tag{14}$$

where λ_h and λ_r control the trade-off between historical consistency and parameter regularization. By selectively enforcing constraints based on structural alignment $\alpha_{sro}^{(t)}$, PAAL retains valid inductive biases while flexibly adapting to evolving structural distribution.

5 Experiment

In this section, we conduct an extensive series of experiments using seven standard CKGE datasets and six pattern-shift CKGE datasets to evaluate the effectiveness of the proposed PAAL framework. Our empirical analysis addresses the following three research questions:

- **RQ1:** How does the performance of PAAL compare with state-of-the-art methods?
- **RQ2:** How do the individual components contribute to the effectiveness of the PAAL framework?
- **RQ3:** How does PAAL perform regarding efficiency, robustness, and hyperparameter sensitivity?

5.1 Experimental Settings

Datasets

The PAAL framework is evaluated on **seven standard CKGE datasets** and **six pattern-shift CKGE datasets**. Standard datasets (ENTITY, RELATION, FACT, HYBRID, GraphEqual, GraphHigher, GraphLower) capture diverse evolutionary dynamics across five time steps [Cui *et al.*, 2023; Liu *et al.*, 2024a]. To investigate structural shifts, the PS-CKGE and FACT series datasets (derived from FB15k-237 with pattern shift ratios 4:1, 3:2, 3:1:1) are utilized, explicitly incorporating distinct logical pattern shifts [Zhao *et al.*, 2025; Bordes *et al.*, 2013].

Evaluation Metrics

Experimental performance is assessed using standard **Mean Reciprocal Rank (MRR)** and **Hits@N** ($N \in \{1, 3, 10\}$). To comprehensively assess evolutionary quality, we additionally employ **Forward Transfer (FWT)** and **Resistance to Forgetting (RtF)** [Li *et al.*, 2025b; Zhao *et al.*, 2025]. FWT averages zero-shot performance on emerging snapshots to measure transferability, while RtF assesses stability via historical performance retention after final training.

Method	PS-CKGE$_{4:1}$			FACT$^*_{4:1}$			PS-CKGE$_{3:2}$			FACT$^*_{3:2}$			PS-CKGE$_{3:1:1}$			FACT$^*_{3:1:1}$		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
Joint	0.324	0.204	0.553	0.326	0.206	0.555	0.323	0.202	0.554	0.326	0.206	0.556	0.325	0.205	0.555	0.326	0.206	0.556
GEM	0.177	0.102	0.320	0.211	0.121	0.388	0.186	0.108	0.336	0.220	0.127	0.405	0.146	0.082	0.287	0.204	0.113	0.383
EMR	0.186	0.107	0.335	0.204	0.116	0.375	0.197	0.117	0.354	0.213	0.122	0.390	0.149	0.093	0.289	0.189	0.104	0.359
EWC	0.231	0.138	0.410	0.238	0.143	0.423	0.215	0.125	0.388	0.242	0.146	0.428	0.200	0.129	0.373	0.232	0.137	0.421
LKGE	0.235	0.142	0.415	0.242	0.148	0.426	0.232	0.140	0.411	0.248	0.153	0.434	0.208	0.119	0.387	0.238	0.144	0.426
IncDE	0.214	0.122	0.394	0.232	0.137	0.417	0.223	0.134	0.398	0.240	0.143	0.425	0.222	0.133	0.397	0.231	0.138	0.414
SAGE	0.222	0.134	0.395	0.229	0.138	0.407	0.224	0.137	0.394	0.235	0.143	0.414	0.214	0.130	0.381	0.224	0.135	0.401
PAAL	**0.262**	**0.162**	**0.457**	**0.265**	**0.164**	**0.459**	**0.252**	**0.156**	**0.443**	**0.263**	**0.163**	**0.455**	**0.254**	**0.157**	**0.445**	**0.257**	**0.158**	**0.452**
Imp.(↑)	+11.5%	+14.1%	+10.1%	+9.5%	+10.8%	+7.7%	+8.6%	+11.4%	+7.8%	+6.0%	+6.5%	+4.8%	+14.4%	+18.0%	+12.1%	+8.0%	+9.7%	+6.1%

Table 2: Performance comparison on pattern-shift CKGE datasets explicitly designed to evaluate adaptability to structural pattern shift: PS-CKGE$_{4:1}$, FACT$^*_{4:1}$, PS-CKGE$_{3:2}$, FACT$^*_{3:2}$, PS-CKGE$_{3:1:1}$, and FACT$^*_{3:1:1}$. (Ratios represent the degrees of inter-snapshot shift.)

Method	PS-CKGE$_{3:1:1}$		FACT$^*_{4:1}$		FACT	
	RtF	FWT	RtF	FWT	RtF	FWT
GEM	0.452	0.145	0.492	0.149	0.475	0.173
IncDE	0.406	0.131	0.489	0.145	0.413	0.160
LKGE	0.478	0.148	**0.499**	0.153	0.481	0.176
SAGE	0.481	0.159	**0.499**	0.166	0.494	0.182
PAAL	**0.496**	**0.185**	**0.499**	**0.189**	**0.499**	**0.191**

Table 3: Performance comparison on RtF and FWT metrics.

Methods	PS-CKGE$_{3:1:1}$			FACT		
	MRR	H@1	H@10	MRR	H@1	H@10
w/o PAAL	0.211	0.128	0.376	0.215	0.127	0.384
w/o SAA	0.229	0.136	0.405	0.222	0.134	0.397
w/o CMA	0.245	0.152	0.434	0.232	0.145	0.408
w/o HOA	0.237	0.143	0.421	0.229	0.139	0.406
w/o ROA	0.241	0.149	0.430	0.231	0.142	0.405
PAAL	**0.254**	**0.157**	**0.445**	**0.237**	**0.150**	**0.415**

Table 4: Ablation study on different components of PAAL.

Compared Baselines

PAAL is benchmarked against a diverse set of baselines, encompassing fine-tuning with parameter inheritance and the performance upper bound-Joint strategy [Zhao *et al.*, 2025; Li *et al.*, 2025b], **traditional continual learning methods** encompass dynamic architecture-based methods (PNN, CWR) [Rusu *et al.*, 2016; Lomonaco and Maltoni, 2017], memory-replay-based methods (GEM, EMR, DiCGRL) [Lopez-Paz and Ranzato, 2017; Wang *et al.*, 2019; Kou *et al.*, 2020], and regularization-based methods (SI, EWC) [Zenke *et al.*, 2017; Kirkpatrick *et al.*, 2017]. And recent **CKGE-specific methods** (LKGE, IncDE, FastKGE, SAGE) [Cui *et al.*, 2023; Liu *et al.*, 2024a; Liu *et al.*, 2024b; Li *et al.*, 2025b].

5.2 Performance Comparison (RQ1)

To evaluate the performance of PAAL, we conducted systematic link prediction experiments on standard CKGE benchmarks and pattern-shift benchmarks, alongside an analysis of Forward Transfer (FWT) and Resistance to Forgetting (RtF). The quantitative results are presented in Table 1, Table 2, and Table 3. In standard CKGE tasks (Table 1), PAAL achieves competitive performance compared to baselines. Notably, on the RELATION dataset, which features disjoint relation sets, PAAL achieves an 8.3% MRR improvement (0.235 vs. 0.217) over the next best baseline, indicating robustness in relation-dynamic scenarios. Moreover, in pattern-shift scenarios (Table 2), the performance gain of PAAL is even more evident. For instance, on PS-CKGE$_{3:1:1}$, it outperforms the baseline IncDE by 18.0% in H@1 (0.157 vs. 0.133). Furthermore, Table 3 confirms PAAL's favorable evolutionary quality, exceeding the best baselines by 3.1% in RtF and 16.4% in FWT on PS-CKGE$_{3:1:1}$, suggesting enhanced adaptability. These observations yield three key conclusions.

(1) PAAL shows effectiveness across different evolutionary scenarios. The consistent performance on both standard and pattern-shift benchmarks indicates that structural anchors provide stable guidance under different conditions.

(2) PAAL effectively adapts to structural distribution shifts. Compared to baselines, PAAL maintains strong predictive performance on pattern-shift datasets, demonstrating that modulating constraints via structural alignment enables robust adaptation to evolving structural distribution.

(3) PAAL maintains stability during evolution. The favorable RtF and FWT scores demonstrate that the consistency-aware optimization effectively maintains forward transfer and stability during knowledge graph evolution.

5.3 Ablation Study (RQ2)

To assess the contribution of each component, we conduct an ablation study on PS-CKGE$_{3:1:1}$ and FACT, as summarized in Table 4. We compare the full PAAL model against several variants excluding: the whole alignment strategy (**w/o PAAL**), structural anchors alignment (**w/o SAA**), calibration modulation (**w/o CMA**), historical loss optimization (**w/o HOA**), and regularization loss optimization (**w/o ROA**). The removal of any component results in performance degradation. Notably, the **w/o PAAL** variant exhibits the most significant decline (with MRR decreasing from 0.254 to 0.211 on PS-CKGE$_{3:1:1}$), underscoring the importance of the integrated framework. Furthermore, **w/o SAA** suffers a more pronounced decline than **w/o CMA**, specifically on the pattern-shift dataset, highlighting the critical role of structural anchors in capturing evolving structural distribution.

These ablation results show that: (i) The substantial performance decrease in the **w/o SAA** variant under pattern shifts indicates that relation-level anchors are essential for capturing structural evolution. (ii) The performance decline in the **w/o CMA** variant confirms that calibration modulation is necessary to filter noisy observations. (iii) The performance margins observed for **w/o HOA** and **w/o ROA** demonstrate

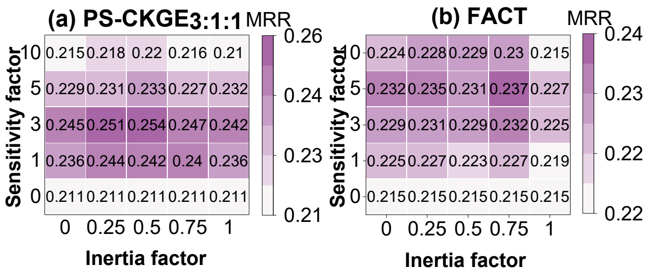

Figure 3: Hyperparameter sensitivity analysis of PAAL.

Dataset	LKGE			SAGE			**PAAL**		
	MRR	FWT	RTF	MRR	FWT	RTF	MRR	FWT	RTF
PS-CKGE$_{3:1:1}$	0.207	0.148	0.478	<u>0.212</u>	<u>0.159</u>	<u>0.481</u>	**0.254**	**0.185**	**0.496**
PS-CKGE$_{3:1:1}$(N)	0.181	0.126	0.473	<u>0.185</u>	<u>0.132</u>	<u>0.475</u>	**0.236**	**0.168**	**0.495**
Drop (↓)	-12.56%	-17.57%	-1.05%	-12.74%	-16.98%	-1.25%	-7.09%	-9.19%	-0.20%
FACT	0.210	0.176	0.481	<u>0.222</u>	<u>0.182</u>	<u>0.494</u>	**0.237**	**0.191**	**0.496**
FACT(N)	0.184	0.157	0.476	<u>0.195</u>	<u>0.163</u>	<u>0.490</u>	**0.218**	**0.177**	**0.495**
Drop (↓)	-12.38%	-10.80%	-1.04%	-12.16%	-10.44%	-0.81%	-8.02%	-7.33%	-0.20%

Table 5: Performance robustness comparison. Drop: Percentage of decline caused by noise (lower is better).

that applying alignment signals to both historical consistency and regularization facilitates stability on evolution scenarios, ensuring that historical constraints are enforced adaptively based on structural consistency.

5.4 Analysis of Hyperparameters (RQ3)

To investigate the influence of our anchor alignment mechanism, we examine its two key hyperparameters: **the sensitivity factor** β and **the inertia factor** λ_h. In the PAAL framework, these parameters are designed to guide the alignment process. Specifically, β adjusts the sharpness of the modulation function, influencing how the model distinguishes between historical patterns and new observations. Meanwhile, λ_h sets the baseline intensity for historical constraints, serving as a reference for the model's stability under evolution. As shown in Figure 3, the variations in optimal configurations suggest that pattern-shift scenarios (e.g., PS-CKGE) tend to favor settings promoting plasticity, whereas standard scenarios (e.g., FACT) generally benefit from higher inertia. This observation indicates that PAAL can tune these parameters aids the model in adapting to different degrees of structural evolution.

5.5 Analysis of Robustness and Efficiency (RQ3)

To evaluate practical applicability, we conduct a joint analysis of robustness and efficiency (Table 5 and Figure 4). We introduce structural noise via techniques such as missing links, head-tail entity inversion, and other perturbations. Regarding robustness against structural noise, baseline methods suffer significant degradation with MRR drops exceeding 12% due to their sensitivity to interference. In contrast, PAAL demonstrates considerable resilience, restricting the MRR decline to approximately 7%-8%, validating that our alignment mechanism effectively filters noise. In terms of efficiency, although PAAL incurs a marginal computational increase due to anchor calculation, the training time remains comparable to

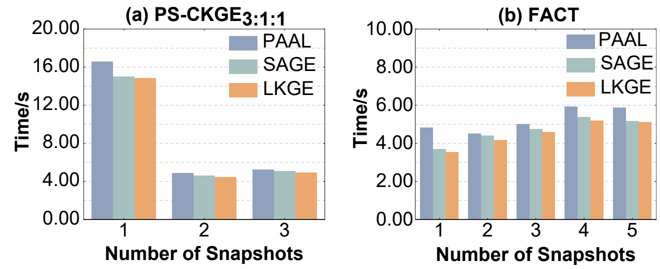

Figure 4: Runtime analysis of PAAL and baselines for a single epoch across various snapshots of different datasets.

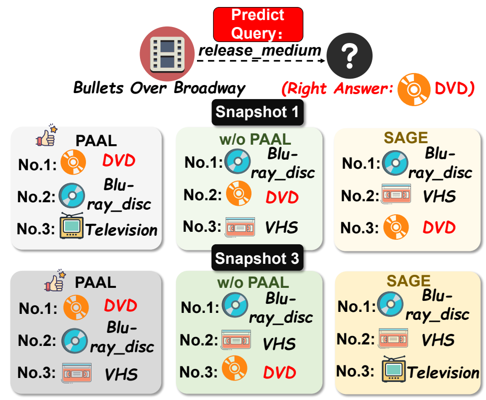

Figure 5: Case examples of PAAL.

baselines. This confirms that PAAL achieves strong gains in stability and accuracy with only a marginal increase in computational overhead.

5.6 Case Study

Figure 5 illustrates a case study for the query (`Bullets Over Broadway, release medium, ?)` in PS-CKGE$_{3:1:1}$ dataset. As shown, baseline models like SAGE and w/o PAAL incorrectly identify `Blu-ray disc` or `VHS` as the top prediction. In contrast, PAAL consistently identifies the right answer `DVD` as the Top-1 prediction in both snapshot 1 and snapshot 3. This indicates PAAL's effectiveness in the generalization of evolving knowledge, supporting consistent prediction across evolving snapshots.

6 Conclusion

In this paper, we address the challenge of structural distribution shifts in Continual Knowledge Graph Embedding (CKGE) by proposing **P**attern-**A**nchor **Al**ignment (PAAL). PAAL utilizes relational anchors to model structural distributions, aiming to achieve stable continual embedding through anchor alignment and optimization. Experimental results on both standard and pattern-shift benchmarks demonstrate that PAAL effectively adapts to structural distribution shifts in evolving scenarios. For future work, although efficiency analysis indicates marginal overhead, this cost constrains large-scale, latency-sensitive applications like real-time recommendation systems, necessitating future optimization.

Ethical Statement

There are no ethical issues.

In strict compliance with the conference's policy, Large Language Models (LLMs) were employed exclusively as a grammar checking and language refinement tool to ensure standard English usage. The model functioned solely to correct grammatical errors and improve phrasing. No part of the paper's text was generated from scratch by the AI, and the authors verify that the manuscript is their own original writing.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 62276196) and the Hubei Provincial Natural Science Foundation of China (No. 2026AFC0615).

References

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, pages 2787–2795, 2013.

[Cui *et al.*, 2023] Yuanning Cui, Yuxin Wang, Zequn Sun, Wenqiang Liu, Yiqiao Jiang, Kexin Han, and Wei Hu. Lifelong embedding learning and transfer for growing knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4217–4224, 2023.

[Daruna *et al.*, 2021] Angel Andres Daruna, Mehul Gupta, Mohan Sridharan, and Sonia Chernova. Continual learning of knowledge graph embeddings. *IEEE Robotics and Automation Letters*, 6(2):1128–1135, 2021.

[Dinan *et al.*, 2019] Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. Wizard of wikipedia: Knowledge-powered conversational agents. In *Proceedings of The 7th International Conference on Learning Representations*, 2019.

[Guo *et al.*, 2022] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3549–3568, 2022.

[Hamaguchi *et al.*, 2017] Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, and Yuji Matsumoto. Knowledge transfer for out-of-knowledge-base entities: a graph neural network approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1802–1808, 2017.

[Jian *et al.*, 2025a] Yue Jian, Xiangyu Luo, Zhifei Li, Miao Zhang, Yan Zhang, Kui Xiao, and Xiaoju Hou. APKGC: noise-enhanced multi-modal knowledge graph completion with attention penalty. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 15005–15013, 2025.

[Jian *et al.*, 2025b] Yue Jian, Miao Zhang, Ziyue Qin, Chuyuan Xie, Kui Xiao, Yan Zhang, and Zhifei Li. Adaptive modality interaction transformer for multimodal knowledge graph completion. *ACM Trans. Knowl. Discov. Data*, 19(8):158:1–158:24, 2025.

[Kirkpatrick *et al.*, 2017] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

[Kou *et al.*, 2020] Xiaoyu Kou, Yankai Lin, Shaobo Liu, Peng Li, Jie Zhou, and Yan Zhang. Disentangle-based continual graph representation learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2961–2972, 2020.

[Lan *et al.*, 2021] Yunshi Lan, Gaole He, Jinhao Jiang, Jing Jiang, Wayne Xin Zhao, and Ji-Rong Wen. A survey on complex knowledge base question answering: Methods, challenges and solutions. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 4483–4491, 2021.

[Lehmann *et al.*, 2015] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195, 2015.

[Li *et al.*, 2024] Xiner Li, Shurui Gui, Youzhi Luo, and Shuiwang Ji. Graph structure extrapolation for out-of-distribution generalization. In *Proceedings of the 41st International Conference on Machine Learning*, pages 27846–27874, 2024.

[Li *et al.*, 2025a] Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. Out-of-distribution generalization on graphs: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[Li *et al.*, 2025b] Yifei Li, Lingling Zhang, Hang Yan, Tianzhe Zhao, Zihan Ma, Muye Huang, and Jun Liu. SAGE: Scale-aware gradual evolution for continual knowledge graph embedding. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1600–1611, 2025.

[Li *et al.*, 2026] LinYu Li, Zhi Jin, Yuanpeng He, Dongming Jin, Yichi Zhang, Haoran Duan, Xuan Zhang, Zhengwei Tao, and Tashi Nyima. Learning to evolve: Bayesian-guided continual knowledge graph embedding. In Hakim Hacid, Yoelle Maarek, Francesco Bonchi, Ido Guy, and Emine Yilmaz, editors, *Proceedings of the ACM Web Conference 2026, WWW 2026*, pages 4372–4383, 2026.

[Lin *et al.*, 2022] Qika Lin, Jun Liu, Fangzhi Xu, Yudai Pan, Yifan Zhu, Lingling Zhang, and Tianzhe Zhao. Incorporating context graph with logical reasoning for inductive relation prediction. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 893–903, 2022.

[Liu *et al.*, 2021] Huihui Liu, Yiding Yang, and Xinchao Wang. Overcoming catastrophic forgetting in graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 8653–8661, 2021.

[Liu *et al.*, 2024a] Jiajun Liu, Wenjun Ke, Peng Wang, Ziyu Shang, Jinhua Gao, Guozheng Li, Ke Ji, and Yanhe Liu. Towards continual knowledge graph embedding via incremental distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8759–8768, 2024.

[Liu *et al.*, 2024b] Jiajun Liu, Wenjun Ke, Peng Wang, Jiahao Wang, Jinhua Gao, Ziyu Shang, Guozheng Li, Zijie Xu, Ke Ji, and Yining Li. Fast and continual knowledge graph embedding via incremental LoRA. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 2198–2206, 2024.

[Lomonaco and Maltoni, 2017] Vincenzo Lomonaco and Davide Maltoni. Core50: a new dataset and benchmark for continuous object recognition. In *Proceedings of the 1st Conference on Robot Learning*, pages 17–26, 2017.

[Lopez-Paz and Ranzato, 2017] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6467–6476, 2017.

[Pons *et al.*, 2025] Gerard Pons, Besim Bilalli, and Anna Queralt. Improving continual learning of knowledge graph embeddings via informed initialization. *arXiv preprint arXiv:2511.11118*, 2025.

[Rusu *et al.*, 2016] Andrei A. Rusu, Neil C. Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016.

[Song *et al.*, 2026] Tengwei Song, Long Yin, Zhen Han, and Zhiqiang Xu. Improving enzyme prediction with chemical reaction equations by hypergraph-enhanced knowledge graph embeddings. *arXiv preprint arXiv:2601.05330*, 2026.

[Sun *et al.*, 2019] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. RotatE: Knowledge graph embedding by relational rotation in complex space. In *Proceedings of the 7th International Conference on Learning Representations*, pages 1–18, 2019.

[Wang *et al.*, 2019] Hong Wang, Wenhan Xiong, Mo Yu, Xiaoxiao Guo, Shiyu Chang, and William Yang Wang. Sentence embedding alignment for lifelong relation extraction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 796–806, 2019.

[Wang *et al.*, 2024] Yuhan Wang, Qing Xie, Mengzi Tang, Lin Li, Jingling Yuan, and Yongjian Liu. Amazon-kg: A knowledge graph enhanced cross-domain recommendation dataset. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024,*, pages 123–130, 2024.

[Wu *et al.*, 2024] Qitian Wu, Fan Nie, Chenxiao Yang, Tianyi Bao, and Junchi Yan. Graph out-of-distribution generalization via causal intervention. In *Proceedings of the ACM Web Conference 2024*, pages 850–860, 2024.

[Zenke *et al.*, 2017] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3987–3995, 2017.

[Zhang *et al.*, 2023] Xikun Zhang, Dongjin Song, and Dacheng Tao. Hierarchical prototype networks for continual graph representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4622–4636, 2023.

[Zhao *et al.*, 2025] Tianzhe Zhao, Jiaoyan Chen, Yanchi Ru, Qika Lin, Yuxia Geng, Haiping Zhu, Yudai Pan, and Jun Liu. Rethinking continual knowledge graph embedding: Benchmarks and analysis. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 138–147, 2025.

[Zhou and Cao, 2021] Fan Zhou and Chengtai Cao. Overcoming catastrophic forgetting in graph neural networks with experience replay. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4714–4722, 2021.

[Zhu *et al.*, 2023] Zhenfang Zhu, Dianyuan Zhang, Lin Li, Kefeng Li, Jiangtao Qi, Wenling Wang, Guangyuan Zhang, and Peiyu Liu. Knowledge-guided multi-granularity GCN for ABSA. *Inf. Process. Manag.*, 60(2):103223, 2023.

[Zhu *et al.*, 2025a] Junlin Zhu, Bo Fu, and Guiduo Duan. DebiasedKGE: Towards mitigating spurious forgetting in continual knowledge graph embedding. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pages 4519–4528, 2025.

[Zhu *et al.*, 2025b] Lijing Zhu, Qizhen Lan, Qing Tian, Wenbo Sun, Li Yang, Lu Xia, Yixin Xie, Xi Xiao, Tiehang Duan, Cui Tao, and Shuteng Niu. ETT-CKGE: efficient task-driven tokens for continual knowledge graph embedding. In *Machine Learning and Knowledge Discovery in Databases. Research Track - European Conference, ECML PKDD 2025*, pages 481–496, 2025.