SaCa: A Highly Compatible Reinforcing Framework for Knowledge Graph Embedding via Structural Pattern Contrast

Jiashi Lin, Changhong Jiang*, Yixiao Wang, Xinyi Zhu, Zhongtian Hu, Wei Zhang
School of Computer Science, Northwestern Polytechnical University, Xi'an, China

Abstract

Knowledge Graph Embedding (KGE) seeks to learn latent representations of entities and relations to support knowledge-driven AI systems. However, existing KGE approaches often exhibit a growing discrepancy between the learned embedding space and the intrinsic structural semantics of the underlying knowledge graph. This divergence primarily stems from the over-reliance on geometric criteria for assessing triple plausibility, whose effectiveness is inherently limited by the sparsity of factual triples and the disregard of higher-order structural dependencies in the knowledge graph. To overcome this limitation, we introduce Structure-aware Calibration (SaCa), a versatile framework designed to calibrate KGEs through the integration of global structural patterns. SaCa designs two new components: (i) Structural Proximity Measurement, which captures multi-order structural signals from both entity and entity-relation perspectives; and (ii) KG-Induced Soft-weighted Contrastive Learning (KISCL), which assigns soft weights to hard-to-distinguish positive and negative pairs, enabling the model to better reflect nuanced structural dependencies. Extensive experiments on seven benchmarks demonstrate that SaCa consistently boosts performance across ten KGE models on link prediction and entity classification tasks with minimal overhead.

1 Introduction

Knowledge graphs (KGs), which represent entities as nodes and their relationships as edges in a graph-based structure, have widespread applications in diverse fields such as knowledge discovery, question answering, recommendation systems, and clinical prediction. The applicability of KGs faces two primary challenges. Firstly, real-world knowledge graphs suffer from sparsity and incompleteness despite their considerable scales. Secondly, while most knowledge-driven applications operate in numerical spaces, knowledge graphs are predominantly represented in symbolic or logical formats.

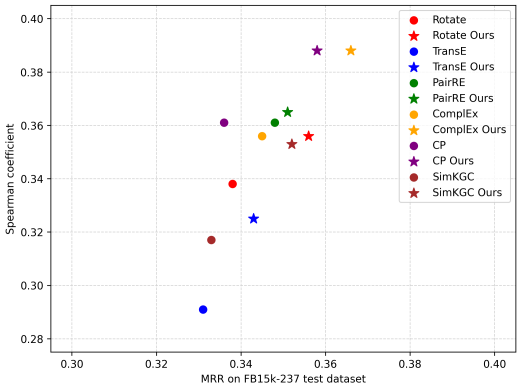

Figure 1: The Spearman correlation coefficient of KG-induced similarities and MRR on FB15k-237 dataset.

To enable efficient reasoning and knowledge acquisition, knowledge graph embedding (KGE) techniques (Sun et al., 2019; Trouillon et al., 2016; Balazevic et al., 2019; Zhang et al., 2020; Wang et al., 2022) have been developed to map entities and relations into low-dimensional vector spaces, which are expected to preserve the underlying structural properties of the original graph. Thus, embeddings are widely utilized in various downstream tasks, including knowledge graph completion (Wang et al., 2022; Fan et al., 2024; Jiang et al., 2024; Wei et al., 2024), entity classification (Xie et al., 2016; Weller and Paulheim, 2021; Liang et al., 2023), and more.

While current KGE models have yielded promising outcomes, most still focus on explaining single triple facts (an intra-triple perspective) while overlooking the broader inter-triple relationships among triples. As a result, researchers raise questions about their ability to effectively capture structural proximity within KGs (Jain et al., 2021; Ilievski et al., 2023; Hubert et al., 2024). Building upon previous work, we further reveal that structural semantic drift is prevalent across existing KGE

* Corresponding author

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 15008–15021

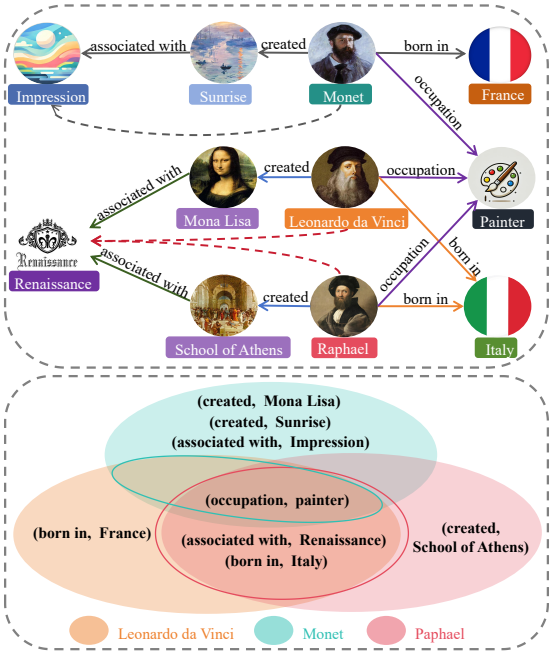

Figure 2: Illustration of graph contexts and their overlap for Leonardo Da Vinci, Raphael, and Monet within a knowledge graph.

models. This is evidenced by varying Spearman correlation coefficients[1] between KG-induced similarities and cosine similarities from learned entity embeddings, as shown in Fig. 1. Such variation reflects differing model abilities to preserve original KG semantics, potentially leading to suboptimal performance where fine-grained structural distinctions are critical.

Recent contrastive strategies (Liang et al., 2023; Qiao et al., 2023; Zhang et al., 2024; Lin et al., 2024) attempt to address this by using standard contrastive learning to learn more discriminative entity embeddings. However, this approach suffers from three key limitations: Firstly, the continuous nature of entity similarity conflicts with rigid binary labeling in simple contrastive learning, necessitating non-discrete modeling. Secondly, over-reliance on individual triples for contrastive sample construction lacks mechanisms for capturing contextual consistency across graph structures. As illustrated in Figs. 2 and 3, while Da Vinci, Raphael, and Monet all share the triple $(ent \xrightarrow{occupation} painter)$, Da Vinci and Raphael exhibit stronger semantic ties due to a greater extent of contextual overlap between them. Consequently, their embedding space distance should be closer than that with Monet. Thirdly, in real-world KGs with multiple relations, entity semantics are inherently multifaceted and vary across different relational contexts. However, existing contrastive methods often emphasize entity-level contrast, ignoring these relation-contextualized semantics.

To address this, we propose Structural Calibration (SaCa), a plug-and-play framework that improves the sensitivity of a model to global structural signals in the KG. SaCa draws inspiration from the distributional hypothesis: entities with similar meanings tend to appear in similar structural contexts. Based on this, we design four strategies to induce similarity metrics grounded in context overlap, which capture multi-scale structural cues from both entity and relation perspectives.

These structural similarities are then used as soft supervision in our KG-induced Soft-weighted Contrastive Learning (KISCL) module. Unlike standard contrastive learning, KISCL introduces a continuous weighting mechanism that adjusts the loss sensitivity based on similarity confidence, enabling more nuanced and structure-aware embedding refinement.

SaCa integrates seamlessly into existing models without adding parameters. Extensive experiments show that it consistently enhances structural fidelity and improves performance across a range of KGE architectures and benchmarks.

2 Related Work

2.1 Knowledge Graph Embedding

Knowledge graph embedding (KGE) aims to represent entities and relations in a continuous vector space based on the existing triple structure. Typically, conventional KGE models employ various scoring functions to evaluate the plausibility of triples.

Translation-based approaches, such as TransE (Bordes et al., 2013), TransR (Lin et al., 2015b), RotatE (Sun et al., 2019), and PairRE(Chao et al., 2021) rely on the translation assumption.

Tensor decomposition-based models rely on low-rank factorization assumptions, such as CP decomposition (Hitchcock, 1927), RESCAL (Nickel et al., 2011), DistMult (Yang et al., 2015), and ComplEx (Trouillon et al., 2016). However, these models have limitations since they can easily overfit the training triples while violating the distributed se-

[1]We adopt the non-parametric Spearman correlation coefficient to measure the monotonic relationship between the similarity rankings predicted from learned embeddings and the KG-induced similarity scores computed based on the structural proximity defined in Section 4.1.

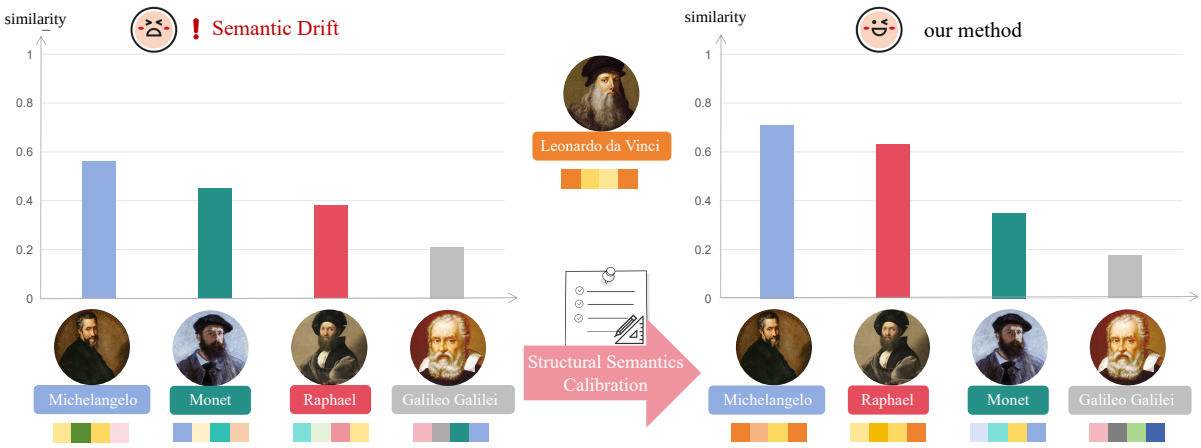

Figure 3: Differences in entity embedding similarity rankings before and after structural semantic calibration.

mantics assumption (Zhang et al., 2020; Cui and Zhang, 2024).

Graph Neural Networks (GNNs) (Ristoski and Paulheim, 2016; Vashishth et al., 2019; Weller and Paulheim, 2021; Tan et al., 2023) have been widely explored for knowledge graph completion (KGC), as they incorporate structural information through iterative message passing. However, recent studies report that message passing is often less effective in KGC (Zhang et al., 2022; Li et al., 2023b). Moreover, the over-smoothing problem, where repeated propagation causes node embeddings to converge and lose discriminability, is well documented in the GNN literature (Oono and Suzuki, 2020; Keriven, 2022). These limitations motivate alternatives that avoid excessive reliance on local message passing while preserving structural information.

Recently, text-based approaches (Wang et al., 2022; Jiang et al., 2023, 2024), which integrate pre-trained language models (PLMs) or large language models (LLMs) into KGE via entity descriptions, have attracted growing attention. While effective in leveraging textual semantics, these methods struggle to capture the structural richness of knowledge graphs. Our framework addresses this gap by explicitly incorporating structural patterns alongside textual signals to enhance structural awareness and improve robustness.

2.2 Contrastive Learning for Knowledge Graph Embedding

Contrastive learning has recently emerged as a powerful paradigm for KGE. Several representative methods have explored this direction. For instance, LP-BERT (Li et al., 2023a) employs triple-level negative sampling to refine relational representations. C-LMKE (Wang et al., 2023) and SimKGC (Wang et al., 2022) apply the InfoNCE loss on PLM-encoded triples with in-batch sampled negatives. StructKGC (Lin et al., 2024) developed a structure-aware contrastive learning method by contrasting triples with their local structural contexts. PMD (Fan et al., 2024) introduces a progressive distillation framework to stabilize training, while GHN (Qiao et al., 2023) generates challenging negatives through a seq2seq model. HaSa+ (Zhang et al., 2024) further emphasizes hard negative mining and proposes strategies to mitigate the impact of false negatives.

Despite these advances, most methods rely on rigid binary labels, limiting their ability to capture fine-grained structural semantics. In contrast, our approach harnesses global structural patterns as soft contrastive signals to achieve higher-order structural awareness.

3 Preliminaries

In this section, we briefly describe the formulation of the knowledge graph embedding (KGE) and the concept of graph context within a knowledge graph (KG). **A Knowledge Graph** (KG): A general KG can be formalized as $\mathcal{KG} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ where $\mathcal{E}, \mathcal{R}$ are the entity set, the relation set respectively. $\mathcal{T} = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$ is the triple set.

Knowledge Graph Embedding. *Knowledge Graph Embedding* (KGE) encompasses encoding entities $\mathcal{E}$ and relations $\mathcal{R}$ into low-dimensional continuous vectors ($d \ll |\mathcal{E}|$). These embeddings are typically optimized by applying a scoring function $\mathcal{F}(h, r, t)$ to evaluate the plausibility of triples.

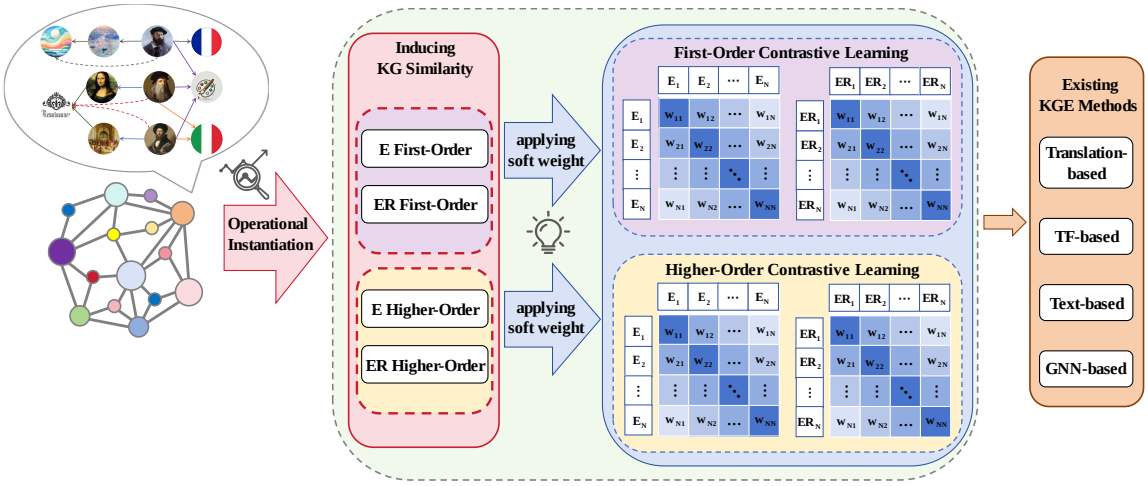

Figure 4: Overview of our proposed SaCa framework to alleviate the structural drift problem in knowledge graphs.

3.1 Graph Context

Graph context encapsulates the relational environments surrounding a subject (entity or entity-relation pair) within a knowledge graph, which is critical for interpreting its structural implications. To achieve a comprehensive understanding, we examine graph context at multiple scales, specifically **first-order** and **higher-order** contexts. Additionally, we consider different subject perspectives, including entity-centric (E) and relation-aware (E-R) views, to ensure a comprehensive understanding of multifaceted semantics across various relational contexts.

For a view S (either E or E-R), its k-order contextual scope $\mathcal{C}^k(S)$ is defined as:

$$\mathcal{C}^k(E) = \left\{ (r_1, \ldots, r_k, E_k) \;\middle|\; E \xrightarrow{[r_1 \circ \cdots \circ r_k]} E_k \right\}, \quad (1)$$

$$\mathcal{C}^k(E, R) = \left\{ (r_1, \ldots, r_k, E_k) \;\middle|\; E \xrightarrow{R \circ [r_1 \circ \cdots \circ r_k]} E_k \right\}, \quad (2)$$

where k denotes the contextual scale. At $k = 1$, the framework specializes in capturing the most fundamental and localized relationships associated with the subject. When $k \geq 2$, the **higher-order graph context** extends beyond direct neighbors, encoding logical patterns along multi-hop relation chains and revealing implicit structural interactions. In practice, these higher-order contexts are automatically extracted using depth-first search (DFS) with a maximum path length of 4, as empirical evidence from prior research (Sun et al., 2018) indicates that semantically similar entities are rarely separated by extensive relation paths.

4 Methodology

In this section, we present our framework, SaCa, as illustrated in Fig. 4. Four combination strategies are designed to induce the structural proximity inherent in knowledge graphs, capturing both direct and indirect cues from both entity and relation perspectives. Then, we propose KG-induced Soft-weighted Contrastive Learning (KISCL), which leverages KG-derived similarities as continuous weighting factors in contrastive learning to refine the knowledge embedding space. Finally, we seamlessly integrate our framework with existing KGE methods by jointly optimizing the proposed contrastive loss and their original scoring functions, leading to a more comprehensive and effective knowledge representation.

4.1 Multi-scale Structural Proximity Measurement

Measuring similarity within complex, multi-layered KGs is challenging. Inspired by the distributional hypothesis (Suresh et al., 2023), we design a multi-scale structural proximity measurement based on overlapping graph contexts. This measurement operates at varying granularities—from first-order ($k = 1$) analysis of direct connections using $\mathcal{C}^1(S)$, to higher-order ($k \geq 2$) exploration of implicit structural chains within broader contexts $\mathcal{C}^k(S)$ (identified via graph traversals)—thereby enabling the comprehensive modeling of both direct connections and multi-hop dependencies. We then unify this similarity measurement using the contextual scope operator $\mathcal{C}^k(S)$ (from Sec. 3.1) in a parameterized Jaccard formulation:

$$\text{Sim}^k(S_1, S_2) = \frac{|\mathcal{C}^k(S_1) \cap \mathcal{C}^k(S_2)|}{|\mathcal{C}^k(S_1) \cup \mathcal{C}^k(S_2)|}. \quad (3)$$

4.1.1 Relation-aware Operational Instantiation

Unlike previous contrastive KGE methods that focus solely on entity-level discrimination (Liang et al., 2023; Qiao et al., 2023; Zhang et al., 2024), we consider that the relations also play a crucial role in knowledge graphs. Thus, we simultaneously consider two perspectives: the entity-centric (E) mode, which examines all surrounding structural contexts to form a comprehensive entity profile, and the relation-aware (ER) mode, which focuses on entity structure within specific relations, emphasizing relation-contextualized semantics. Both perspectives analyze similarity at multiple granularities: first-order measures capture direct connections, while higher-order ones encode indirect semantic associations. Consequently, we define four operational instantiations for the Jaccard formulation as follows.

E-First: First-order entity-centric similarity for entity pairs (E_1, E_2):

$$\text{Sim}^{\text{e-first}}(E_1, E_2) = \frac{|\mathcal{C}^1(E_1) \cap \mathcal{C}^1(E_2)|}{|\mathcal{C}^1(E_1) \cup \mathcal{C}^1(E_2)|}. \quad (4)$$

E-High: Higher-order ($k > 1$) entity-centric similarity:

$$\text{Sim}^{\text{e-high}}(E_1, E_2) = \frac{|\mathcal{C}^k(E_1) \cap \mathcal{C}^k(E_2)|}{|\mathcal{C}^k(E_1) \cup \mathcal{C}^k(E_2)|}. \quad (5)$$

ER-First: First-order relation-constrained similarity, where the context for entity E is refined to $\mathcal{C}^1(E, R)$:

$$\text{Sim}^{\text{er-first}}((E_1, R_1), (E_2, R_2)) = \frac{|\mathcal{C}^1(E_1, R_1) \cap \mathcal{C}^1(E_2, R_2)|}{|\mathcal{C}^1(E_1, R_1) \cup \mathcal{C}^1(E_2, R_2)|}. \quad (6)$$

ER-High: Higher-order relation-constrained similarity using k-th order paths initiating with a specific relation R:

$$\text{Sim}^{\text{er-high}}((E_1, R_1), (E_2, R_2)) = \frac{|\mathcal{C}^k(E_1, R_1) \cap \mathcal{C}^k(E_2, R_2)|}{|\mathcal{C}^k(E_1, R_1) \cup \mathcal{C}^k(E_2, R_2)|}. \quad (7)$$

All similarity measures follow the Jaccard index definition with values in [0,1], where 0 indicates completely disjoint context sets and 1 represents identical sets.

4.1.2 Strategies for Accelerating Similarity Computing

The naive pairwise Jaccard similarity computation has prohibitive complexity of $O(N^2 \cdot M)$ for N entities with average context size M, limiting scalability to large knowledge graphs. We address this with a feature-context inverted index that transforms global comparisons into local, feature-specific intersections. The effectiveness of this strategy leverages the inherent long-tail distribution in knowledge graphs, where only a few features are frequent and most appear rarely. As a result, the computational cost is dominated by the feature frequency m_f, and for any feature f, $m_f \ll N$ due to sparsity. Thus, the overall complexity becomes $\sum_{f \in F} O(m_f^2) \ll O(N^2 \cdot M)$, yielding substantial efficiency gains, especially for large-scale graphs. More empirical results are provided in Appendix D.

4.2 Similarity-Guided Continuously Weighted Contrastive Learning

We introduce the **KG-induced Soft-weighted Contrastive Learning (KISCL)**, which leverages KG-derived similarity as continuous weighting factors in contrastive learning. KISCL dynamically adjusts embedding distances by bringing similar entities closer and pushing dissimilar ones apart, guided by the KG-induced similarity. Based on the two essential characteristics of contrastive loss—alignment and uniformity—KISCL consists of two components: **Dynamic Alignment Loss** and **Mutual Dissimilarity Loss**.

To dynamically align the embedding space produced by KGE models with the inherent semantics of the original knowledge graph, we introduce the dynamic alignment loss as:

$$\mathcal{L}_d^{Sim} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{k} \sum_{j \in P_i(k)} Sim(s_i, s_j) \cdot f(s_i, s_j) \right). \quad (8)$$

Where N denotes the total number of subjects, and k represents the number of considered positive samples. $P_i(k)$ refers to the set of k positive samples for the i-th entity. The function $f(\cdot, \cdot)$ computes the Euclidean distance between two subjects' embeddings, s_i and s_j. $Sim(s_i, s_j)$ denotes the structural similarity calculated through specific operational instantiation (e.g., E-First, ER-First, E-High, ER-High), with values in [0,1] following the Jaccard index definition.

To increase the uniformity of the embedding space, we introduce a mutual dissimilarity loss that

penalizes the similarity between embeddings of dissimilar entities. This promotes the even distribution of entities in the latent space, preventing overly dense embeddings. The Mutual Dissimilarity Loss is defined as:

$$\mathcal{L}_m([s_1, \cdots, s_B]) = \frac{2}{B(B-1)} \sum_{i=1}^{B} \sum_{j=i+1}^{B} \frac{s_i}{\|s_i\|} \cdot \frac{s_j}{\|s_j\|}, \tag{9}$$

where B denotes the batch size, The terms $\|s_i\|$ and $\|s_j\|$ represent the norms of embeddings vectors s_i and s_j.

4.3 Model Training and Inference

Our framework captures structural granularity at different levels. Therefore, we define two types of similarity-based losses—one for first-order and another for higher-order information—each incorporating terms from both entity (E) and entity-relation (ER) perspectives. These losses are formulated as follows:

$$\mathcal{L}_{\text{KISCL}}^{\text{First/High}} = \alpha \mathcal{L}_d^{\text{E-First/High}} + \beta \mathcal{L}_d^{\text{ER-First/High}} + \lambda \mathcal{L}_m. \tag{10}$$

The total loss function, combining the KGE backbone loss and the weighted contrastive learning loss $\mathcal{L}_{\text{KISCL}}$, is given by:

$$\mathcal{L}o = \mathcal{L}_{kge} + \mathcal{L}_{\text{KISCL}}. \tag{11}$$

To mitigate structural interference, our proposed decoupled-integrated learning framework utilizes a two-phase optimization. It first allows for separate training to leverage diverse similarity-type strengths, then integrates local and global structural information via an embedding fusion strategy for joint prediction:

$$\mathbf{z} = \gamma \cdot \mathbf{z}_{\text{first}} + (1-\gamma) \cdot \mathbf{z}_{\text{high}}. \tag{12}$$

5 Experiment

5.1 Dataset

For the link prediction task, we adopt four widely-used benchmark datasets: **WN18RR** (Dettmers et al., 2018), **FB15k-237** (Toutanova and Chen, 2015), **YAGO3-10**, **UMLS**, and **Kinship** (Kok and Domingos, 2007). For the entity classification task, we adopt two benchmark datasets, including AIFB (Bloehdorn and Sure, 2007) and MUTAG (Ristoski et al., 2016). These datasets vary in size and complexity, allowing for a comprehensive assessment of our method across different types of knowledge graphs. We use the training set to compute KG similarity and train the model. For further details, please refer to Appendix B.

5.2 Baselines

In our study, we conducted a comparative analysis of our methods against various KGEs. The translation-based methods we considered encompass TransE (Bordes et al., 2013), RotatE (Sun et al., 2019), PairRE (Chao et al., 2021) and HAKE (Li et al., 2022). The tensor decomposition-based methods include CP (Hitchcock, 1927), ComplEx (Trouillon et al., 2016), DisMult (Yang et al., 2015), N3 (Lacroix et al., 2018) and DURA (Zhang et al., 2020). The text-based methods we evaluated include SimKGC (Wang et al., 2022), LP-BERT (Li et al., 2023a), C-LMKE (Wang et al., 2023), GHN (Qiao et al., 2023), HaSa (Zhang et al., 2024), PMD(Fan et al., 2024) and KG-FIT(Jiang et al., 2024). The GNN-based methods we include RDF2Vec (Ristoski and Paulheim, 2016), R-GCN (Schlichtkrull et al., 2018), E-R-GCN (Weller and Paulheim, 2021) and MQuinE (Shang et al., 2024).

5.3 Evaluation Metrics

The evaluation framework employs two complementary assessment tasks to comprehensively validate model performance: **link prediction** for knowledge graph completion and **entity classification** for structural categorization. This task evaluates the embedding quality through rank-based metrics, with Mean Reciprocal Rank (MRR) and Hits@N, where N$\in \{1, 3, 10\}$. The entity classification task adopts classification accuracy as its core metric, quantifying the proportion of correctly categorized entities against a predefined taxonomy.

5.4 Implementation Detail

Our framework SaCa is implemented based on Pytorch (Paszke et al., 2019). For fair competition, we ensure consistency with the baseline configurations as presented in the original papers. For further details about the novel weighting parameters, please refer to Appendix A. The implementation is publicly available at `https://github.com/ninjaX2o/SaCa`.

5.5 Main Result

Table 1 summarizes the link prediction results for our SaCa framework when applied to various baseline models on the WN18RR and FB15k-237 benchmark datasets. Detailed link prediction results for the smaller UMLS and Kinship datasets are provided in Appendix E.1. To evaluate its effectiveness and generalizability, SaCa was integrated

Methods	WN18RR				FB15k-237			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
Translation Based Models								
TransE	0.202	0.014	0.367	0.496	0.329	0.230	0.368	0.526
TransE-SaCa	**0.229**	**0.041**	**0.388**	**0.523**	**0.343**	**0.244**	**0.382**	**0.539**
RotatE	0.476	0.428	0.492	0.571	0.335	0.238	0.372	0.530
RotatE-SaCa	**0.481**	**0.435**	**0.498**	**0.576**	**0.356**	**0.260**	**0.393**	**0.548**
PairRE	0.451	0.405	0.463	0.545	0.345	0.252	0.381	0.542
PairRE-SaCa	**0.461**	**0.416**	**0.476**	**0.555**	**0.353**	**0.257**	**0.390**	**0.552**
Tensor Decomposition Based Models								
CP	0.438	0.414	0.444	0.485	0.333	0.247	0.364	0.508
CP-SaCa	**0.486**	**0.447**	**0.498**	**0.559**	**0.358**	**0.268**	**0.393**	**0.543**
DisMult	0.444	0.414	0.454	0.504	0.343	0.251	0.376	0.525
DisMult-SaCa	**0.476**	**0.431**	**0.493**	**0.564**	**0.351**	**0.262**	**0.385**	**0.531**
ComplEx	0.460	0.428	0.471	0.522	0.346	0.256	0.382	0.525
ComplEx-SaCa	**0.495**	**0.447**	**0.512**	**0.581**	**0.365**	**0.272**	**0.405**	**0.554**
Text Based Models								
C-LMKE	0.619	0.523	0.671	0.789	0.306	0.218	0.331	0.484
LP-BERT	0.482	0.343	0.563	0.752	0.310	0.223	0.336	0.490
GHN	0.678	0.596	0.719	0.821	0.339	0.251	0.364	0.518
HaSa+	0.538	0.444	0.588	0.713	0.304	0.220	0.325	0.483
SimKGC	0.671	0.587	0.731	0.817	0.333	0.246	0.362	0.510
SimKGC-SaCa	0.689	**0.619**	0.728	0.822	**0.355**	**0.260**	0.391	0.545
PMD	0.678	0.588	0.737	0.832	0.331	0.241	0.363	0.518
PMD-SaCa	**0.690**	0.615	**0.740**	**0.834**	0.350	0.255	**0.395**	**0.548**

Table 1: Performance Comparison on WN18RR and FB15k-237 Datasets.

Methods	AIFB	MUTAG
RDF2Vec	0.889	0.672
RGCN	0.931	0.682
RGCN-SaCa	**0.944**	**0.706**
E-R-GCN	0.913	0.682
E-R-GCN-SaCa	0.936	0.693

Table 2: Comparative evaluation of SaCa against baseline KGE models on entity classification tasks.

into popular KGE baselines across different categories. The results demonstrate that applying SaCa generally leads to notable improvements over the respective baseline models across evaluation metrics. For instance, SaCa boosts the performance of Translation Based Models on WN18RR by an average of 5.54% in MRR and a significant 65.74% in Hit@1 relative to their baselines. For Tensor Decomposition Based Models on WN18RR, the average relative gains were 8.59% for MRR and 5.51% for Hit@1. When applied to SimKGC (Text Based) on FB15k-237, SaCa increased MRR by 6.61% and Hit@1 by 5.69% relatively. This underscores that SaCa's proposed graph-induced weighting mechanism effectively introduces nuanced structural supervision, leading to enhanced differentiation between semantically similar entities.

Further analysis, as shown in Table 2, compares SaCa with GNN-based methods on the entity classification task. Unlike GNNs, which mainly depend on neighborhood aggregation and are prone to oversmoothing, SaCa leverages its distinctive KISCL loss to more effectively align and distinguish entities across diverse structural patterns and relational contexts. This approach resonates with prior studies Zhang et al. (2022); Li et al. (2023b), which highlight the critical role of loss function design over complex information propagation schemes in enhancing KGE performance.

In summary, the consistent improvements observed when applying SaCa—across datasets of varying sizes and for different tasks —demonstrate its robustness and versatility as a powerful enhancement for a diverse range of KGE models.

5.6 Semantic Confusion Rate Analysis

To quantitatively evaluate the model's capacity in distinguishing semantically proximate entities, we evaluated our method against three categories of baseline approaches using a curated challenge subset consisting of 1,200 high-similarity entity pairs based on pretrained embeddings from the WN18RR dataset. Then, we propose the **Semantic**

Confusion Rate (SCR) metric:

$$\text{SCR} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\left(\text{rank}_{\text{correct}}^{(i)} > \text{rank}_{\text{nearest\_distractor}}^{(i)}\right), \tag{13}$$

where $\text{rank}_{\text{correct}}$ denotes the rank of the ground-truth entity in prediction results, and nearest_distractor represents the most semantically similar competing entity identified through cosine similarity in the embedding space. Lower SCR values indicate superior discrimination capability.

As shown in Table 3, the baseline models exhibit increasing SCR with higher semantic similarity. However, our SaCa demonstrates progressively greater advantages in higher similarity intervals. The relative SCR reduction reaches 7.5% in the >0.9 range, compared to 5.2% reduction in 0.8-0.9 and 4.1% in 0.7-0.8. This pattern suggests our method particularly excels at resolving subtle semantic distinctions.

Method	Similarity Range		
	0.7–0.8	0.8–0.9	>0.9
TransE	24.3 ± 1.2	31.7 ± 1.5	43.2 ± 2.1
RotatE	21.8 ± 1.1	28.4 ± 1.3	39.6 ± 1.9
ComplEx	18.9 ± 0.9	23.7 ± 1.1	34.1 ± 1.6
KG-BERT	19.7 ± 1.0	25.3 ± 1.2	36.4 ± 1.7
SimKGC	16.2 ± 0.8	20.5 ± 1.0	29.8 ± 1.4
SimKGC-SaCa	**12.1 ± 0.6**	**15.3 ± 0.7**	**22.3 ± 0.9**

Table 3: Semantic Confusion Rate (SCR) results (%).

5.7 Ablation Analysis

Our ablation study on ComplEx, presented in Table 4, confirms the necessity of each component in our framework, as their removal consistently degrades performance. A balanced combination of Dynamic Alignment Loss terms ($\mathcal{L}_d^{\text{E}}$ and $\mathcal{L}_d^{\text{ER}}$) is particularly important, since using either term alone results in substantial performance drops, with the removal of $\mathcal{L}_d^{\text{ER}}$ having the most pronounced effect on WN18RR. On the sparse UMLS dataset, higher-order terms prove more effective, corroborating our strategy of leveraging broader graph contexts.

To further validate our design, we examined two variants: SaCa-D, which discretizes similarity scores, and SaCa-S, which restricts the model to a single relational context. Both variants underperform, underscoring the importance of continuous similarity scores and diverse relational contexts. Overall, the full SaCa model, by integrating multiple structural levels, delivers the best results, demonstrating the complementary synergy of its components.

Setting	WN18RR		UMLS	
	MRR	Hit@10	MRR	Hit@10
Only $\mathcal{L}_d^{First}$	0.488	0.567	0.844	0.967
w/o $\mathcal{L}_d^{\text{e-first}}$	0.472	0.538	0.699	0.898
w/o $\mathcal{L}_d^{\text{er-first}}$	0.461	0.526	0.775	0.932
w/o $\mathcal{L}_m$	0.483	0.558	0.838	0.965
Only $\mathcal{L}_d^{High}$	0.492	0.577	0.850	0.976
w/o $\mathcal{L}_d^{\text{e-high}}$	0.484	0.556	0.798	0.959
w/o $\mathcal{L}_d^{\text{er-high}}$	0.469	0.538	0.836	0.972
w/o $\mathcal{L}_m$	0.487	0.572	0.848	0.973
SaCa-D	0.472	0.549	0.832	0.961
SaCa-S	0.478	0.562	0.837	0.968
Full SaCa Model	**0.495**	**0.581**	**0.861**	**0.982**

Table 4: Ablation study results on WN18RR and UMLS datasets.

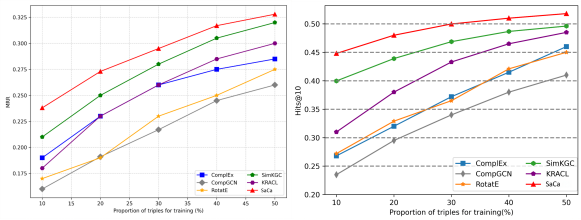

Figure 5: Link prediction performance on the FB15k-237 dataset under low-resourse setting.

5.8 Sparsity Analysis

We evaluated our model's robustness in sparse scenarios through a detailed sparsity analysis on the FB15k-237 dataset. As shown in Fig. 5, our framework consistently demonstrated superior and more stable performance compared to baseline models, even when trained on significantly reduced data subsets. This resilience is largely attributed to its ability to leverage higher-order structural contexts, which our analysis indicates remain relatively rich even in highly sparse graph conditions where first-order contexts are diminished. For a comprehensive description of the experimental setup, detailed performance metrics, and the statistical analysis of graph properties under varying sparsity levels, please refer to Appendix C.

5.9 Relational Mode Analysis

The semantics of an entity are inherently multi-faceted and exhibit variations across different relational contexts. As depicted in Fig. 6, the SCR

distribution across various relation types demonstrates how these variations manifest. This ability can be attributed to our explicit modeling of relational contexts, enhanced by the incorporation of relational-aware contrastive loss.

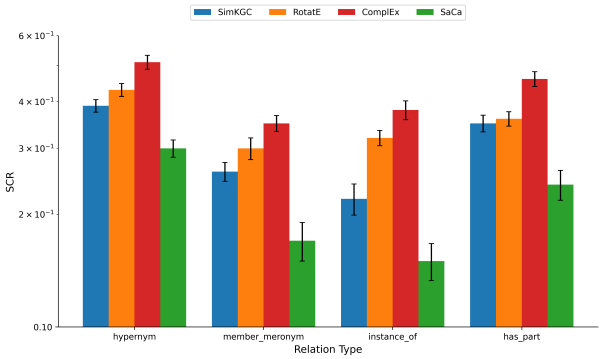

Figure 6: SCR distribution across relation types.

5.10 Scalability Analysis

Time Complexity. Our framework has linear time complexity $O(N(B+K-1)d)$, which scales efficiently with the number of triplets. Empirically, training on FB15k-237 with TB-based models incurs an average additional time of 13.6 minutes, whereas on WN18RR the average additional time is 5.6 minutes. Additional efficiency analyses are provided in Appendix D.

Larger-scale Results. On the larger-scale YAGO3-10 dataset (∼1M triples), SaCa achieves notable relative improvements over standard models. Additional results and detailed comparisons are provided in Appendix E.2.

6 Conclusion

In this study, we introduce Structure-aware Calibration (SaCa), a novel framework designed to mitigate structural drift in knowledge graph embeddings. First, we develop a structural proximity metric by quantitatively leveraging the concept of overlapping graph context. Second, we propose KG-induced Soft-weighted Contrastive Learning (KISCL) to enhance the refinement of embeddings. Our extensive experiments demonstrate that SaCa consistently enhances the performance of various baseline models, underscoring its high compatibility, efficiency, and effectiveness with existing KGE architectures. Our analysis further validates the framework's robustness, showing that SaCa preserves structural consistency while enhancing discrimination among semantically proximate entities.

Limitations

The simplicity and versatility of the SaCa framework, along with its ability to consistently improve KGE models, demonstrate its practical value in mitigating structural drift and enhancing knowledge graph embeddings. Despite the strengths of the framework, there are some limitations worth noting. Firstly, although SaCa leverages global structural patterns to compensate for the sparsity of valid triples, its performance may fluctuate on extremely sparse knowledge graphs due to limited available context signals. Secondly, as an emerging approach, the current SaCa framework primarily focuses on static knowledge graph representations and lacks mechanisms for modeling temporal evolution in dynamic knowledge graphs. Future work will explore adapting the framework to dynamic systems, enabling its extension to time-aware scenarios that require temporal calibration.

References

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5185–5194, Hong Kong, China. Association for Computational Linguistics.

Stephan Bloehdorn and York Sure. 2007. Kernel methods for mining instance data in ontologies. In *International Semantic Web Conference*, pages 58–71. Springer.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Linlin Chao, Jianshan He, Taifeng Wang, and Wei Chu. 2021. Pairre: Knowledge graph embeddings via paired relation vectors. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4360–4369.

Wanyun Cui and Linqiu Zhang. 2024. Modeling knowledge graphs with composite reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8338–8345.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Cunhang Fan, Yujie Chen, Jun Xue, Yonghui Kong, Jianhua Tao, and Zhao Lv. 2024. Progressive distillation based on masked generation feature method for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8380–8388.

Frank L Hitchcock. 1927. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189.

Nicolas Hubert, Heiko Paulheim, Armelle Brun, and Davy Monticolo. 2024. Do similar entities have similar embeddings? In *European Semantic Web Conference*, pages 3–21. Springer.

Filip Ilievski, Kartik Shenoy, Hans Chalupsky, Nicholas Klein, and Pedro Szekely. 2023. A study of concept similarity in wikidata. *Semantic Web*, (Preprint):1–20.

Nitisha Jain, Jan-Christoph Kalo, Wolf-Tilo Balke, and Ralf Krestel. 2021. Do embeddings actually capture knowledge graph semantics? In *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18*, pages 143–159. Springer.

Pengcheng Jiang, Shivam Agarwal, Bowen Jin, Xuan Wang, Jimeng Sun, and Jiawei Han. 2023. Text augmented open knowledge graph completion via pre-trained language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 11161–11180, Toronto, Canada. Association for Computational Linguistics.

Pengcheng Jiang, Lang Cao, Cao Danica Xiao, Parminder Bhatia, Jimeng Sun, and Jiawei Han. 2024. Kg-fit: Knowledge graph fine-tuning upon open-world knowledge. *Advances in Neural Information Processing Systems*, 37:136220–136258.

Nicolas Keriven. 2022. Not too little, not too much: a theoretical analysis of graph (over) smoothing. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 2268–2281.

Stanley Kok and Pedro Domingos. 2007. Statistical predicate invention. In *Proceedings of the 24th international conference on Machine learning*, pages 433–440.

Timothée Lacroix, Nicolas Usunier, and Guillaume Obozinski. 2018. Canonical tensor decomposition for knowledge base completion. In *International Conference on Machine Learning*, pages 2863–2872. PMLR.

Da Li, Boqing Zhu, Sen Yang, Kele Xu, Ming Yi, Yukai He, and Huaimin Wang. 2023a. Multi-task pre-training language model for semantic network completion. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(11).

Juanhui Li, Harry Shomer, Jiayuan Ding, Yiqi Wang, Yao Ma, Neil Shah, Jiliang Tang, and Dawei Yin. 2023b. Are message passing neural networks really helpful for knowledge graph completion? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10696–10711.

Yong-Lu Li, Xinpeng Liu, Xiaoqian Wu, Yizhuo Li, Zuoyu Qiu, Liang Xu, Yue Xu, Hao-Shu Fang, and Cewu Lu. 2022. Hake: A knowledge engine foundation for human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8494–8506.

Ke Liang, Yue Liu, Sihang Zhou, Wenxuan Tu, Yi Wen, Xihong Yang, Xiangjun Dong, and Xinwang Liu. 2023. Knowledge graph contrastive learning based on relation-symmetrical structure. *IEEE Transactions on Knowledge and Data Engineering*, 36(1):226–238.

Jiashi Lin, Lifang Wang, Xinyu Lu, Zhongtian Hu, Wei Zhang, and Wenxuan Lu. 2024. Improving knowledge graph completion with structure-aware supervised contrastive learning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13948–13959.

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015a. Modeling relation paths for representation learning of knowledge bases. *arXiv preprint arXiv:1506.00379*.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015b. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel, and 1 others. 2011. A three-way model for collective learning on multi-relational data. In *Icml*, volume 11, pages 3104482–3104584.

Kenta Oono and Taiji Suzuki. 2020. Graph neural networks exponentially lose expressive power for node classification. In *International Conference on Learning Representations (ICLR)*.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, and 2 others. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Zile Qiao, Wei Ye, Dingyao Yu, Tong Mo, Weiping Li, and Shikun Zhang. 2023. Improving knowledge

graph completion with generative hard negative mining. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5866–5878.

Petar Ristoski, Gerben Klaas Dirk De Vries, and Heiko Paulheim. 2016. A collection of benchmark datasets for systematic evaluations of machine learning on the semantic web. In *The Semantic Web–ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II 15*, pages 186–194. Springer.

Petar Ristoski and Heiko Paulheim. 2016. Rdf2vec: Rdf graph embeddings for data mining. In *International semantic web conference*, pages 498–514. Springer.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15*, pages 593–607. Springer.

Bin Shang, Yinliang Zhao, Jun Liu, and Di Wang. 2024. Mixed geometry message and trainable convolutional attention network for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8966–8974.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent knowledge graph embedding for effective recommendation. In *Proceedings of the 12th ACM conference on recommender systems*, pages 297–305.

Siddharth Suresh, Kushin Mukherjee, Xizheng Yu, Wei-Chun Huang, Lisa Padua, and Timothy Rogers. 2023. Conceptual structure coheres in human cognition but not in large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 722–738, Singapore. Association for Computational Linguistics.

Zhaoxuan Tan, Zilong Chen, Shangbin Feng, Qingyue Zhang, Qinghua Zheng, Jundong Li, and Minnan Luo. 2023. Kracl: Contrastive learning with graph context modeling for sparse knowledge graph completion. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 2548–2559, New York, NY, USA. Association for Computing Machinery.

Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd workshop on continuous vector space models and their compositionality*, pages 57–66.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2071–2080, New York, New York, USA. PMLR.

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. 2019. Composition-based multi-relational graph convolutional networks. *arXiv preprint arXiv:1911.03082*.

Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. 2022. SimKGC: Simple contrastive knowledge graph completion with pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4281–4294, Dublin, Ireland. Association for Computational Linguistics.

Xintao Wang, Qianyu He, Jiaqing Liang, and Yanghua Xiao. 2023. Language models as knowledge embeddings. *Preprint*, arXiv:2206.12617.

Yanbin Wei, Qiushi Huang, James T Kwok, and Yu Zhang. 2024. Kicgpt: Large language model with knowledge in context for knowledge graph completion. *arXiv preprint arXiv:2402.02389*.

Tobias Weller and Heiko Paulheim. 2021. Evidential relational-graph convolutional networks for entity classification in knowledge graphs. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3533–3537.

Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. 2016. Representation learning of knowledge graphs with entity descriptions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*.

Honggen Zhang, June Zhang, and Igor Molybog. 2024. Hasa: Hardness and structure-aware contrastive knowledge graph embedding. In *Proceedings of the ACM on Web Conference 2024*, pages 2116–2127.

Zhanqiu Zhang, Jianyu Cai, and Jie Wang. 2020. Duality-induced regularizer for tensor factorization based knowledge graph completion. *Advances in Neural Information Processing Systems*, 33:21604–21615.

Zhanqiu Zhang, Jianyu Wang, Jiawei Ye, Xiang Ren, Meng Zhang, and Wei Chen. 2022. Rethinking graph convolutional networks in knowledge graph completion. In *Proceedings of the ACM Web Conference 2022*, pages 798–807. ACM.

Model	WN18RR		FB15k-237		Kinship		UMLS		YAGO3-10	
	α	β	α	β	α	β	α	β	α	β
TB-based ($\lambda = 0.05$)										
CP	0.05	0.15	0.025	0.075	0.025	0.075	0.025	0.075	0.005	0.015
ComplEx	0.05	0.15	0.025	0.075	0.025	0.075	0.025	0.075	0.005	0.015
DisMult	0.05	0.15	0.025	0.075	0.025	0.075	0.025	0.075	0.005	0.015
Translation-based ($\lambda = 0.05$)										
TransE	0.7	0.1	0.4	0.25	0.8	0.1	0.4	0.1	-	-
RotatE	0.7	0.1	0.4	0.25	0.8	0.1	0.8	0.2	-	-
PairE	0.7	0.1	0.6	0.1	0.8	0.1	0.8	0.2	-	-
PLM-based ($\lambda = 0.05$)										
SimKGC	0.4	0.6	0.5	0.5	-	-	-	-	-	-
PMD	0.4	0.6	0.5	0.5	-	-	-	-	-	-

Table 5: Models with corresponding α_1/β_1 and α_2/β_2 values for different datasets.

A Hyperparameters

Table 5 provides the optimal settings for the hyperparameters used. Specifically, for a fair comparison, we adhere to the setups of various KGE models and use the same hyperparameters as presented in the original papers. The newly introduced hyperparameters α, β, and λ are selected via grid search. The parameter λ is consistently set to 0.05 across all experiments. For different KGE model classes, we adjust the search range accordingly: for TB-based models, since excessive regularization can hinder convergence, we search α and β within $\{0.3, 0.15, 0.075, 0.05, 0.025, 0.015, 0.005\}$. While for Translation-based and PLM-based models, the range is set from 0.1 to 1.0. Based on our experiments, we found that sharing hyperparameters within the same type of model generally yields better performance.

B Datasets

In this work, we evaluate our proposed methods on several widely used benchmark datasets. These datasets cover both link prediction and entity classification tasks, and vary in terms of size, number of entities, relations, and edges. Detailed statistics for each dataset are summarized in Tables 6 and 7. The diversity and scale of these datasets allow for a comprehensive assessment of the effectiveness and generalizability of our approach.

Datasets	Entities	Relations	Train	Valid	Test	Total Triples
WN18RR	40,943	11	86,835	3,034	3,134	93,003
FB15k-237	14,541	237	272,115	17,535	20,466	310,116
YAGO3-10	123,182	37	1,079,040	5,000	5,000	1,089,040
UMLS	135	46	5,216	652	661	6,529
Kinship	104	25	8,544	1,068	1,074	10,686

Table 6: Six benchmark datasets for link prediction.

Datasets	Entities	Relations	Edges	Classes
AIFB	8,285	45	29,043	4
MUTAG	23,644	23	74,227	2

Table 7: Four benchmark datasets for entity classification.

C More Details on Sparsity Analysis

To evaluate the robustness of our method in sparse scenarios, we conducted a sparsity analysis by randomly selecting factual triples to create reduced training subsets of the FB15k-237 dataset. As shown in Fig. 5, as the size of the training set decreased, our model consistently exhibited low standard deviations in performance metrics, outperforming the baseline models across all settings. This demonstrates the SaCa framework's ability to maintain robust embeddings even when trained with limited data, underscoring its effectiveness in capturing meaningful semantic relationships in knowledge graphs under sparse conditions.

Furthermore, we performed a statistical analysis of the out-degree and relationship context quantities at various sparsity levels for the FB15k-237 dataset, as shown in Table 8. In sparse KGs, the sparsity ratio directly impacts the out-degree of entities and the availability of 1-hop contexts. Sparse graphs typically feature fewer entity connections, leading to a reduction in the richness of 1-hop contexts. However, even in highly sparse datasets, such as the FB15k-237 dataset with 10% sparsity, the 2-hop contexts remain relatively rich. This ability to maintain rich higher-hop contexts is a key strength of our approach, ensuring that the model can still capture meaningful relationships despite data sparsity.

Data Subset	Out-degree	Avg 1-hop for E	Avg 1-hop for ER	Avg 2-hop for E	Avg 2-hop for ER
Original	18.76	37.52	18.76	1330.91	230.95
ratios 0.1	2.35	4.71	1.70	176.85	63.97
ratios 0.2	4.15	8.30	2.04	277.25	88.29
ratios 0.3	5.96	11.92	2.32	396.82	110.72
ratios 0.4	7.78	15.56	2.56	547.71	149.28

Table 8: Sparsity levels for the FB15k-237 datasets.

Models	UMLS				Kinship			
	MRR	**Hit@1**	**Hit@3**	**Hit@10**	**MRR**	**Hit@1**	**Hit@3**	**Hit@10**
Translation Based Models								
TransE	0.704	0.547	0.835	0.935	0.298	0.113	0.375	**0.684**
TransE-SaCa	**0.708**	**0.553**	**0.839**	**0.936**	**0.306**	**0.120**	**0.394**	0.682
RotatE	0.760	0.621	**0.877**	**0.953**	0.645	0.500	0.741	0.925
RotatE-SaCa	**0.764**	**0.632**	0.874	0.951	**0.662**	**0.521**	**0.758**	**0.930**
PairRE	0.835	0.740	0.913	0.979	0.543	0.379	0.631	0.890
PairRE-SaCa	**0.844**	**0.756**	**0.917**	**0.980**	**0.559**	**0.392**	**0.659**	**0.920**
Tensor Decomposition Based Models								
CP	0.789	0.679	0.879	0.956	0.633	0.484	0.736	0.928
CP-N3	0.819	0.717	0.910	0.966	0.664	0.519	0.768	0.940
CP-DURA	0.810	0.706	0.897	0.962	0.688	0.562	0.792	0.948
CP-SaCa	**0.849**	**0.760**	**0.932**	**0.972**	**0.711**	**0.580**	**0.809**	**0.958**
ComplEx	0.783	0.690	0.848	0.948	0.647	0.497	0.749	0.938
ComplEx-N3	0.796	0.707	0.864	0.950	0.654	0.502	0.763	0.943
ComplEx-DURA	0.812	0.715	0.872	0.962	0.662	0.497	0.778	0.944
ComplEx-SaCa	**0.861**	**0.767**	**0.920**	**0.982**	**0.704**	**0.567**	**0.806**	**0.955**

Table 9: Performance Comparison on UMLS and Kinship Datasets

Model	MRR	Hit@1	Hit@10
RotatE	0.495	0.402	0.670
HAKE (Li et al., 2022)	0.546	0.462	0.694
MQuinE (Shang et al., 2024)	0.566	0.492	0.711
KG-FIT (Jiang et al., 2024)	0.568	0.472	**0.718**
CP	0.531	0.467	0.687
CP-SaCa	0.583	0.511	0.710
ComplEx	0.541	0.482	0.693
ComplEx-SaCa	**0.589**	**0.518**	0.715

Table 10: Link prediction on YAGO3-10.

D Efficiency Analysis

The efficiency of our framework is evaluated in two stages: structural proximity measurement during preprocessing and the model training phase. In preprocessing, the Jaccard computation is performed once per dataset, as relational contexts are deterministic for a fixed knowledge graph. Directly computing Jaccard similarity for every entity pair can be computationally expensive, especially with large datasets. To address this, we employ three synergistic mechanisms: (1) a feature-context inverted index that precomputes entity affiliations per structural feature, reducing search space from global comparisons to local feature-specific intersections; (2) frequency-based filtering to exclude high-frequency, low-discriminative features, and (3) parallel processing to distribute the workload across multiple cores. Additionally, we limit the maximum path length to $k \leq 4$. Previous works (Lin et al., 2015a; Sun et al., 2018) show that considering excessively long relation paths is often unnecessary, as semantically similar entities are usually not extremely far apart. In contrast to the conventional Jaccard similarity approach—which generally incurs a time complexity of $O(N^2 \cdot M)$ for N entities with an average context size of M—our method effectively reduces the complexity to $\sum_{f \in F} O({m_f}^2)$, where m_f denotes the frequency of feature f. Given that most features occur infrequently, and high-frequency features are filtered out, the summation $\sum_{f \in F} O({m_f}^2)$ is typically much smaller than N^2, leading to significant efficiency gains, especially when enhanced by parallel execution. Specifically, for larger datasets such as FB15k-237 and WN18RR, the context extraction takes approximately 35 minutes and 12 minutes, respectively, on an Intel Xeon Platinum 8352V CPU.

Entity	Similar Entity	Joint 1-hop Relational Context	Joint 2-hop Relational Context
Taylor Swift	Miley Cyrus (0.84), Joe Jonas (0.79), Justin Bieber (0.76)	('/common/topic/webpage/category', 'Official Website'), ('/people/person/profession', 'Musician-GB'), ('/people/person/profession', 'Singer-songwriter-GB')	('celebrity/friendship/participant', '/people/person/nationality', 'USA'), ('/people/person/profession', '/people/specialization_of', 'Artist-GB')
Bill Clinton	Hillary Rodham (0.87), George W. Bush (0.71), Barack Obama (0.68)	('/people/person/profession', 'Politician-GB'), ('/government_position_office', 'USA')	('/government_position_of_office', '/location/statistical_region/exported_to', 'Angola')

Table 11: Nearest neighbor analysis in FB15k-237.

In the training phase, our framework achieves linear time complexity, specifically $O(N(B+K-1)d)$. Here, N denotes the number of triplets, B is the batch size, and d is the feature dimension. This complexity arises from two main components: (1) the Dynamic Alignment Loss, with $O(NKd)$ complexity for processing K positive sample pairs per triplet and (2) the Dissimilarity Loss, with $O(N(B-1)d)$ complexity for in-batch negative sample comparisons. In other words, the computational complexity of our SemCa framework is comparable to that of a supervised contrastive loss function applied to K positive samples, ensuring computational efficiency. Empirically, for tensor factorization models, the average extra training time is 5.6 minutes for WN18RR and 13.6 minutes for FB15k-237, with negligible overhead for smaller datasets like Kinship and UMLS. For text-based KGE methods with dual-encoder architectures (e.g., SimKGC and PMD), integrating SaCa increased running times by merely 15 minutes on WN18RR and 24 minutes on FB15k-237, respectively. This highlights that the computational overhead of SaCa is modest and manageable, facilitating substantial performance gains without a significant impact on overall efficiency.

E Supplementary Link Prediction Results

E.1 Results on UMLS and Kinship

We present supplementary link prediction results on the UMLS and Kinship datasets, further evaluating our SaCa framework. These smaller benchmarks provide additional insights into SaCa's performance when integrated with various KGE baselines.

As detailed in Table 9, applying SaCa generally enhances the performance of the baseline models. For instance, PairRE-SaCa, CP-SaCa, and ComplEx-SaCa demonstrate consistent improvements across all reported metrics (MRR, Hit@1, Hit@3, Hit@10) on both UMLS and Kinship datasets when compared to their original versions. Notably, for the CP and ComplEx baselines, the SaCa-enhanced versions also clearly outperform other comparative enhancement techniques like N3 and DURA across all metrics. The overall results on these datasets complement the findings on larger benchmarks, illustrating SaCa's broad applicability and significant benefits.

E.2 Results on YAGO3-10

To further assess the scalability of SemCa, we conduct experiments on the large-scale YAGO3-10 dataset ($\sim$1M triples), as shown in Table 10. Notably, YAGO3-10 poses significant demands on both model capacity and computational efficiency. Specifically, it boosts CP and ComplEx by an average of 9.4% in MRR and 8.5% in Hit@1. These results highlight SemCa's robust performance and practical value for large knowledge graphs.

F Case Study

To gain a clearer understanding of how the embedding space encodes semantic relationships, we conducted a nearest neighbor case study, as shown in Table 11. We examined *Taylor Swift* and *Clinton*, who, despite both being American figures, represent distinct semantic prototypes: the former aligns with cultural domains such as singers and actors, while the latter is consistently associated with political figures. This contrast underscores the model's capacity to capture fine-grained semantic distinctions between entities across different domains.