

Learning to Evolve: Bayesian-Guided Continual Knowledge Graph Embedding

Linyu Li¹, Zhi Jin^{1*}, Yuanpeng He¹, Dongming Jin¹, Yichi Zhang², Haoran Duan³, Nyima Tashi⁴

¹School of Computer Science, Peking University, Beijing, China

²College of Computer Science and Technology, Zhejiang University, Hangzhou

³School of Cyber Science and Engineering, Wuhan University

⁴School of Information Science and Technology, Tibet University

* corresponding author

Abstract

Since knowledge graphs (KG) will continue to evolve in real scenarios, traditional KGE models are only suitable for static knowledge graphs. Therefore, continual knowledge graph embedding (CKGE) has attracted the attention of researchers. Currently, a key challenge facing CKGE is that the model is prone to "catastrophic forgetting", resulting in the loss of previously learned knowledge. In order to effectively alleviate this problem, we propose a new CKGE model BAKE. First, we note that the Bayesian posterior update principle provides a natural continual learning strategy that is insensitive to data order and can theoretically effectively resist the forgetting of previous knowledge during data evolution. Different from the existing CKGE method, BAKE regards each batch of new data as a Bayesian update of the model prior. Under this framework, as long as the posterior distribution of the model is maintained, the model can better preserve the knowledge of early snapshots even after evolving through multiple time snapshots. Secondly, we propose a continual clustering method for CKGE, which further directly combats knowledge forgetting by constraining the evolution difference (or change amplitude) between new and old knowledge between different snapshots. We conduct extensive experiments on BAKE on multiple datasets, and the results show that BAKE significantly outperforms existing baseline models.

Introduction

Knowledge graphs (KGs)(Liang et al. 2024b)(Ji et al. 2021) have become indispensable tools for representing structured information, supporting a wide range of knowledge-driven applications such as LLM based question answering(Pan et al. 2024)(Xu et al. 2025) and recommender systems(Zhao et al. 2024)(Yuan et al. 2025b). Knowledge graph embedding (KGE) aims to encode entities and relations in a KG as continual vectors, and the quality of these embeddings directly affects downstream performance. Classic KGE models (Bordes et al. 2013)(Li et al. 2023)(Sun et al. 2019)(Xiao, He, and Cao 2023)(Long et al. 2024)(Li et al. 2025c) perform excellently on static KGs, yet real-world graphs usually evolve continuously. For example, the YAGO knowledge base evolved from YAGO3(Mahdisoltani, Biega, and Suchanek 2013) to YAGO4.5(Suchanek et al. 2024), with the number of entities soaring from about 4.6 million to 49 million and facts (triples) from 40 million to 132 million. For such continually evolving knowledge graphs (CKGs),

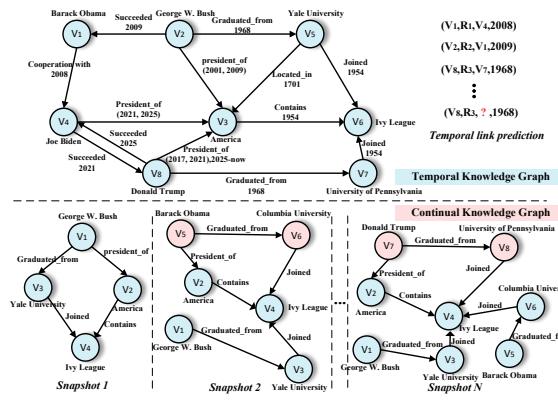


Figure 1: TKG typically employs triplets combined with timestamps to explicitly annotate temporal information on all relational edges, exhibiting a strong temporal dimension. In contrast, CKG decomposes real KGs into a series of time unmarked evolutionary snapshots, focusing on continuously updating and transferring learned knowledge as snapshots evolve.

existing KGE models must retrain the entire graph after each update—a computationally prohibitive and impractical strategy for large-scale graphs like YAGO. Hence, this work focuses on CKGE models capable of learning CKGs efficiently.

Recent CKGE studies optimize along three axes: training structure, parameter efficiency, and regularization and masking mechanisms. Structure: IncDE(Liu et al. 2024a) employs a hierarchical strategy with incremental distillation to exploit graph structure. Parameter efficiency: FastKGE(Liu et al. 2024b) reduces trainable parameters via incremental low-rank adapters, whereas ETT-CKGE(Zhu et al. 2025) leverages task-driven tokens for efficient knowledge transfer. Regularization and masking model: existing approaches include lifelong learning based on masked auto-encoders(Cui et al. 2023), energy-based alignment of old and new knowledge(Cao 2024), a bio-inspired dual-mask mechanism(Song et al. 2024), and flexible regularization grounded on the Fisher information matrix(Zhu et al. 2024). However, these methods largely rely on heuristic regularizers or masks to passively relieve forgetting and

have not yet actively learned from the perspective of continuous data evolution by letting the prior guide the posterior.

As illustrated in Figure 1, CKGs differ from temporal KGs(Zhang et al. 2025)(Wang et al. 2023), whose time-stamped relations explicitly model temporal dependencies. CKGs instead focus on structural increments, such as newly emerging entities or relations(Cui et al. 2023). We observe that, because a CKG lacks explicit temporal semantics, its evolution naturally aligns with the sequential update mechanism of the Bayesian framework. According to Bayes’ rule(Lee et al. 2024)(Bonnet et al. 2025), we treat the posterior distribution of snapshot t as the prior for snapshot $t + 1$, thereby accumulating knowledge in a principled manner. This sequential update is order-insensitive, effectively resists forgetting during evolution, and—through uncertainty quantification(Ahn et al. 2019) (e.g., probabilistic weight distributions)—balances learning and forgetting, avoids catastrophic memory, retains critical knowledge, and gradually discards obsolete information.

Moreover, without constraints, representations of entities and relations may drift across snapshots, exacerbating catastrophic forgetting because the model lacks guidance on how knowledge should evolve while remaining consistent with historical representations. To address this, we propose Continual Clustering, which introduces a contrastive clustering regularizer that keeps entity embeddings in compact clusters during evolution, ensuring that an entity “remembers who it is” and “whom it is related to,” so that semantically similar entities maintain their relative positions over time. Our contributions are as follows:

- We propose BAKE, a Bayesian-guided continual knowledge graph embedding approach that formulates CKGE as a sequential Bayesian inference problem, providing theoretical guarantees against catastrophic forgetting.
- We introduce a continual clustering approach that constrains the evolution of knowledge between KG snapshots, preserving semantic consistency while allowing controlled adaptation to new knowledge.
- We conduct extensive experiments on multiple CKGE benchmarks and demonstrate that BAKE significantly outperforms existing baseline models in terms of both knowledge preservation and adaptability.

Related Work

Knowledge Graph Embedding

Early KGE models mainly follow the translation paradigm, such as TransE(Bordes et al. 2013) and TransH(Wang et al. 2014) in the real vector space, and their extensions to complex/quaternion spaces, including RotatE(Sun et al. 2019), QuatE(Zhang et al. 2019), HAKE(Zhang et al. 2020), MRME(Li et al. 2025b), and GIE(Cao et al. 2022). These approaches introduced the idea of representing a triple (h, r, t) through simple algebraic operations in a vector space. Since then, both the efficiency and the expressive power of KGE models have improved along multiple dimensions. RecPiece(Liang et al. 2024a) clusters relation centres to select anchors, reducing memory consumption while

improving link-prediction accuracy; RAA-KGC(Yuan et al. 2025a) pulls queries of a pretrained language model toward relation-aware neighbour anchors to construct more discriminative embeddings; NFE(Xiao, He, and Cao 2023) treats entities and relations as permutations of random variables, unifying point and probabilistic embeddings under group theory and enabling logical reasoning with uncertainty. Recently, KGE research has begun to leverage diffusion models to recast the task as conditional entity generation. For example, KGDM(Long et al. 2024) trains a diffusion model that outputs the distribution of tail entities conditioned on the head entity and relation, thereby capturing multimodal relational semantics without relying on complicated probabilistic mechanisms.S2DN(Ma et al. 2025) proposes a semantic-aware denoising network and filters out unreliable interactions to enhance the structural reliability of target links.

Continual Knowledge Graph Embedding

As a knowledge graph (KG) evolves, embedding models must learn new facts while avoiding catastrophic forgetting of previously acquired knowledge. Consequently, research on continual KGE (CKGE) has advanced rapidly in recent years. IncDE(Liu et al. 2024a) orders new triples hierarchically and employs layer-wise incremental distillation to preserve old knowledge; FastKGE(Liu et al. 2024b) accelerates model updates with an adaptive low-rank adapter (IncLoRA), substantially shortening training time while isolating parameters for new snapshots; LKGE(Cui et al. 2023) proposes a masked auto-encoder combined with transfer regularisation to extend embeddings across multiple snapshots, enabling lifelong learning; ETT-CKGE(Zhu et al. 2025) replaces costly node scoring with lightweight task-driven labels and aligns CKGs through simple matrix operations; CLKGE(Cao 2024) couples continual learning with an energy-based manifold so that new and old knowledge reinforce each other, whereas CMKGE(Song et al. 2024) introduces a biologically inspired dual inhibition excitation mask to balance plasticity and stability; FMR(Zhu et al. 2024) rotates the parameter space based on Fisher information and assigns flexible regularisation to emulate the refinement of human memory and further mitigate forgetting.

Preliminaries

Continual knowledge graph

A Continual Knowledge Graph (CKG) is represented as a series of snapshots $\mathcal{G}_c = \mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_N$, where each snapshot $\mathcal{S}_i = \mathcal{E}_i, \mathcal{R}_i, \mathcal{T}_i$ describes the state of the knowledge graph at time i . Each triple in \mathcal{S}_i is represented as $(e_h, r, e_t) \in \mathcal{T}_i$, where e_h and e_t are the head and tail entities drawn from \mathcal{E}_i , and r is a relation from the relation set \mathcal{R}_i . This study focuses on scenarios where relations remain constant over time (i.e., $\mathcal{R}_i = \mathcal{R}$ for all i), but entities expand gradually. The new entities and triples at time i are denoted by $\Delta\mathcal{E}_i = \mathcal{E}_i \setminus \mathcal{E}_{i-1}$ and $\Delta\mathcal{T}_i = \mathcal{T}_i \setminus \mathcal{T}_{i-1}$, respectively. This setting aligns with typical real-world scenarios and is applied in current continual knowledge graph embedding methods.

Continual Knowledge Graph Embedding

The goal of Knowledge Graph Embedding is to embed entities and relations into a low-dimensional vector space while preserving their semantic meanings. For a CKG $\mathcal{G}_c = \mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_N$, CKGE progressively builds an embedding model \mathcal{M}_i as new snapshots \mathcal{S}_i arrive. It leverages prior embeddings from snapshots $\mathcal{S}0, \dots, \mathcal{S}_{i-1}$, optimizing training efficiency and overall performance. Completely retraining \mathcal{M}_i from scratch on all triples in \mathcal{S}_i is a basic yet inefficient choice unsuitable for practical scenarios. Current CKGE methods adapt the existing model \mathcal{M}_{i-1} to new data \mathcal{M}_i by creating embeddings for new entities $\Delta\mathcal{E}_i$ and refining embeddings for existing relations \mathcal{R} and prior entities \mathcal{E}_{i-1} when receiving new triples $\Delta\mathcal{T}_i$. CKGE is evaluated through continual link prediction: given an incomplete triple such as $(h, r, ?)$, the model scores each candidate tail entity and outputs the highest-ranked one, maintaining accuracy while learning new knowledge and remembering past knowledge.

Continual learning and Bayesian neural networks

Continual learning (CL) aims to train a neural network on a sequence of tasks $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_T$ without forgetting previously acquired knowledge. In standard CL, the model parameters θ are updated by stochastic gradient descent, which often overwrites older information.

Bayesian neural networks (BNNs) (Bonnet et al. 2025) provide a probabilistic framework for CL by treating the parameters θ as random variables with a prior distribution $p(\theta)$. The goal is to compute the posterior distribution $p(\theta | \mathcal{D})$ after observing data \mathcal{D} , according to Bayes' rule: $p(\theta | \mathcal{D}) = \frac{p(\mathcal{D} | \theta) p(\theta)}{p(\mathcal{D})}$. Here, $p(\mathcal{D} | \theta)$ is the likelihood, and $p(\mathcal{D})$ is the marginal likelihood. For a task sequence, the posterior can be updated recursively as: $p(\theta | \mathcal{D}_{1:t}) \propto p(\mathcal{D}_t | \theta) p(\theta | \mathcal{D}_{1:t-1})$, meaning that the previous posterior serves as the new prior.

Exact inference is usually intractable for deep networks, so an independent Gaussian (mean-field) approximation is commonly adopted (Lee et al. 2024; Ahn et al. 2019): $q_t(\omega) = \prod_{i=1}^P \mathcal{N}(\omega_i; \mu_{t,i}, \sigma_{t,i}^2)$, with $\theta_t = (\mu_t, \sigma_t)$ taken as the learnable variational parameters. The independence assumption greatly simplifies the KL-divergence computation and the reparameterization gradient estimator, enabling iterative posterior updates. In practice, the reparameterization trick $\omega_i = \mu_{t,i} + \sigma_{t,i} \epsilon$, $\epsilon \sim \mathcal{N}(0, 1)$ is often employed to estimate gradients.

Methodology

Framework Overview

The overall framework of BAKE is shown in Figure 2. Its core idea is to formalize the continual knowledge graph embedding (CKGE) problem as a serialized Bayesian inference process. Given a knowledge graph sequence $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_T$ that evolves over time, the model receives a set of newly added triples $\Delta\mathcal{T}_t$ at each time snapshot t . BAKE works together through the following three modules: First, Bayesian-guided knowledge evolution uses the parameter posterior

distribution of the previous snapshot $t - 1$ as a prior, integrates new knowledge and quantifies uncertainty through Bayesian updating; second, the continual clustering module introduces fairness regularization to maintain the semantic consistency of the embedding space; finally, the joint optimization objective integrates the above modules to achieve a balance between new knowledge learning and old knowledge retention. This framework aims to continually accumulate knowledge and alleviate catastrophic forgetting guided by the Bayesian rule.

Bayesian-Guided Knowledge Evolution Learning

Inspired by related work such as uncertainty regularized (Li et al. 2025a) continual learning (Ahn et al. 2019) (Lee et al. 2024) (Bonnet et al. 2025), we regard the embedding parameters of entities and relations as random variables in probability distributions rather than fixed point estimates. This approach not only facilitates quantification of parameter uncertainty but also protects important old knowledge when learning new knowledge, avoiding catastrophic forgetting.

Entity and Relation Distributional Representations

For each entity e and relation r , their embedding vectors at time snapshot t are modeled as Gaussian distributions. The embedding of entity e is represented as $\mathbf{e}_t \sim \mathcal{N}(\boldsymbol{\mu}_{e,t}, (\text{diag}(\boldsymbol{\lambda}_{e,t}))^{-1})$, where $\boldsymbol{\mu}_{e,t} \in \mathbb{R}^d$ is the mean vector, and $\boldsymbol{\lambda}_{e,t} \in \mathbb{R}^d$ is the diagonal precision vector (inverse of variance); the embedding of relation r is similar, represented as $\mathbf{r}_t \sim \mathcal{N}(\boldsymbol{\mu}_{r,t}, (\text{diag}(\boldsymbol{\lambda}_{r,t}))^{-1})$. Each element of the precision $\boldsymbol{\lambda}$ reflects the certainty of the corresponding dimension: a higher value indicates more reliable knowledge in that dimension, which should be better protected in subsequent updates.

Sequential Bayesian Update At time snapshot t , we use the posterior distribution $\{\boldsymbol{\mu}_{t-1}, \boldsymbol{\lambda}_{t-1}\}$ from time $t - 1$ as the prior and perform online Bayesian update using new data $\Delta\mathcal{T}_t$. To simplify inference, the embeddings $\hat{\Theta}_t$ obtained from training on the current snapshot are treated as new observations to the prior mean $\boldsymbol{\mu}_{t-1}$, generated by the classic KGE model TransE (Bordes et al. 2013). TransE evaluates the plausibility of triples through a scoring function and optimizes the embeddings $\hat{\Theta}_t = \{\hat{\mathbf{e}}_t, \hat{\mathbf{r}}_t\}$ by minimizing the margin-based ranking loss \mathcal{L}_{KGE} :

$$\mathcal{L}_{\text{KGE}} = \sum_{(h, r, t) \in \mathcal{T}} \sum_{(h', r, t') \in \mathcal{T}} [\gamma + f(h, r, t) - f(h', r, t')], \quad (1)$$

where $f(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2$ represents the score function of TransE (Bordes et al. 2013). Based on the conjugate property of Gaussian distributions, the update rules for the embedding parameters of entity e are:

$$\boldsymbol{\lambda}_{e,t} = \boldsymbol{\lambda}_{e,t-1} + \lambda_{obs}, \quad (2)$$

$$\boldsymbol{\mu}_{e,t} = \frac{\boldsymbol{\lambda}_{e,t-1} \odot \boldsymbol{\mu}_{e,t-1} + \lambda_{obs} \odot \hat{\mathbf{e}}_t}{\boldsymbol{\lambda}_{e,t}}, \quad (3)$$

where $\hat{\mathbf{e}}_t$ is the entity embedding trained by TransE, λ_{obs} is the fixed observation precision that controls the influence of new knowledge, and \odot denotes element-wise multiplication. New entities are initialized with uninformative priors

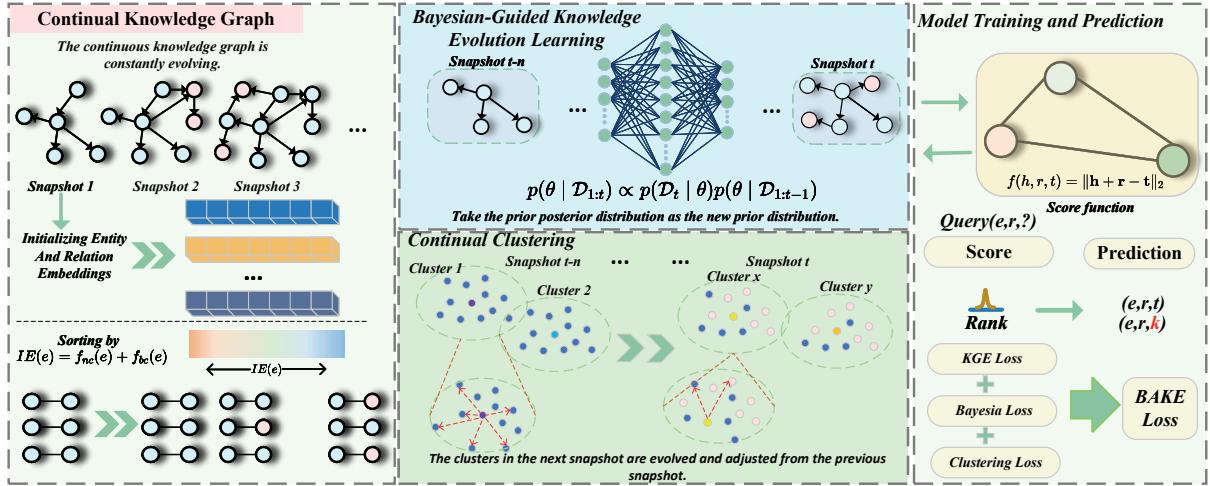


Figure 2: An overview of our proposed BAKE framework.

(random means, small precisions), and relation embeddings are updated in the same way. In experiments, we treat the value of λ_{obs} as a tunable hyperparameter of the model to balance the importance of new and old knowledge.

To guide the embeddings towards the posterior mean, we use a regularization term (Lee et al. 2024) \mathcal{L}_{Bayes} , which is equivalent to minimizing the KL divergence between the current distribution $q(\Theta_t)$ and the target posterior:

$$\mathcal{L}_{Bayes} = \sum_{e \in \mathcal{E}_t \cup \mathcal{R}_t} \beta \cdot \left\| \sqrt{\lambda_{i,t-1}} \odot (\hat{\theta}_{i,t} - \mu_{i,t-1}) \right\|_2^2, \quad (4)$$

where $\hat{\theta}_{i,t}$ is the current embedding, and β is the regularization hyperparameters. This term is weighted by the precision λ_{t-1} , limiting changes to highly certain knowledge, thereby preserving old knowledge.

Continual Clustering

Bayesian updates prevent forgetting at the parameter level, but the evolution of knowledge representations between KG snapshots needs to be constrained; otherwise, entity and relation embeddings in the latent space may undergo uncertainty drift across snapshots. Inspired by contrastive learning and clustering methods (Liang et al. 2024a)(Truong et al. 2025)(Nguyen et al. 2021), we propose a continual clustering method as a constraint to maintain the geometric structural consistency of the embedding space at the semantic level.

Sequential Contrastive Clustering Learning In continual learning, imbalances in data volume and class distribution across temporal snapshots may lead to under-learning of minority classes. To address this, we propose the sequential contrastive clustering method (\mathcal{L}_{FCC}), which achieves fairness by dynamically adjusting the contributions of major and minor classes. Specifically, for each cluster k , we maintain a centroid vector $\mathbf{c}_k \in \mathbb{R}^d$ (where d is the embedding dimension) to represent the central features of the cluster. First, we acknowledge the perspective proposed in (Liu et al. 2024a) that the learning order is also crucial in the process

of learning KG structures. Second, since new entities emerge in each snapshot, making it difficult to apply k-means clustering, we sort the entities based on their importance, then fix the size of each cluster k , and take the mean of the embeddings in the current cluster k as the clustering centroid \mathbf{c}_k . Specifically, the sorting process is as follows: we comprehensively consider the node’s centrality and betweenness centrality in the graph structure to compute its importance score. The calculation formulas for these two metrics are as follows:

$$f_{nc}(e) = \frac{f_{neighbor}(e)}{N-1}, \quad f_{bc}(e) = \sum_{s,t \in \mathcal{E}, s \neq t} \frac{\sigma(s,t|e)}{\sigma(s,t)} \quad (5)$$

where $f_{nc}(e)$ reflects the tightness of local connections for the entity, and $f_{bc}(e)$ measures its bridging role in information propagation. The variables in the formulas are defined as follows: $f_{neighbor}(e)$ represents the number of neighbors of entity e , N is the total number of entities in the current knowledge graph snapshot, \mathcal{E} is the set of entities in the current time snapshot i , $\sigma(s,t)$ is the total number of shortest paths between entities s and t , and $\sigma(s,t|e)$ is the number of those paths that pass through entity e . We combine the two to define the importance score for entity e as: $IE(e) = f_{nc}(e) + f_{bc}(e)$. We calculate the IE score for each entity and sort them in descending order of scores. Then, we assign the sorted entities to clusters of fixed size in sequence, and compute the initial embedding for each cluster as $\mathbf{c}_k = \frac{1}{|K|} \sum_{e_i \in K} \mathbf{e}_i$. The continual clustering then constrains the representations of entities across different snapshots through contrastive loss, with the loss function defined as:

$$\mathcal{L}_{FCC} = - \sum_{k=1}^K \left(\alpha_k \sum_{\mathbf{e}_i \in \text{Cluster}_k} \mathcal{L}_{cont}(\mathbf{e}_i, \mathbf{c}_k) + \mathcal{L}_{cont}(\mathbf{v}_k, \mathbf{c}_k) \right) \quad (6)$$

where \mathbf{e}_i is the entity embedding vector belonging to cluster k , $\text{sim}(\cdot, \cdot)$ is the cosine similarity function, τ is the temperature parameter (controlling the smoothness of softmax), K

Dataset	Snapshot 0			Snapshot 1			Snapshot 2			Snapshot 3			Snapshot 4		
	N_E	N_R	N_T	N_E	N_R	N_T	N_E	N_R	N_T	N_E	N_R	N_T	N_E	N_R	N_T
ENTITY	2,909	233	46,388	5,817	236	72,111	8,275	236	73,785	11,633	237	70,506	14,541	237	47,326
RELATION	11,560	48	98,819	13,343	96	93,535	13,754	143	66,136	14,387	190	30,032	14,541	237	21,594
HYBRID	8,628	86	57,561	10,040	102	20,873	12,779	151	88,017	14,393	209	103,339	14,541	237	40,326
GraphEqual	2,908	226	57,636	5,816	235	62,023	8,724	237	62,023	11,632	237	62,023	14,541	237	66,411
GraphHigher	900	197	10,000	1,838	221	20,000	3,714	234	40,000	7,467	237	80,000	14,541	237	160,116
GraphLower	7,505	237	160,000	11,258	237	80,000	13,134	237	40,000	14,072	237	20,000	14,541	237	10,116
FB-CKGE	7,505	237	186,070	11,258	237	31,012	13,134	237	31,012	14,072	237	31,012	14,541	237	31,010
WN-CKGE	24,567	11	55,801	28,660	11	9,300	32,754	11	9,300	36,848	11	9,300	40,943	11	9,302

Table 1: The statistics of datasets. N_E , N_R and N_T denote the number of entities and relations, and the current snapshot i

is the total number of clusters (equal to the number of entity classes in the current snapshot), and $\mathcal{L}_{\text{cont}}(\mathbf{e}_i, \mathbf{c}_k)$ is the standard contrastive loss, defined as:

$$\mathcal{L}_{\text{cont}}(\mathbf{e}_i, \mathbf{c}_k) = -\log \frac{\exp(\text{sim}(\mathbf{e}_i, \mathbf{c}_k) / \tau)}{\sum_{j=1}^K \exp(\text{sim}(\mathbf{e}_i, \mathbf{c}_j) / \tau)}. \quad (7)$$

The scaling factor α_k is an adjustable parameter for each cluster, used to balance inter-cluster contributions (e.g., dynamically set as $\alpha_k = 1/N_k$ based on the cluster sample size N_k to boost the weight of minority classes). The learnable vector $\mathbf{v}_k \in \mathbb{R}^d$ serves as the proxy vector for cluster k , used to optimize the cluster center position, thereby balancing class contributions and promoting cluster compactness. This loss, by minimizing \mathcal{L}_{FCC} , pulls embeddings of entities in the same cluster closer to the centroid while pushing apart those from different clusters, achieving semantic consistency. Hyperparameters include the scaling factor α_k , the temperature parameter τ , and the cluster sample size N_k .

Cluster Maintenance and Update At snapshot t , the centroids \mathbf{c}_k of old classes are inherited and fixed from $t-1$, serving as anchors for old knowledge. New classes initialize centroids via feature means. During training, entities are dynamically assigned to the nearest cluster, and centroids are slowly adjusted through a momentum update:

$$\mathbf{c}_{k,t} = (1 - \eta)\mathbf{c}_{k,t-1} + \eta \cdot \text{mean}(\{\mathbf{e}_i \in \text{Cluster}_k\}), \quad (8)$$

where η is a hyperparameter balancing the stability and adaptability of cluster centroids.

Final Training Objective

The optimal objective of BAKE at snapshot t is the combined loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{KGE}} + \mathcal{L}_{\text{Bayes}} + \mathcal{L}_{\text{FCC}}. \quad (9)$$

Experiments

Experimental Setup

Datasets To thoroughly verify the effectiveness and stability of BAKE, we conducted extensive experiments on a wide variety of datasets, including: ENTITY (Cui et al. 2023), RELATION (Cui et al. 2023), HYBRID (Cui et al. 2023), GraphLower (Liu et al. 2024a), GraphEqual (Liu et al. 2024a), GraphHigher (Liu et al. 2024a), FB-CKGE (Liu et al. 2024b), and WN-CKGE (Liu et al. 2024b). These eight datasets encompass different evolution patterns, each with

unique characteristics. Detailed statistics of the datasets are listed in Table 1. Each dataset contains 5 snapshots.

Implementation Details All experiments are implemented in PyTorch(Paszke et al. 2019) and conducted on 8 NVIDIA Tesla V100 GPUs. We tune all hyperparameters via grid search, including embedding dimensions for entities and relations 50, 100, 200, 300, batch sizes 256, 512, 1024, and number of clusters K 256, 512, 1024, among others. We use Adam as the optimizer with learning rates selected from 1e-5, 1e-4, 1e-3. To ensure fairness, all reported results are averaged over five runs.

Baselines To fully demonstrate the effectiveness of BAKE, we compare it with all recent CKGE models, including Fin-tune (Cui et al. 2023) (based on fine-tuning with new triples), InDE (Liu et al. 2024a) and FastKGE (Liu et al. 2024b) (based on incremental distillation and LoRA fine-tuning), LKGE (Cui et al. 2023), EWC (Kirkpatrick et al. 2017), SI (Zenke, Poole, and Ganguli 2017), DiCGRL (Kou et al. 2020), EMR (Wang et al. 2019), CWR (Lopez-Paz and Ranzato 2017), PNN (Rusu et al. 2016), FMR (Zhu et al. 2024), as well as energy-based and fine-grained token-driven models ETT-CKGE (Zhu et al. 2025) and CLKGE (Cao 2024).

Metrics We evaluate link prediction performance using MRR, Hits@1, Hits@3, and Hits@10, where higher is better. Following (Bordes et al. 2013) for negative sampling, we rank candidate triples by replacing the head or tail entity. For each snapshot i , we average scores over the current and all previous test sets. Results are reported using the fully trained model.

Results and Analysis

Main Results and Analysis. The experimental results are shown in Table 2 and Table 3. Overall, our proposed BAKE achieves the best performance on all datasets, demonstrating the effectiveness of the Bayesian-guided continual learning framework in knowledge graph embedding tasks. First, BAKE achieves the highest MRR scores on all eight datasets, with significant improvements over the strongest baseline methods. Specifically, on the four datasets shown in Table 2, BAKE improves MRR by 10.3% (ENTITY), 4.0% (RELATION), 10.8% (FB-CKGE), and 14.7% (WN-CKGE) compared to the next best method (mainly IncDE). Similar improvements are observed in Hits@1, Hits@3, and Hits@10. Notably, on the large-scale initial KG dataset FB-CKGE, BAKE achieves an MRR of 0.257, a 10.8% improve-

Model	ENTITY				RELATION				FB-CKGE				WN-CKGE			
	MRR	H@1	H@3	H@10												
PNN	0.229	0.13	0.265	0.425	0.167	0.096	0.191	0.305	0.215	0.122	0.245	0.403	0.134	0.002	0.241	0.342
CWR	0.088	0.028	0.114	0.202	0.021	0.01	0.024	0.043	0.075	0.011	0.105	0.192	0.005	0.002	0.007	0.012
GEM	0.165	0.085	0.188	0.321	0.093	0.04	0.106	0.196	0.188	0.103	0.212	0.359	0.119	0.002	0.215	0.297
EMR	0.171	0.090	0.195	0.33	0.111	0.052	0.126	0.225	0.180	0.097	0.204	0.346	0.114	0.002	0.205	0.286
DiCGRL	0.107	0.057	0.110	0.211	0.133	0.079	0.147	0.241	0.149	0.091	0.16	0.261	0.057	0.001	0.155	0.166
SI	0.154	0.072	0.179	0.311	0.113	0.055	0.131	0.224	0.187	0.102	0.211	0.359	0.115	0.001	0.209	0.289
EWC	0.229	0.130	0.264	0.423	0.165	0.093	0.190	0.306	0.218	0.124	0.247	0.410	0.136	0.003	0.248	0.338
LK	0.234	0.136	0.269	0.425	0.192	0.106	0.219	0.366	0.208	0.113	0.238	0.403	0.144	0.007	0.259	0.347
FastKGE	0.239	0.146	0.271	0.427	0.185	0.107	0.213	0.359	0.223	0.131	0.257	0.405	0.159	0.015	0.287	0.356
IncDE	0.253	0.151	0.291	0.448	0.199	0.110	0.221	0.368	0.232	0.134	0.271	0.428	0.150	0.003	0.278	0.366
BAKE	0.279	0.172	0.326	0.484	0.207	0.126	0.239	0.377	0.257	0.157	0.299	0.451	0.172	0.019	0.295	0.404

Table 2: Performance comparison of various models on ENTITY, RELATION, FB-CKGE, and WN-CKGE datasets. The best results are in **bold**, and the second-best results are underlined.

Model	GraphLower			GraphEqual			GraphHigher			HYBRID		
	MRR	H@1	H@10									
Fine-tune	0.185	0.098	0.363	0.183	0.096	0.358	0.198	0.108	0.375	0.135	0.069	0.262
PNN	0.213	0.119	0.407	0.212	0.118	0.405	0.186	0.097	0.364	0.185	0.101	0.349
CWR	0.032	0.005	0.080	0.122	0.041	0.277	0.189	0.096	0.374	0.037	0.015	0.077
GEM	0.170	0.084	0.346	0.189	0.099	0.372	0.197	0.109	0.372	0.136	0.070	0.263
EMR	0.188	0.101	0.362	0.185	0.099	0.359	0.202	0.113	0.379	0.141	0.073	0.267
DiCGRL	0.102	0.039	0.222	0.104	0.040	0.226	0.116	0.041	0.242	0.149	0.083	0.277
SI	0.186	0.099	0.366	0.179	0.092	0.353	0.190	0.099	0.371	0.111	0.049	0.229
EWC	0.210	0.116	0.405	0.207	0.113	0.400	0.198	0.106	0.385	0.186	0.102	0.350
LKGE	0.210	0.116	0.403	0.214	0.118	0.407	0.207	0.120	0.382	0.207	0.121	0.379
IncDE	0.228	0.129	0.426	0.234	0.134	0.432	0.227	0.132	0.412	0.224	0.131	0.401
BAKE	0.247	0.147	0.448	0.257	0.154	0.463	0.241	0.142	0.439	0.228	0.146	0.419

Table 3: Performance comparison of various models on GraphLower, GraphEqual, GraphHigher, and HYBRID datasets.

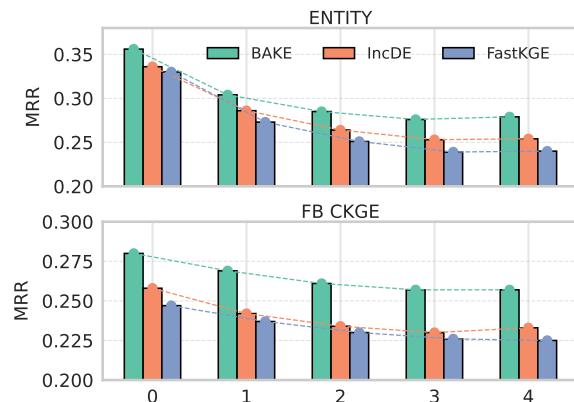


Figure 3: The change of model performance with the evolution of snapshots under ENTITY and FB-CKGE datasets.

ment over IncDE’s 0.232. These results strongly support that our Bayesian-guided framework can effectively suppress catastrophic forgetting while learning new knowledge.

Secondly, BAKE outperforms all CKGE baselines in most cases. Compared to dynamic architecture methods (PNN and CWR), BAKE improves MRR by 1.7%–19.8%, Hits@1 by 1.4%–13.0%, and Hits@10 by 2.3%–34.5%; compared to replay-based methods (GEM, EMR, and DiCGRL), MRR improves by 2.8%–14.7%, Hits@1 by 2.3%–9.5%, and Hits@10 by 3.8%–23.9%; compared to regularization-based methods (SI, EWC, and LKGE), MRR improves

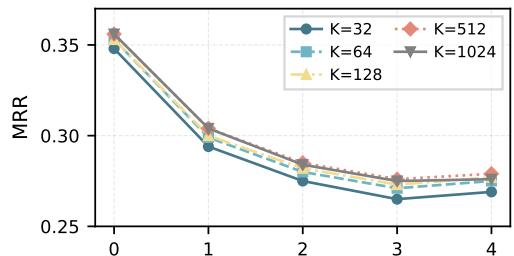


Figure 4: Parameter Sensitivity of BAKE to Cluster Number K under Different Snapshots on the ENTITY Dataset.

by 0.9%–11.0%, Hits@1 by 0.8%–8.3%, and Hits@10 by 0.7%–17.0%. BAKE also exhibits strong robustness and adaptability on datasets with different knowledge growth patterns. On balanced growth datasets (e.g., GraphEqual) and unbalanced ones (e.g., GraphHigher and GraphLower), BAKE consistently delivers optimal performance. Compared to the second-best baselines, MRR improves by 9.8%, 6.2%, and 8.3%, respectively, on these three datasets. This indicates that our proposed continual clustering constraint mechanism can effectively maintain semantic consistency in the embedding space, regardless of the evolution pattern of the KG.

Finally, comparing our method with baselines from different technical paradigms, BAKE consistently shows superior performance. Whether against dynamic architecture-based methods (PNN, CWR), replay-based methods (GEM, EMR,

Model	ENTITY				RELATION				FB-CKGE				WN-CKGE			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
BAKE	0.279	0.172	0.326	0.484	0.207	0.126	0.239	0.377	0.257	0.157	0.299	0.451	0.172	0.019	0.295	0.404
w/o Bayesian	0.259	0.165	0.301	0.458	0.196	0.119	0.223	0.354	0.236	0.151	0.279	0.422	0.159	0.018	0.272	0.382
w/o cluster	0.264	0.153	0.313	0.476	0.197	0.113	0.228	0.368	0.240	0.137	0.287	0.446	0.161	0.017	0.280	0.398

Table 4: Ablation study results on ENTITY, RELATION, FB-CKGE, and WN-CKGE datasets.

DiCGRL), or regularization-based methods (SI, EWC, LKGE), BAKE achieves the best results across all metrics. This strongly suggests that framing continual KGE as a sequential Bayesian inference problem, enhanced with explicit semantic consistency constraints, is a more fundamental and effective approach than existing solutions.

Ablation Results and Analysis. In ablation studies conducted on the ENTITY, RELATION, FB-CKGE, and WN-CKGE datasets, removing the Bayesian (Use the distillation ideas from IncDE) resulted in significant overall performance degradation, particularly in Hits@10 and MRR metrics. This highlights its crucial role in preserving prior knowledge through sequential posterior updates and mitigating catastrophic forgetting across evolving snapshots. By contrast, removing the continual clustering mechanism led to an even steeper decline in Hits@1 accuracy. This outcome demonstrates its effectiveness in maintaining semantic consistency and constraining embedding drift for precise link prediction. The integration of these components within BAKE achieves the optimal balance, showcasing synergistic benefits in adaptability and knowledge retention.

Case 1: (<i>Holland Taylor, profession, ?</i>)	
BAKE	Correct tail rank: "Actor" : <u>1</u>
IncDE	Correct tail rank: "Actor" : <u>2</u>
FastKGE	Correct tail rank: "Actor" : <u>5</u>
Case 2: (<i>Blade Runner, genre, ?</i>)	
BAKE	Correct tail rank: "Drama" : <u>1</u>
IncDE	Correct tail rank: "Drama" : <u>26</u>
FastKGE	Correct tail rank: "Drama" : <u>3</u>
Case 3: (<i>Netherlands, contains, ?</i>)	
BAKE	Correct tail rank: "Alkmaar" : <u>1</u>
IncDE	Correct tail rank: "Alkmaar" : <u>25</u>
FastKGE	Correct tail rank: "Alkmaar" : <u>69</u>
Case 4: (<i>Raoul Walsh, place_of_birth, ?</i>)	
BAKE	Correct tail rank: "New York" : <u>3</u>
IncDE	Correct tail rank: "New York" : <u>179</u>
FastKGE	Correct tail rank: "New York" : <u>336</u>

Table 5: Case study experiments on the ENTITY dataset.

Results and analysis on different snapshots. Figure 3 plots the MRR trends of BAKE across 5 snapshots on the ENTITY and FB-CKGE benchmark datasets. As shown, BAKE consistently outperforms the baselines, achieving the highest and most stable metric values across all snapshots. Examining the curve slopes, BAKE exhibits the flattest decline in performance, indicating its superior resistance to forgetting compared to IncDE and FastKGE. In the continual learning scenario, this capability is demonstrated by

BAKE’s Bayesian ”Posterior-as-Prior” update strategy and the semantic cluster consistency regularization via continual clustering. The former facilitates efficient knowledge acquisition, while the latter constrains embedding drift, collectively achieving an optimal trade-off between plasticity (adaptation to new facts) and stability (retention of old knowledge).

The parameter sensitivity experiment of cluster number K. In a Continual learning scenario, traditional K-means clustering directly applied to each snapshot may disrupt stable cluster formation due to the dynamic emergence of new entities. To address this, we fix the cluster size K as a hyperparameter. As shown in Figure 4, the MRR performance of the BAKE model seems to improve with larger K values, indicating that a relatively higher number of clusters enhances semantic consistency and knowledge retention by allowing finer-grained embedding grouping. The performance gap between different K values is most significant in early snapshots and gradually converges as the model evolves over multiple time steps. However, if K is too large, performance may degrade despite increased overhead.

Case Study. To visually demonstrate BAKE’s practical advantages, we conducted an in-depth analysis of four link prediction cases from the ENTITY dataset, comparing BAKE against strong baselines (IncDE and FastKGE). As shown in Table 5, these cases cover diverse relationship types, including profession, category, geographical affiliation, and birthplace. For queries such as (*Blade Runner, genre, ?*) and (*Netherlands, contains, ?*), BAKE correctly ranked the ground-truth answers ("Drama" and "Alkmaar") at the top position. In contrast, competing models degraded the true answers to positions beyond the top-100. Remarkably, even for highly sparse relations like the query for Raoul Walsh’s birthplace, BAKE achieved a competitive rank (3rd position), significantly outperforming the baselines (179th and 336th ranks).

Conclusion

In this paper, we propose a novel Bayesian-guided CKGE model that effectively mitigates catastrophic forgetting in evolving knowledge graphs by treating sequential updates as Bayesian inference and incorporating continual clustering to preserve semantic consistency. Extensive experiments on multiple benchmarks demonstrate that our model outperforms existing baselines in terms of knowledge preservation and adaptability.

References

- Ahn, H.; Cha, S.; Lee, D.; and Moon, T. 2019. Uncertainty-based continual learning with adaptive regularization. *Advances in neural information processing systems*, 32.
- Bonnet, D.; Cottart, K.; Hirtzlin, T.; Januel, T.; Dalgaty, T.; Vianello, E.; and Querlioz, D. 2025. Bayesian continual learning and forgetting in neural networks. *arXiv preprint arXiv:2504.13569*.
- Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.
- Cao, Z. 2024. Continual Learning Knowledge Graph Embeddings for Dynamic Knowledge Graphs.
- Cao, Z.; Xu, Q.; Yang, Z.; Cao, X.; and Huang, Q. 2022. Geometry interaction knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 5521–5529.
- Cui, Y.; Wang, Y.; Sun, Z.; Liu, W.; Jiang, Y.; Han, K.; and Hu, W. 2023. Lifelong embedding learning and transfer for growing knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4217–4224.
- Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; and Philip, S. Y. 2021. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2): 494–514.
- Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13): 3521–3526.
- Kou, X.; Lin, Y.; Liu, S.; Li, P.; Zhou, J.; and Zhang, Y. 2020. Disentangle-based continual graph representation learning. *arXiv preprint arXiv:2010.02565*.
- Lee, S.; Jeon, H.; Son, J.; and Kim, G. 2024. Learning to Continually Learn with the Bayesian Principle. In *International Conference on Machine Learning*, 26621–26639. PMLR.
- Li, L.; Jin, Z.; He, Y.; Jin, D.; Duan, H.; Tao, Z.; Zhang, X.; and Li, J. 2025a. Rethinking regularization methods for knowledge graph completion. *arXiv preprint arXiv:2505.23442*.
- Li, L.; Jin, Z.; Zhang, X.; Duan, H.; Wang, J.; Tao, Z.; Zhao, H.; and Zhu, X. 2025b. Multi-view riemannian manifolds fusion enhancement for knowledge graph completion. *IEEE Transactions on Knowledge and Data Engineering*.
- Li, L.; Jin, Z.; Zhang, Y.; Jin, D.; Dou, C.; He, Y.; Zhang, X.; and Zhao, H. 2025c. Towards Structure-aware Model for Multi-modal Knowledge Graph Completion. *arXiv preprint arXiv:2505.21973*.
- Li, L.; Zhang, X.; Jin, Z.; Gao, C.; Zhu, R.; Liang, Y.; and Ma, Y. 2023. Knowledge graph completion method based on quantum embedding and quaternion interaction enhancement. *Information Sciences*, 648: 119548.
- Liang, K.; Liu, Y.; Li, H.; Meng, L.; Liu, S.; Wang, S.; Zhou, S.; and Liu, X. 2024a. Clustering then propagation: Select better anchors for knowledge graph embedding. *Advances in Neural Information Processing Systems*, 37: 9449–9473.
- Liang, K.; Meng, L.; Liu, M.; Liu, Y.; Tu, W.; Wang, S.; Zhou, S.; Liu, X.; Sun, F.; and He, K. 2024b. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Liu, J.; Ke, W.; Wang, P.; Shang, Z.; Gao, J.; Li, G.; Ji, K.; and Liu, Y. 2024a. Towards continual knowledge graph embedding via incremental distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8759–8768.
- Liu, J.; Ke, W.; Wang, P.; Wang, J.; Gao, J.; Shang, Z.; Li, G.; Xu, Z.; and Ji, K. 2024b. Fast and Continual Knowledge Graph Embedding via Incremental LoRA. In *IJCAI*.
- Long, X.; Zhuang, L.; Li, A.; Wei, J.; Li, H.; and Wang, S. 2024. KGDM: A diffusion model to capture multiple relation semantics for knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8850–8858.
- Lopez-Paz, D.; and Ranzato, M. 2017. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30.
- Ma, T.; Chen, Y.; Wang, L.; Lin, X.; Song, B.; and Zeng, X. 2025. S²DN: Learning to Denoise Unconvincing Knowledge for Inductive Knowledge Graph Completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12355–12363.
- Mahdisoltani, F.; Biega, J.; and Suchanek, F. M. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.
- Nguyen, X.-B.; Bui, D. T.; Duong, C. N.; Bui, T. D.; and Luu, K. 2021. Clusformer: A transformer based clustering approach to unsupervised large-scale face and visual landmark recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10847–10856.
- Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; and Wu, X. 2024. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7): 3580–3599.
- Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Rusu, A. A.; Rabinowitz, N. C.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Pascanu, R.; and Hadsell, R. 2016. Progressive neural networks. *arXiv preprint arXiv:1606.04671*.
- Song, A.; Chen, Y.; Wang, Y.; Zhong, S.; and Xu, M. 2024. Orchestrating Plasticity and Stability: A Continual Knowledge Graph Embedding Framework with Bio-Inspired Dual-Mask Mechanism. In *The 16th Asian Conference on Machine Learning (Conference Track)*.

- Suchanek, F. M.; Alam, M.; Bonald, T.; Chen, L.; Paris, P.-H.; and Soria, J. 2024. Yago 4.5: A large and clean knowledge base with a rich taxonomy. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 131–140.
- Sun, Z.; Deng, Z.-H.; Nie, J.-Y.; and Tang, J. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*.
- Truong, T.-D.; Prabhu, U.; Raj, B.; Cothren, J.; and Luu, K. 2025. FALCON: Fairness Learning via Contrastive Attention Approach to Continual Semantic Scene Understanding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 15065–15075.
- Wang, H.; Xiong, W.; Yu, M.; Guo, X.; Chang, S.; and Wang, W. Y. 2019. Sentence embedding alignment for lifelong relation extraction. *arXiv preprint arXiv:1903.02588*.
- Wang, J.; Wang, B.; Qiu, M.; Pan, S.; Xiong, B.; Liu, H.; Luo, L.; Liu, T.; Hu, Y.; Yin, B.; et al. 2023. A survey on temporal knowledge graph completion: Taxonomy, progress, and prospects. *arXiv preprint arXiv:2308.02457*.
- Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28.
- Xiao, C.; He, X.; and Cao, Y. 2023. Knowledge graph embedding by normalizing flows. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4756–4764.
- Xu, D.; Li, X.; Zhang, Z.; Lin, Z.; Zhu, Z.; Zheng, Z.; Wu, X.; Zhao, X.; Xu, T.; and Chen, E. 2025. Harnessing large language models for knowledge graph question answering via adaptive multi-aspect retrieval-augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 25570–25578.
- Yuan, D.; Zhou, S.; Chen, X.; Wang, D.; Liang, K.; Liu, X.; and Huang, J. 2025a. Knowledge graph completion with relation-aware anchor enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 15239–15247.
- Yuan, M.; Zhang, Z.; Chen, W.; Zhao, C.; Cai, T.; Wang, D.; Liu, R.; and Zhuang, F. 2025b. Hek-cl: Hierarchical enhanced knowledge-aware contrastive learning for recommendation. *ACM Transactions on Information Systems*.
- Zenke, F.; Poole, B.; and Ganguli, S. 2017. Continual learning through synaptic intelligence. In *International conference on machine learning*, 3987–3995. PMLR.
- Zhang, S.; Liang, X.; Niu, S.; Niu, Z.; Wu, B.; Hua, G.; Wang, L.; Guan, Z.; Wang, H.; Zhang, X.; et al. 2025. Integrating Large Language Models and Möbius Group Transformations for Temporal Knowledge Graph Embedding on the Riemann Sphere. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 13277–13285.
- Zhang, S.; Tay, Y.; Yao, L.; and Liu, Q. 2019. Quaternion knowledge graph embeddings. *Advances in neural information processing systems*, 32.
- Zhang, Z.; Cai, J.; Zhang, Y.; and Wang, J. 2020. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*, 3065–3072. AAAI Press.
- Zhao, Z.; Jing, Y.; Feng, F.; Wu, J.; Gao, C.; and He, X. 2024. Leave no patient behind: Enhancing medication recommendation for rare disease patients. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 533–542.
- Zhu, L.; Jeon, D. H.; Sun, W.; Yang, L.; Xie, Y.; and Niu, S. 2024. Flexible Memory Rotation (FMR): Rotated Representation with Dynamic Regularization to Overcome Catastrophic Forgetting in Continual Knowledge Graph Learning. In *2024 IEEE International Conference on Big Data (Big-Data)*, 6180–6189. IEEE.
- Zhu, L.; Lan, Q.; Tian, Q.; Sun, W.; Yang, L.; Xia, L.; Xie, Y.; Xiao, X.; Duan, T.; Tao, C.; et al. 2025. ETT-CKGE: Efficient Task-driven Tokens for Continual Knowledge Graph Embedding. *arXiv preprint arXiv:2506.08158*.