BiCore: Bicomplex and Fisher-Rao Fusion for Multimodal Knowledge Graph Completion

Abstract. Knowledge base completion is a fundamental task in knowledge graph research, aimed at predicting unseen relationships between entities in Knowledge Graphs (KGs). Traditional Knowledge Graph Completion (KGC) models have shown promising results but struggle when incorporating multimodal information—especially textual and visual data. These models are typically designed for a single modality, limiting their ability to capture complex interactions in multimodal KGs and ultimately reducing reasoning accuracy. To address these limitations, we propose BiCore, a novel framework for multimodal KGC tasks, including analogy reasoning and link prediction. BiCore leverages BiComplex Embeddings to represent relationships in a richer, more expressive space, enhancing reasoning capabilities in multimodal environments. By effectively integrating textual and visual modalities using Transformer-based models and Riemannian fusion, BiCore provides a structured and robust multimodal knowledge representation. Experiments on benchmark datasets including DB15K, MKG-W, MKG-Y, and MARS demonstrate substantial improvements in performance metrics such as Mean Reciprocal Rank (MRR) and H@K. BiCore surpasses traditional models, achieving accuracy improvements up to 27.5%. More importantly, BiComplex Embeddings enable a more effective fusion of heterogeneous data sources, addressing a key challenge in multimodal KGC.

Keywords: Multimodal Knowledge Base Completion, BiComplex Embeddings, Analogy Reasoning, Riemannian Fusion

1. Introduction

A Knowledge Graph (KG) is a structured representation of knowledge where entities (nodes) are linked by relations (edges) to form a network that enables semantic reasoning and knowledge retrieval. KGs have been widely adopted in recommendation systems (Xian et al., 2019; J. Yu et al., 2022; Zhao et al., 2020), question answering (QA) (Kaiser et al., 2021), semantic search (Y. Li, 2017; Opdahl et al., 2022), and automated reasoning (Bellomarini et al., 2022; Qin & Chow, 2019), providing a structured framework for AI applications. Traditionally, KGs store information as triplets (*head entity*, *relation*, *tail entity*), effectively capturing structured knowledge. However, conventional KGs primarily rely on symbolic relational data, often omitting multimodal information such as textual descriptions, images, and sensory data—critical components of real-world knowledge representation.

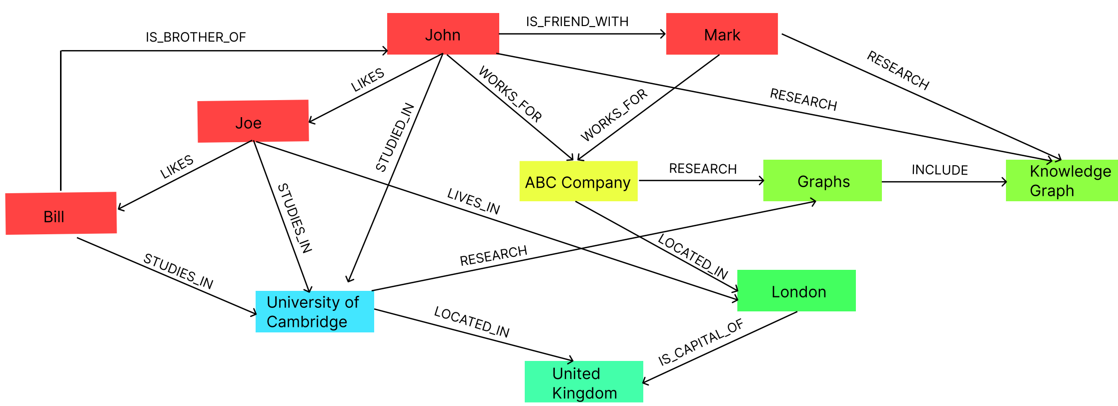

Fig. 1. An illustrative example of a Knowledge Graph capturing semantic relationships between entities such as individuals (e.g., *John*, *Joe*, *Mark*), organizations (e.g., *ABC Company*), locations (e.g., *London*, *United Kingdom*), and concepts. It demonstrates diverse relation types including *IS_BROTHER_OF*, *STUDIES_IN*, *WORKS_FOR*, and *RESEARCH*, highlighting the rich structural and semantic heterogeneity inherent in KGs.

To overcome this limitation, Multimodal Knowledge Graphs (MKGs) have been introduced as an enhanced extension of KGs, integrating text, images, and other sensory data alongside structured triplets. Unlike traditional KGs, which rely solely on textual and structural information, MKGs provide richer, context-aware representations, improving their ability to support AI-driven reasoning. Fig. 1 illustrates a conventional KG, showing structured triplets such as (*John*, *WORKS_FOR*, *ABC Company*) or (*London*, *IS_CAPITAL_OF*, *United Kingdom*). In contrast, Fig. 2 depicts an MKG, where entities are enriched with visual elements, such as an image of London complementing its textual description, thereby providing deeper contextual information.

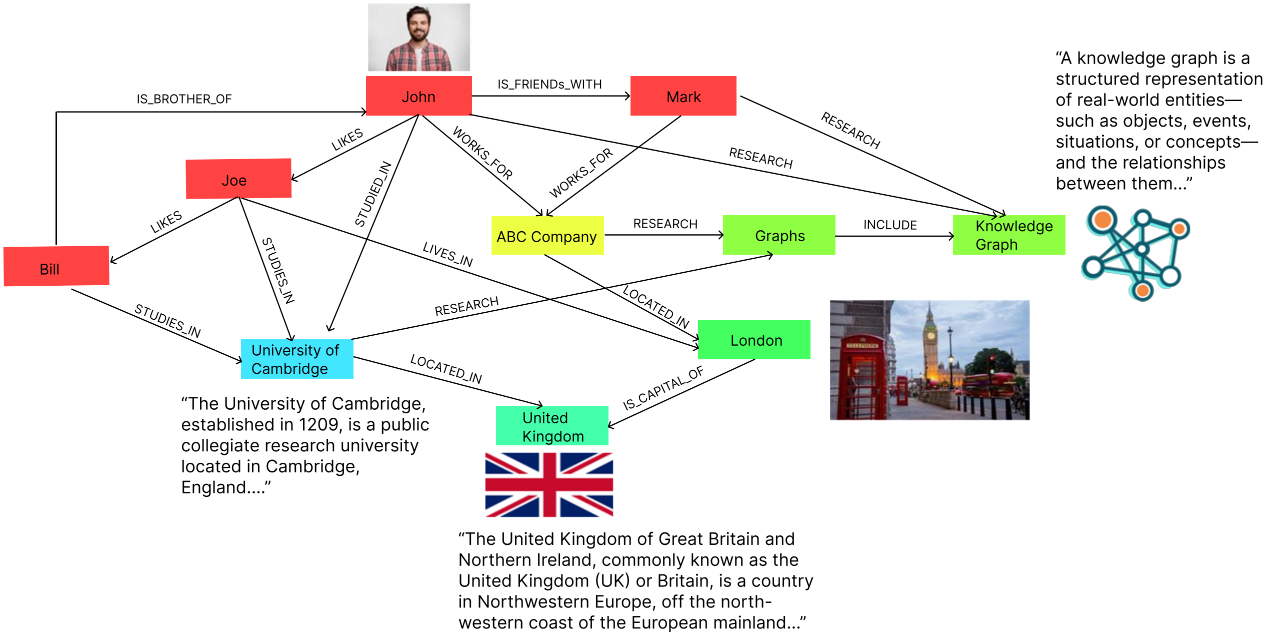

Fig. 2. A Multimodal Knowledge Graph that integrates both textual descriptions and images to enrich entity representations. This example illustrates how entities—such as people, organizations, and locations—can be augmented with modality-specific context, where images provide visual grounding and text offers semantic depth. By combining structured triples with heterogeneous data sources, the graph highlights the expressive power and contextual richness enabled by multimodal knowledge representation.

Despite their advantages, MKGs suffer from knowledge incompleteness, where essential relationships between entities remain missing, limiting their effectiveness in real-world AI applications. Knowledge Graph Completion (KGC) has emerged as a fundamental research challenge, aiming to predict missing links between entities and enhance the reasoning capability of KGs. While existing KGC methods have significantly improved link prediction accuracy, they predominantly focus on single-modal settings, leveraging only structural and textual information. However, with the increasing integration of multimodal data, modern KGs demand more advanced KGC techniques capable of handling heterogeneous information. The inability of traditional KGC models to effectively incorporate multimodal data presents a critical bottleneck, limiting their reasoning capacity in real-world applications. Most existing models treat multimodal information as auxiliary features, failing to fully capture cross-modal dependencies. Addressing this challenge requires novel methodologies that can seamlessly align and integrate multiple modalities in a unified and meaningful way.

In recent years, several embedding-based models have been proposed to enhance link prediction in KGs. Methods such as TransE (Bordes et al., 2013), DistMult (B. Yang et al., 2015), and ComplEx (Trouillon et al., 2016) have demonstrated significant improvements by learning low-dimensional representations of entities and relations in Euclidean or complex-valued spaces. While these models are effective in modeling structured knowledge, they struggle when dealing with multimodal information. The primary challenge stems from the inherent differences between textual and visual representations, which operate in fundamentally distinct semantic and spatial domains. Existing models fail to establish a coherent framework that can bridge this gap, leading to suboptimal performance in multimodal knowledge inference. Additionally, traditional KGC models often exhibit difficulties in capturing asymmetric and cyclic relationships, which are essential for modeling real-world KGs. Although ComplEx partially addresses this issue by leveraging complex-valued embeddings, it remains insufficient for handling multimodal data, where the interaction between textual and visual modalities introduces additional layers of complexity.

Another significant limitation of existing KGC methods is their inability to effectively fuse information from multiple sources while preserving the unique properties of each modality. Recent deep learning approaches, such as ConvE (Dettmers et al., 2018) has attempted to incorporate multimodal data by employing convolutional neural networks (CNNs) or graph neural networks (GNNs). However, these models often suffer from inefficient cross-modal synchronization, as they lack a principled mechanism for integrating textual and visual embeddings in a unified space. The fundamental challenge lies in aligning heterogeneous representations while maintaining the integrity of their respective semantic structures. Without an effective fusion strategy, these models fail to fully exploit the potential of multimodal data, leading to inconsistencies in reasoning and limiting their applicability to real-world KGC tasks.

Moreover, recognizing these challenges, several recent studies have proposed advanced deep learning-based models to improve multimodal knowledge graph completion (MKGC). For instance, MoCi (Modality Circular Fusion) (J. Yang et al., 2024) introduces a triplets-prompt modality contrastive pre-training strategy to align semantic meaning across different modalities before fusion. Unlike previous methods that directly merge multimodal embeddings, MoCi employs modality circular fusion, allowing for inter-entity multimodal interactions rather than limiting fusion within individual entities. This method significantly enhances the model's ability to capture contextual relationships across different data modalities, leading to improved link prediction accuracy. Another promising approach is M2KGRL (Multimodal Knowledge Graph Representation Learning) (Chen et al., 2025), which integrates VGG16, BERT, and SimplE to extract and fuse multimodal features from structured triples, images, and textual data. M2KGRL utilizes a customized autoencoder to enhance the quality of multimodal embeddings and introduces a similarity-based scoring function to guide representation learning. This framework has demonstrated good performance over traditional KGC models by improving entity representation quality and enhancing semantic matching across modalities. Similarly, ML-Former (Multimodal Learning Transformer) (Wu et al., 2024) presents a hybrid transformer-based architecture designed to address imbalanced modality contributions. Traditional multimodal models often suffer from "modality laziness," where dominant modalities (e.g., text) overshadow less informative ones (e.g., images), leading to biased learning. ML-Former mitigates this issue by implementing deep interaction and perceptual learning modules, ensuring balanced contributions from all modalities. By leveraging self-attention mechanisms and a multi-tiered fusion strategy, ML-Former achieves a more robust and interpretable multimodal representation, making it particularly effective in knowledge graph reasoning tasks. While these models have made significant progress in multimodal KGC, they still face key challenges in capturing complex entity relationships and ensuring robust multimodal alignment, highlighting the need for a more structured and principled solution.

To overcome these challenges, we propose BiCore, a novel multimodal KGC framework that introduces BiComplex embeddings and Riemannian fusion to enhance knowledge graph reasoning. BiComplex embeddings extend traditional complex-valued embeddings, enabling more expressive representations of asymmetric and cyclic relationships in KGs. Unlike existing models that struggle with multimodal fusion, BiCore employs Riemannian fusion, a novel geometric integration technique that allows for structured alignment of textual and visual information. This approach not only preserves the distinct properties of each modality but also facilitates their seamless integration in a unified embedding space, leading to more effective knowledge representation. More importantly, the integration of BiComplex embeddings and Riemannian fusion enables BiCore to address the long-standing issue of heterogeneous data fusion, making it a robust and scalable solution for MKGC. By tackling the fundamental limitations of existing models, BiCore presents a significant advancement in MKGC research, offering a principled and scalable approach for enhancing link prediction, analogy reasoning, and multimodal knowledge representation.

1.1. Research objectives

The primary objective of this research is to develop an innovative framework for MKGC that integrates textual, visual, and structural data to enhance the accuracy and expressiveness of knowledge graph embeddings. This research aims to achieve three key objectives: First, we aim to integrate multimodal data on a specific manifold, creating a robust representation space capable of handling diverse data types such as text, images, and structural relationships. Second, we seek to extend complex embeddings to model various relational patterns, including symmetric, asymmetric, and cyclic relations, within KGs, thereby improving relational reasoning in multimodal environments. Third, we focus on optimizing the alignment of multimodal embeddings using some special distances, ensuring that both the semantic and spatial properties of the data are preserved across different modalities.

1.2. Contributions

Our main contributions are as follows:

- We propose BiCore, a novel multimodal KGC framework integrating BiComplex embeddings with Riemannian fusion. BiCore extends complex-valued representations to capture both symmetric and asymmetric relationships, enhancing reasoning in multimodal environments.
- We introduce a principled approach for integrating textual and visual modalities while preserving their semantic and spatial integrity. BiCore employs Riemannian fusion to align and merge heterogeneous data within a geometrically structured space, improving cross-modal reasoning.
- We address modality imbalance by designing a balanced fusion mechanism. BiCore ensures that both dominant (e.g., text) and less prominent (e.g., images) modalities contribute meaningfully to knowledge representation, improving model robustness.
- We empirically validate BiCore on benchmark datasets, showing that it outperforms state-of-the-art models in link prediction and analogy reasoning on DB15K, MKG-W, MKG-Y, and MARS, achieving notable improvements in MRR and H@K.

The remainder of this paper is structured as follows. Section 2 reviews related work on multimodal KGC and advancements in embedding techniques. Section 3 provides the theoretical background on multimodal knowledge graphs, BiComplex numbers, and Riemannian manifolds, establishing the foundation for our proposed approach. Section 4 introduces BiCore, detailing its architecture, methodology, and multimodal fusion strategy. Section 5 presents comprehensive experiments, evaluating BiCore against baseline models on multiple benchmark datasets, and analyzes the influence of regularization and embedding dimensions. Section 6 is a case study. Finally, Section 7 concludes the paper, summarizing the key contributions and discussing future research directions.

2. Related work

2.1. Knowledge graph completion

Knowledge Graph Completion (KGC) task is fundamental in predicting missing relationships between entities. KGC methods have evolved across three primary categories: embedding-based methods, graph-based methods, and deep learning-based methods.

One of the earliest and most influential KGC models is TransE (Bordes et al., 2013), which represents relationships as translations in Euclidean space. However, TransE struggles with one-to-many, many-to-one, and many-to-many relationships, as simple translations fail to capture complex relational structures. To address these shortcomings, RotatE (Sun et al., 2019) extended relational modeling using rotations in the complex space, allowing the model to capture symmetric, anti-symmetric, and cyclic relations. Similarly, ComplEx (Trouillon et al., 2016) introduced complex-valued embeddings to handle hierarchical relationships, while TuckER (Balazevic et al., 2019) applied tensor decomposition to enhance the model's ability to capture non-linear relations. Despite their effectiveness, these methods are limited to unimodal structured data, making them insufficient for tasks requiring multimodal knowledge representation or external reasoning sources.

Beyond embedding-based approaches, graph-based methods have also significantly contributed to KGC by utilizing graph structures for relational reasoning. One of the earliest such methods is Path Ranking Algorithm (PRA) (Lao et al., 2011), which explores multi-step paths in KGs to predict missing links. However, PRA suffers from poor scalability due to its high computational complexity. To address this, R-GCN (Schlichtkrull et al., 2018) introduced Graph Neural Networks (GNNs) (J. Zhou et al., 2020) for learning structured representations in KGs, allowing for message passing between neighboring entities to improve generalization. In addition, hypercomplex embedding methods have gained traction in KGC. QuatE (S. Zhang et al., 2019) extended complex-valued embeddings to quaternions for capturing richer relational structures, while OctonionE (M. Yu et al., 2022) introduced higher-dimensional number spaces to enhance expressiveness. However, these methods remain limited by their reliance on Euclidean space, which struggles to model hierarchical and structured knowledge efficiently. To overcome this limitation, recent works such as MuRP (Balažević et al., 2019), HBE (Pan & Wang, 2021), and CAKE (Niu et al., 2022) have leveraged Riemannian geometry to embed KGs into curved spaces, improving their ability to represent complex, hierarchical knowledge structures. However, these approaches do not integrate bicomplex embeddings or multimodal representations, which are central to our work on BiCore

Despite these advances, long-tail entity learning remains a significant challenge in KGC, as most models fail to generalize well to sparse and low-resource entities. This has led to the emergence of hybrid KGC approaches that integrate advanced learning mechanisms. For example, MMRNS (Xu et al., 2022) introduced Riemannian fusion with adaptive attention mechanisms, allowing them to capture complex relations more efficiently in large-scale KGs. Recently, KG-TRICK (Z. Zhou et al., 2025) proposed an innovative approach by unifying KGC and Knowledge Graph Enhancement (KGE) into a single framework. Traditionally, KGC focuses on predicting missing relations, while KGE aims to predict missing textual attributes of entities. Prior methods treated these as separate tasks, but KG-TRICK hypothesizes that they are interdependent and can enhance each other. By leveraging multilingual textual data, KG-TRICK improves both relational link prediction and textual entity completion, making KGs more complete.

With the rise of deep learning, Transformer-based models have also become widely adopted for KGC. KG-BERT (Yao et al., 2019) was among the first methods to apply Transformers for knowledge graph representation learning, enabling contextualized entity and relation embeddings. However, KG-BERT suffers from high computational costs, making it difficult to scale to large KGs. To address scalability and reasoning limitations, retrieval-augmented approaches like KICGPT (Wei et al., 2023) have emerged. KICGPT integrates a knowledge graph retriever with large language models (LLMs) to mitigate the long-tail entity problem. Unlike purely embedding-based or text-based methods, KICGPT leverages in-context learning (ICL) strategies to combine structured KG knowledge with LLM reasoning, ensuring more effective entity ranking. By structuring KGC as a re-ranking process, KICGPT prevents LLM hallucination while reducing computational overhead.

Overall, while traditional methods have significantly improved link prediction and KGC, they still fail to fully leverage structured knowledge alongside external reasoning resources, which are crucial for real-world applications requiring multimodal integration and large-scale reasoning. This gap has driven the evolution of retrieval-augmented and multimodal-enhanced KGC, a new research direction that aims to integrate structured knowledge with external reasoning for improved entity prediction and link inference.

2.2. Multimodal knowledge graph completion

Multimodal Knowledge Graph Completion (MKGC) extends traditional KGC by incorporating heterogeneous data modalities, such as text, images, and audio, alongside structured relational data. Unlike conventional KGC, which primarily relies on structured triplets, MKGC enriches representation learning by integrating multiple data sources, enabling more expressive and contextually grounded reasoning. This is particularly important because real-world knowledge is inherently multimodal, where entities are associated with textual descriptions, visual representations, and other unstructured data.

Despite its potential, MKGC still faces several critical challenges that hinder its effectiveness in large-scale applications. One of the most significant challenges is the alignment and fusion of multimodal information. Traditional KGE models like TransE, ComplEx, and RotatE were designed for structured relational data, making them unsuitable for processing heterogeneous multimodal inputs, which leads to suboptimal reasoning performance in MKGs.

Additionally, existing MKGC models struggle to capture intricate cross-modal dependencies. Many methods rely on static concatenation-based fusion, as seen in VilBERT (Lu et al., 2019), where textual and visual embeddings are fused at a fixed stage. While these approaches have demonstrated success in image captioning and multimodal question answering, they lack flexibility in capturing complex interdependencies across modalities, which is crucial for multimodal knowledge graph reasoning.

To address these challenges, researchers have explored adaptive multimodal fusion mechanisms and graph-aware multimodal reasoning techniques. LAFA (Vu et al., 2024), for example, integrates multimodal information by propagating signals from neighboring nodes within the graph structure. While this method enhances local feature propagation, it remains limited by the inflexibility of traditional KG embeddings, which struggle to model dynamic multimodal interactions. Another promising research avenue focuses on adaptive multimodal fusion with curvature-aware embeddings. RSME (M. Wang et al., 2021) introduces modality-specific relevance mechanisms, ensuring that only the most informative images and text are utilized for link prediction. This prevents irrelevant multimodal data from diluting reasoning signals. However, while RSME and similar models improve modality filtering, they lack the incorporation of geometric dynamics in embedding learning, which is a core strength of BiCore.

An alternative approach to multimodal knowledge reasoning leverages graph-based methods by integrating Graph Neural Networks (GNNs) with multimodal embeddings to enhance link prediction accuracy. MMKGR (Zheng et al., 2023) employs reinforcement learning-driven multi-hop reasoning, enabling models to traverse MKGs and uncover deeper relational structures.

To further bridge multimodal integration with real-world applications, Multi-KG4Rec (J. Wang et al., 2025) proposes a unified multimodal fusion framework tailored for personalized recommendation systems. Unlike conventional feature-based and entity-based methods, Multi-KG4Rec decomposes multimodal data into individual modality-specific graphs, ensuring a more precise alignment between textual, visual, and structured information. By leveraging pre-trained LLMs and GNNs to refine multimodal embeddings and incorporating a cross-modal attention mechanism for fine-grained feature extraction, Multi-KG4Rec significantly improves personalized recommendation accuracy, demonstrating the growing importance of adaptive multimodal reasoning in AI-driven applications.

3. Background

3.1. Preliminaries

A Multimodal Knowledge Graph (MKG) is an enriched knowledge base where nodes represent entities and edges denote relationships between these entities, supplemented with information from multiple modalities such as text, images, and audio. Unlike traditional KGs that rely solely on structured textual triplets, MKGs integrate multimodal data to provide a more comprehensive representation of real-world knowledge. Formally, an MKG can be represented as $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$, where $\mathcal{E}$ is the set of entities, $\mathcal{R}$ is the set of relations, and $\mathcal{T}$ is the set of triples in the graph. Table 1 summarizes the notations and their corresponding descriptions used throughout this paper.

Table 1
Summary of notations and their descriptions used throughout our framework.

Notation	Description
$\mathcal{G}$	A knowledge graph
$\mathcal{E}$	A set of entities
$\mathcal{R}$	A set of relations
$\mathcal{F}$	A set of triplets
h	The head entity of a triplet
r	A relation that connects the head entity and the tail entity of a triplet
t	The tail entity of a triplet
e_h	The head entity embedding vector in representation space $\mathbb{R}^d$
e_r	The relation embedding vector in representation space $\mathbb{R}^d$
e_t	The tail entity embedding vector in representation space $\mathbb{R}^d$
Z	A bicomplex number (representation of entities and relations in the bicomplex space)
ψ_1, ψ_2	Elements on the unit sphere used for Fisher-Rao distance calculation
e_{lv}	The fused multimodal embedding representing the integration of linguistic and visual modalities
e_F	The fused multimodal embedding for an entity after integration of different modality-specific embeddings
e_{TF}	The transformer-refined multimodal embedding
$f(h, r, t)$	Scoring function for the knowledge graph triple (head entity, relation, tail entity)

For instance, a multimodal triple (*Donald Trump, is_president_of, US*) may include a textual description explaining the relationship, alongside an image depicting *Donald Trump* and a visual representation of the *United States*. However, in MKGs, multimodal data sources are often incomplete or misaligned, leading to missing modalities or inconsistencies across different data types. To address these challenges, MKG embedding models aim to project entities, relations, and multimodal attributes into a shared low-dimensional embedding space, ensuring that the semantic and multimodal properties of the data are preserved.

The core objective of Multimodal Knowledge Graph Embedding (MKGE) is to facilitate knowledge completion by predicting missing entities and relationships across modalities. Given a query $(h, r, ?)$ or $(?, r, t)$, the embedding model predicts the missing entity using a scoring function $f: \mathcal{E} \times \mathcal{R} \times \mathcal{E} \rightarrow \mathbb{R}$, which quantifies the plausibility of a multimodal relationship. High-scoring multimodal triples indicate valid relationships, forming new links within the graph. A non-linear activation function, such as sigmoid (σ), is applied to convert the computed score into a probability $p = \sigma(f) \in [0, 1]$, determining the likelihood of the relationship's existence in a given modality.

Building an effective MKGE model involves four primary components:

Embedding Vectors. The MKG embedding process maps entities, relations, and modalities (e.g., text, image, audio) into a shared low-dimensional space, ensuring that the learned representations preserve semantic consistency across modalities. The embeddings for multimodal triples (h, r, t) are represented as vectors in $\mathbb{R}^{d}$, where d is the embedding dimension.

Scoring Function. A scoring function evaluates the plausibility of a multimodal triple. A valid triple, such as (*Donald Trump, is_president_of, US*), should receive a higher score than an incorrect one, such as (*Donald Trump, is_president_of, France*). The choice of scoring function influences the ability of the model to capture relational dependencies across modalities.

Negative Sampling. To enhance model training, negative sampling generates incorrect triples, forcing the model to distinguish between true and false relationships. In a multimodal setting, negative examples must respect modality constraints—for example, avoiding mismatched image-text pairings that would create unrealistic triples. Advanced techniques such as adversarial negative sampling can further improve the robustness of the model by generating contextually and temporally relevant negative samples.

Loss Function. A loss function $\mathcal{L}$ optimizes the model by adjusting embeddings so that valid multimodal triples receive higher scores than invalid ones. This ensures that learned representations encode both semantic and multimodal relationships. Common optimization methods, including Adam (Kingma & Ba, 2014), Adagrad (Duchi et al., 2011), and SGD (P. Zhou et al., 2020), are employed to minimize the loss, enabling the model to generalize and predict new links effectively.

3.2. Bicomplex numbers

One of the key challenges in MKGC is the need for a representation space that effectively models complex relational structures while integrating diverse modalities, such as text and images. Traditional complex-valued embeddings have been widely used to enhance knowledge representation, as they naturally support asymmetric and cyclic relations in KGs. However, standard complex numbers operate in a two-dimensional space, limiting their ability to capture intricate multimodal interactions. To address this limitation, bicomplex numbers extend complex numbers by introducing a second imaginary unit, resulting in a four-dimensional representation space. This extension provides a more powerful framework for encoding higher-order relationships in KGs. Unlike traditional complex embeddings, constrained to a single imaginary dimension, bicomplex embeddings allow for richer algebraic structures, making them particularly effective for representing multimodal information. By leveraging bicomplex embeddings, we enhance relational modeling in MKGC, enabling more expressive link prediction and analogy reasoning tasks.

In the bicomplex space, entities and relations in a knowledge graph are represented as bicomplex numbers. Each component of a bicomplex number corresponds to different aspects of an entity's attributes and interactions with other entities. This four-dimensional structure accommodates more complex relationships and seamlessly integrates information from multiple modalities. For instance, in multimodal knowledge graphs, bicomplex embeddings can simultaneously capture textual semantics and visual features, allowing for a unified representation of diverse data types. The algebraic properties of bicomplex numbers, including their ability to support transformations like rotations and similarity-based mappings, further enhance their utility in modeling entity relationships, offering richer and more flexible representations than traditional embeddings.

A bicomplex number is defined as:

$$Z = z_1 + jz_2 = (a_1 + ia_2) + j(b_1 + ib_2) \tag{1}$$

where $a_1, a_2, b_1, b_2 \in \mathbb{R}$, and i and j are two distinct imaginary units, satisfying the following relations:

$$ij = ji, i^2 = j^2 = -1 \tag{2}$$

This formulation extends the traditional complex number system, which operates in a two-dimensional space, to a four-dimensional space. The bicomplex number consists of two components, z_1 and z_2, where the real and imaginary parts can independently encode different aspects of the entities and their interactions in the knowledge graph.

Bicomplex numbers form a commutative ring under standard addition and multiplication operations, preserving essential algebraic properties such as associativity, commutativity, and distributivity:

- Additive Group: Bicomplex numbers form an Abelian group, meaning the sum of any two bicomplex numbers remains within the set, and every bicomplex number has an additive inverse.
- Multiplicative Group: Bicomplex multiplication is both associative and commutative.
- Distributive Property: Multiplication distributes over addition, ensuring consistency in algebraic operations.

These algebraic properties make bicomplex numbers a robust and flexible tool for knowledge graph embeddings, particularly when capturing complex relational structures across diverse data types.

The four-dimensional structure of bicomplex numbers enables the modeling of more intricate relationships in KGs. Bicomplex numbers can undergo various transformations that preserve structural properties, such as similarity and rotation, which are essential for knowledge graph embedding and entity alignment.

Two bicomplex numbers Z_1 and Z_2 are similar if there exists a nonzero bicomplex number U such that:

$$Z_2 = UZ_1U^{-1} \tag{3}$$

This similarity transformation ensures that Z_1 and Z_2 share common structural and spectral properties, which remain invariant under conjugation by U. This property is useful for aligning entities across different representation spaces.

Another critical transformation is the rotation of bicomplex numbers, defined as:

$$f(Z) = QZQ^{-1} \tag{4}$$

where Q is a bicomplex unit. These transformations, particularly similarity and rotation, provide powerful mechanisms for modeling and manipulating entity relationships in KGs, especially when dealing with multimodal data.

Bicomplex numbers support a variety of algebraic operations, which extend from complex numbers but are adapted to the four-dimensional space. These operations are as follows:

- For two bicomplex numbers $Z_1 = z_1 + jz_2$ and $Z_2 = z_3 + jz_4$, the sum is:

$$Z_1 + Z_2 = (z_1 + z_3) + j(z_2 + z_4) \tag{5}$$

- The difference between two bicomplex numbers is:

$$Z_1 - Z_2 = (z_1 - z_3) + j(z_2 - z_4) \tag{6}$$

- The product of two bicomplex numbers $Z_1 = z_1 + jz_2$ and $Z_2 = z_3 + jz_4$ is given by:

$$Z_1 Z_2 = (z_1z_3 - z_2 z_4) + j(z_1z_4 + z_2z_3) \tag{7}$$

The conjugate of a bicomplex number is defined as:

$$Z^* = z_1 - jz_2 \tag{8}$$

which negates the j-component while preserving the i-component. This conjugate operation is important for defining the norm and inverse of bicomplex numbers.

The norm of a bicomplex number Z is given by::

$$\|Z\| = ZZ^* = z_1^2 - z_2^2 \tag{9}$$

This norm is useful for ensuring stable embeddings and maintaining consistency in the structural representation of entities in the bicomplex space. The inverse of a bicomplex number Z is computed as:

$$Z^{-1} = \frac{Z^*}{\|Z\|} \tag{10}$$

where the norm $\|Z\| \neq 0$.

Overall, by embedding entities and relations in a bicomplex vector space, we gain several key advantages for MKGC. The four-dimensional representation captures intricate dependencies in KGs, enabling more robust modeling of symmetric, asymmetric, and cyclic relations. Additionally, bicomplex embeddings facilitate a structured integration of textual and visual data, preserving both semantic and geometric properties. This characteristic makes them particularly effective for tasks involving multimodal fusion, where traditional Euclidean embeddings often struggle to align heterogeneous data sources. Furthermore, the algebraic properties of bicomplex numbers enable more robust analogy reasoning, which is essential for predicting missing links in MKGs. Thus, the use of bicomplex embeddings provides a mathematically grounded approach to overcoming the limitations of conventional embedding methods, making them a powerful tool for advancing multimodal knowledge representation. By extending embeddings into a four-dimensional space, bicomplex numbers offer a new perspective on how multimodal information can be encoded and leveraged for more accurate and interpretable reasoning in KGs.

3.3. Riemannian manifold

In mathematical analysis and geometry, the concept of a manifold provides a generalized framework for studying spaces that locally resemble Euclidean space but may have more complex global structures. Intuitively, a manifold is a space in which, for every point, there exists a neighborhood that can be continuously mapped to an open subset of $\mathbb{R}^n$. This property allows manifolds to be analyzed using familiar mathematical tools from differential geometry while extending these techniques to non-Euclidean spaces.

A smooth manifold, also known as a C^∞ manifold, is a topological manifold equipped with a smooth structure. This means that it has an atlas, which is a collection of local coordinate charts, where the transition functions between overlapping charts are infinitely differentiable (smooth functions). The differentiability of these transitions enables the study of fundamental geometric properties, such as curvature, geodesics, and differentiable mappings between manifolds. Without a smooth structure, differentiation and related calculus operations would not be well-defined on the manifold.

A smooth manifold M is always defined together with a maximal atlas, which consists of all possible coordinate charts that are smoothly compatible. This maximal atlas is also referred to as the differentiable structure on M, as it uniquely determines how differentiation can be performed across the manifold. A differentiable structure allows for operations such as computing gradients, analyzing vector fields, and integrating over the manifold, which are essential in differential geometry and physics.

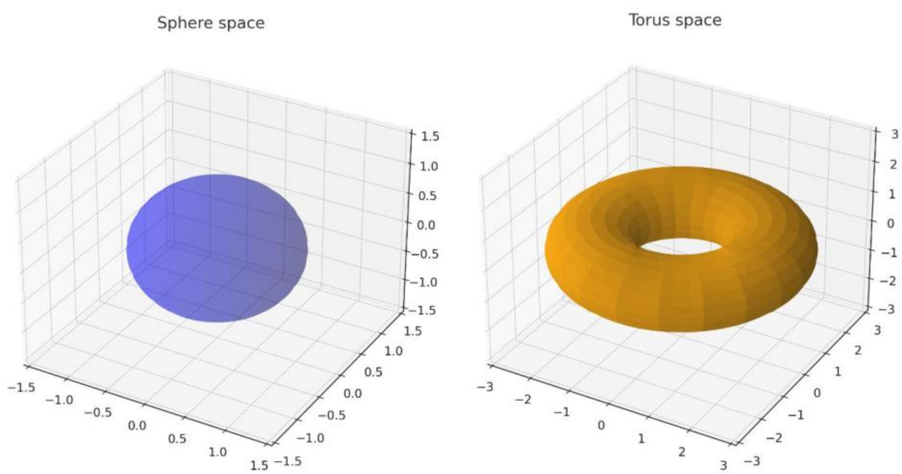

Fig. 3. Illustration of two smooth manifolds commonly used in geometric learning: the sphere and the torus. These spaces admit local coordinate charts with smooth transitions, making them suitable for modeling continuous latent structures. In knowledge graph embedding, such manifolds can be leveraged to represent entities and relations with inherent geometric or topological constraints.

A manifold is said to have dimension n if all of its connected components have dimension n, meaning that locally, it behaves like $\mathbb{R}^n$. In different dimensions, manifolds are classified according to their structure and properties. A 1-dimensional manifold is commonly referred to as a curve, representing objects such as lines or circles. A 2-dimensional manifold is called a surface, which includes familiar examples like spheres and tori (as illustrated in Fig. 3). More generally, an n-dimensional manifold is referred to as an n-manifold, with higher-dimensional manifolds playing a critical role in fields such as algebraic topology, relativity theory, and machine learning.

A maximal smooth atlas ensures that the collection of charts used to describe the manifold is as comprehensive as possible, meaning that it cannot be further extended without losing smooth compatibility. This guarantees that all transition functions between different coordinate charts remain smooth, preserving the differentiable structure of the manifold. The existence of a differentiable structure is essential for many applications, as it allows manifolds to be studied using the powerful tools of differential geometry, making them fundamental objects in modern mathematics and physics (Tu, 2011).

One of the essential structures associated with a smooth manifold is the tangent space. The tangent space at a point p, denoted as $T_p(M)$, consists of all possible tangent vectors to smooth curves passing through p. These tangent vectors represent infinitesimal displacements on the manifold and allow us to describe local geometric behaviors such as velocity, acceleration, and directional derivatives. If a smooth curve $\gamma: \mathbb{R} \to M$ passes through p, its tangent vector at p is defined by:

$$X_p = \left.\frac{d}{dt}\right|_{t=0} \gamma(t) \tag{11}$$

The dimension of the tangent space is equal to the dimension of the manifold itself, meaning that an n-dimensional manifold has an n-dimensional tangent space at each point. The tangent space forms the foundation for defining Riemannian metrics, which introduce inner product structures on manifolds.

A Riemannian manifold (J. M. Lee, 1997) is a smooth manifold that is endowed with a Riemannian metric, which defines a smoothly varying inner product on each tangent space at every point of the manifold. This structure allows for the measurement of geometric properties such as distances, angles, and curvature in a mathematically rigorous way.

Formally, a Riemannian metric on a smooth manifold M is a (0,2)-tensor field $g \in T^2(M)$ that satisfies two key properties. First, it is symmetric, meaning that for any two tangent vectors X,Y at a point $p \in M$, the metric satisfies:

$$g(X,Y) = g(Y,X) \tag{12}$$

This symmetry ensures that the inner product remains invariant under the exchange of vectors, which is a fundamental requirement for defining geometric structures. Second, it is positive definite, meaning that for any nonzero tangent vector X, it satisfies:

$$g(X,X) > 0, \forall X \neq 0 \tag{13}$$

This condition guarantees that the inner product of a vector with itself is always positive, ensuring a well-defined notion of length and distance on the manifold. The inner product associated with a Riemannian metric is typically written as:

$$\langle X,Y \rangle := g(X,Y), \forall X,Y \in T_p(M) \tag{14}$$

A smooth manifold equipped with a Riemannian metric is called a Riemannian manifold. The presence of a Riemannian metric enables the study of fundamental geometric concepts, including curvature, geodesics, and geodesic distance, which generalizes the notion of shortest paths in Euclidean space to curved manifolds. This structure forms the mathematical foundation of Riemannian geometry, which extends classical Euclidean geometry to more complex, non-Euclidean spaces.

In applications such as MKGs, Riemannian manifolds provide a principled framework for embedding entities and relations into curved spaces, improving the ability to model hierarchical and multimodal interactions. By leveraging the geometric properties of Riemannian manifolds, knowledge graph representations can better capture the underlying structure of real-world relationships.

To fully exploit these geometric properties, it is essential to understand how a Riemannian metric defines fundamental measurements such as distances, angles, and curvature on a manifold. The Riemannian metric provides a smoothly varying inner product on each tangent space, allowing for precise computations of vector relationships. Given a tangent vector X, its norm (length) is determined by:

$$\|X\| = \sqrt{g(X,X)} \tag{15}$$

Similarly, the angle between two tangent vectors X and Y at a point p is given by:

$$cos(\theta) = \frac{g(X,Y)}{|X||Y|} \tag{16}$$

These mathematical formulations enable the generalization of Euclidean geometric concepts to curved spaces, allowing KGEs to encode richer structural dependencies. The Riemannian metric naturally includes the standard Euclidean metric on $\mathbb{R}^n$, where the inner product corresponds to the familiar dot product. Moreover, it extends to submanifolds of Euclidean space, such as spheres, where the induced metric is inherited from the surrounding ambient space. These generalizations provide a powerful tool for modeling the latent geometry of MKGs, where relationships between entities may exhibit non-Euclidean characteristics such as hierarchical depth, semantic proximity, or multimodal dependency.

Moreover, in Riemannian geometry, geodesics play a fundamental role in defining distances and paths of minimal length on curved spaces. The geodesic distance between two points p and q on a Riemannian manifold M is formally defined as the infimum (i.e., the shortest possible value) of the lengths of all admissible smooth curves that connect these points. Mathematically, this is written as:

$$d(p,q) \ = inf_{\gamma}\, L(\gamma) \tag{17}$$

where γ is a smooth admissible curve connecting p and q, and $L(\gamma)$ denotes the length of the curve γ. The length of a smooth curve γ: $[a,b] \rightarrow M$ is computed using the Riemannian metric:

$$L(\gamma) = \int_a^b \|\gamma'(t)\|\, dt = \int_a^b \sqrt{g(\gamma'(t), \gamma'(t))} dt \tag{18}$$

One of the key properties of geodesics is that they satisfy the geodesic equation, which ensures that the acceleration of the curve remains zero when measured with respect to the manifold's intrinsic connection. A geodesic $\gamma(t)$ satisfies the Euler-Lagrange equation, which, when expressed in local coordinates ($x^1, x^2, \dots, x^n$), takes the form:

$$\frac{d^2x^k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k \frac{dx^i}{dt}\frac{dx^j}{dt} = 0,\ \ k = 1, \dots, n \tag{19}$$

where, $\Gamma^k{}_{ij}$ are the Christoffel symbols, which encode the curvature and determine how geodesics deviate from straight-line motion. In Euclidean space, these terms vanish, reducing the geodesic equation to the familiar equation for straight-line motion:

$$\frac{d^2x^k}{dt^2} = 0 \tag{20}$$

However, on curved manifolds such as spheres, hyperbolic spaces, and knowledge graph embeddings in Riemannian spaces, geodesics exhibit nontrivial trajectories. For example, on a sphere S^2, geodesics correspond to great circles, which provide the shortest path between any two points on the surface. In MKGs, geodesic distances are crucial for defining similarity measures between embedded entities. When embedding KGs into Riemannian manifolds, geodesic distances naturally reflect hierarchical and relational structures. For instance, hyperbolic embeddings exploit geodesics to model tree-like structures efficiently, where distances grow exponentially as entities move away from the origin. Moreover, geodesic computations play an essential role in optimization on manifolds. Unlike standard Euclidean gradient descent, which updates parameters in a straight-line direction, Riemannian gradient descent updates parameters along geodesics, ensuring that the updates remain consistent with the underlying geometry. This is particularly useful in machine learning models where parameter spaces exhibit non-Euclidean curvature.

4. Proposed model

4.1. Overall architecture of BiCore

MKGC presents significant challenges due to the inherent complexity of integrating diverse data modalities, such as text, images, and structural information. Traditional knowledge graph embedding models primarily operate in Euclidean spaces, which are often insufficient for capturing hierarchical relationships and multimodal dependencies. On the other hand, hyperbolic embeddings provide a more structured representation for hierarchical data but struggle with integrating heterogeneous information sources in a unified manner. Additionally, tensor decomposition approaches, while effective in structured KGC, often overlook the intricate interactions between modalities, leading to suboptimal reasoning performance. To address these limitations, we introduce BiCore, a novel multimodal embedding framework designed to enhance relational reasoning and knowledge completion in MKGC. BiCore is built upon three core innovations: bicomplex embeddings, Riemannian manifold-based fusion, and Transformer-based reasoning mechanisms. Unlike traditional embedding approaches, BiCore extends the representational capacity of complex-valued embeddings by incorporating a bicomplex algebraic structure, enabling richer modeling of symmetric, asymmetric, and cyclic relationships. By leveraging Riemannian manifold geometry, BiCore effectively aligns multimodal embeddings, ensuring that textual and visual information are integrated in a way that preserves semantic consistency and geometric coherence. Fig. 4 illustrates the overall architecture of BiCore.

The architecture of BiCore consists of three primary stages: embedding extraction, fusion via Riemannian manifolds, and bicomplex scoring for relational reasoning. First, multimodal entities and relations from MKG are represented as distinct embedding vectors for textual, visual, and structural information. These embeddings are then projected onto a Riemannian manifold, where they undergo geometric transformations to align different modalities into a shared space. A transformer-based fusion module refines these embeddings by capturing long-range dependencies and higher-order interactions between multimodal features. Finally, the unified representations are transformed into bicomplex embeddings, which provide an expressive mathematical framework for computing relational scores via a dedicated scoring function. For instance, in an MKG where an entity like *Eiffel Tower* is represented through textual descriptions and images, conventional embedding models may fail to establish meaningful connections across modalities. While a text-based embedding may capture historical and descriptive attributes, and a structural embedding may encode its connections in the knowledge graph, existing models struggle to integrate visual semantics effectively. BiCore resolves this by dynamically fusing multimodal inputs into a coherent representation, allowing the model to make accurate relational predictions, such as *Eiffel Tower* → *located_in* → *Paris*.

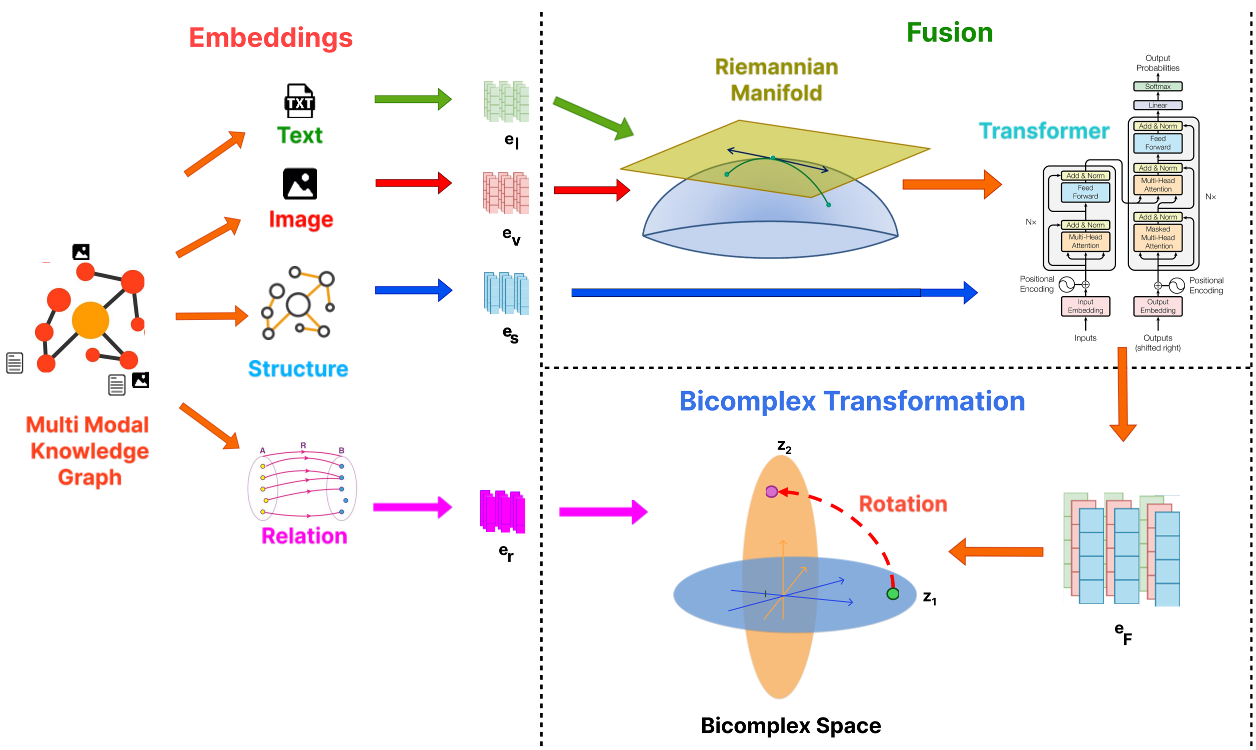

Fig. 4. Overview of the BiCore architecture. The model begins by extracting modality-specific embeddings from text, image, and structural components of a multimodal knowledge graph. These embeddings are projected and fused on a Riemannian manifold to preserve geometric consistency. A Transformer-based fusion module then integrates the information across modalities. The resulting unified representations are encoded as Bicomplex embeddings, which are subsequently processed by a scoring function to perform link prediction.

4.2. Fusion via Riemannian manifolds

Integrating multimodal data into a unified representation space presents significant challenges due to the heterogeneous nature of different modalities (e.g., text and images). While traditional Euclidean spaces provide a straightforward embedding approach, they fail to capture the intrinsic geometric structures present in multimodal relationships. To overcome this limitation, BiCore employs Riemannian manifold-based fusion, which preserves the non-Euclidean properties of multimodal embeddings while facilitating robust alignment between different data sources. In MKGs, entities and relations often exist in non-Euclidean spaces, where hierarchical structures and complex interactions make simple Euclidean embeddings ineffective. Riemannian manifolds provide a geometric framework that adapts to the curvature of data distributions, allowing embeddings to be mapped onto curved spaces where distances more accurately reflect semantic and structural similarities. The key challenge in multimodal fusion is ensuring that embeddings from different modalities (text, image, and structure) are properly aligned. This alignment is particularly difficult due to the fundamental differences in how textual and visual information are structured. BiCore addresses this challenge by employing Fisher-Rao Distance as a measure of similarity in the Riemannian manifold, ensuring that the geometric distances between embeddings reflect their true multimodal relationships.

The Fisher-Rao metric is induced by the Fisher Information Matrix (FIM) (Miyamoto et al., 2024), which defines a local Riemannian structure on a statistical manifold. This metric captures the intrinsic geometry of probability distributions, allowing us to measure distances between multimodal embeddings in a way that respects their statistical properties. Since the Fisher-Rao metric defines a Riemannian manifold, the Fisher-Rao distance is computed as the geodesic distance between two points on this manifold. In our model, this geodesic distance is essential for aligning textual and visual representations, ensuring that their similarities are measured in a curvature-aware manner.

A practical approach to computing the Fisher-Rao distance is through the square-root transformation, which embeds probability densities into a Hilbert space L_2 with the standard inner product. This transformation allows us to represent probability distributions in a more stable and computationally efficient manner. Given two functions $f, g \in L_2$, their inner product is defined as:

$$\langle f, g \rangle = \int f(x) g(x) dx \tag{21}$$

By applying the square-root transformation, probability densities are mapped onto the infinite-dimensional unit sphere S^∞. If $f(x)$ represents a transformed probability density function, then:

$$\|f\|^2 = \langle f, f \rangle = \int f(x)^2 dx = 1 \tag{22}$$

Since these transformed functions lie on the unit sphere, the geodesic curve between two points is given by the great circle connecting them. Let ψ_1 and ψ_2 be two elements on S^∞, then their geodesic distance is computed as:

$$d(\psi_1, \psi_2) = \arccos(\langle \psi_1, \psi_2 \rangle) \tag{23}$$

This approach ensures that multimodal embeddings remain normalized within a structured space, simplifying distance computations while maintaining a stable representation for geometric similarity measurements.

When embeddings are mapped onto the unit sphere in Hilbert space, the Fisher-Rao distance between two normalized vectors v_1 and v_2 is computed as the geodesic distance on this sphere. Since the unit sphere is itself a Riemannian manifold, this geodesic distance naturally respects the curvature of the space. The square-root transformation used in our model ensures that multimodal embeddings (e.g., text and image representations) remain on this manifold, thereby preserving statistical relationships between distributions.

The Fisher-Rao distance between two points on the unit sphere is given by:

$$d_{Fisher} = \arccos(\langle v_1, v_2 \rangle) \tag{24}$$

where $\langle \cdot, \cdot \rangle$ is the dot product of the two normalized embeddings, and $\arccos(\cdot)$ maps the cosine similarity to an angular distance in the range $[0, \pi]$. This formulation ensures that distances between multimodal embeddings correctly reflect their geometric relationships, leading to more accurate link predictions and reasoning in MKGC.

In practical terms, consider a MKG $\mathcal{G}$, where each entity e is associated with three types of embeddings: a structural embedding (e_s), a linguistic embedding (e_l), and a visual embedding (e_v). Traditional unimodal KGE methods focus solely on structural embeddings (e_s), whereas MKGE incorporates multiple modalities to enhance reasoning capabilities.

First, to ensure that multimodal embeddings reside in a common geometric space, both the linguistic embedding e_l and the visual embedding e_v are normalized onto the unit sphere.

$$
\begin{aligned}
e_{l_{norm}} &= \text{normalize}(e_l) \\
e_{v_{norm}} &= \text{normalize}(e_v)
\end{aligned}
\tag{25}
$$

Next, we compute the Fisher-Rao distance between the normalized visual and linguistic embeddings, ensuring they are properly aligned within the Riemannian manifold:

$$d_{Fisher} = \arccos\left(\langle e_{v_{norm}}, e_{l_{norm}} \rangle\right) \tag{26}$$

Using this distance, we construct a multimodal fusion embedding e_{lv} by applying an exponential interpolation mechanism:

$$e_{lv} = \exp(d_{Fisher}) \cdot e_l + (1 - \exp(d_{Fisher})) \cdot e_v \tag{27}$$

The Fisher-Rao distance acts as an adaptive weight, dynamically adjusting the influence of each modality. If e_l and e_v are highly similar, their fusion is nearly balanced, whereas if they differ significantly, the model assigns a greater weight to the more informative modality.

After obtaining e_{lv}, we introduce a curvature-based transformation to further refine the embedding, ensuring that multimodal representations are geometrically aligned. The curvature adjustment is based on the Euclidean distance between e_l and e_v:

$$curve = \text{normalize}(e_v - e_l) \tag{28}$$

This curvature factor is then used to adjust the fused embedding:

$$e_{clv} = \sigma(curve) \cdot e_{lv} + (1 - \sigma(curve)) \cdot e_v \tag{29}$$

where σ is a sigmoid function, ensuring that the weighting remains within a stable range. This adjustment fine-tunes the fusion process, preventing significant distortions when combining heterogeneous modalities.

The fused multimodal embedding e_{lv} is then integrated with the structural embedding e_s using a strategy hyperparameter α, which controls the balance between multimodal and structural features:

$$e_F = \alpha \cdot e_s + (1 - \alpha) \cdot e_{clv} \tag{30}$$

To extract higher-level representations, the final multimodal entity embedding undergoes transformer-based refinement:

$$e_{TF} = \text{Transformer}(e_F) \tag{31}$$

This ensures that the model effectively captures long-range dependencies and complex interactions within MKGs.

4.3. Projection of unified embeddings into bicomplex space

Once the unified multimodal embedding e_{TF} is obtained through fusion of textual, visual, and structural information, it serves as the foundation for entity representation in bicomplex space. Since knowledge graph reasoning requires modeling complex relational structures, we map e_{TF} into the bicomplex number space $\mathbb{BC}$, which allows for more expressive transformations, including rotations and reflections. Each entity embedding, whether it acts as a head entity (e_h) or a tail entity (e_t) in a knowledge graph triple (e_h , r, e_t), is represented as a bicomplex number with two components:

$$
\begin{aligned}
e_h &= z_{h1} + j z_{h2} \\
e_t &= z_{t1} + j z_{t2}
\end{aligned}
\tag{32}
$$

Similarly, each relation r in the knowledge graph is also embedded in the bicomplex space, allowing it to encode both semantic transformations and geometric relationships:

$$e_r = z_{r1} + jz_{r2} \tag{33}$$

The transition from e_{TF} to e_h and e_t ensures that the entity's multimodal features are preserved while adapting them to the relational space. Since an entity can function as both a head or a tail depending on its position in different triples, the unified embedding e_{TF} is decomposed into bicomplex components accordingly. If an entity appears as the head in a triple, its bicomplex embedding is derived as:

$$(z_{h1}, z_{h2}) = \text{split}(e_{hTF}) \tag{34}$$

Similarly, when the entity is used as the tail, its embedding follows the same transformation:

$$(z_{t1}, z_{t2}) = \text{split}(e_{tTF}) \tag{35}$$

The function $\text{split}(\cdot)$ decomposes the multimodal representation into two complementary components, enabling the entity to retain its multimodal information while allowing relational transformations. This ensures that embeddings remain consistent across different contexts while enabling more expressive reasoning in KGs.

The bicomplex representation provides an advantage over traditional Euclidean embeddings by offering a richer mathematical framework for encoding relational dependencies. Unlike real-valued embeddings, which may struggle with encoding non-commutative properties, bicomplex embeddings effectively model hierarchical relations and asymmetric dependencies. Furthermore, the separation of real and imaginary components allows for distinct but complementary representations of multimodal information, preserving both linguistic and visual features while enabling structured transformations in the relational space.

4.4. Relational transformation in the multimodal bicomplex space

Once the head entity embedding e_h and relation embedding e_r are established, BiCore models entity transformations through bicomplex multiplication:

$$e_{hr} = e_h \times e_r = (z_{h1} + jz_{h2}) \times (z_{r1} + jz_{r2}) = z_{hr1} + jz_{hr2} \tag{36}$$

This transformation effectively encodes rotation, scaling, and reflection properties, ensuring that the model captures asymmetric relationships commonly observed in real-world KGs.

After obtaining the relation-dependent entity embedding e_{hr} via bicomplex multiplication, the next step is to evaluate the plausibility of a given triple (e_h, e_r, e_t). This is achieved through a scoring function, which measures how well the transformed head entity e_{hr} aligns with the tail entity e_t in the bicomplex space. The scoring function in BiCore is based on the bicomplex dot product, which ensures that both the real and imaginary components of the embeddings are aligned in a way that captures semantic and geometric relationships. The score for a given triple is computed as:

$$f(h, r, t) = e_{hr} \cdot e_t = (z_{hr1} + jz_{hr2}) \cdot (z_{t1} + jz_{t2}) = \langle z_{hr1}, z_{t1} \rangle + \langle z_{hr2}, z_{t2} \rangle \tag{37}$$

This scoring function allows the model to compare the transformed head entity e_{hr} with the tail entity e_t and determine whether the relation e_r is valid for the given pair. The real part of the score captures semantic similarity, while the imaginary part accounts for complex transformations such as rotation and asymmetry.

4.5. Learning process

To optimize the parameters of BiCore, we minimize the following objective function:

$$\mathcal{L} = \sum_{(h,r,t) \in \mathcal{T} \cup \mathcal{T}^-} \log(1 + \exp(y \cdot f(h, r, t))) + \lambda \left(\|e_h^i\|_3^3 + \|e_r^i\|_3^3 + \|e_t^i\|_3^3 \right) \tag{38}$$

where $y \in \{-1; 1\}$ denotes the label of the triple (h, r, t). $\mathcal{T}$ denotes the set of observed triples then let $\mathcal{T}^-$ be the set of unobserved triples, while $\|\cdot\|_3^3$ represents the N3 norm used for regularization, with λ as the regularization coefficient. Since the multimodal embeddings are pre-trained, only the structural embeddings are updated during training.

Additionally, we propose the use of the N3F regularization, which combines the N3 regularizer with the Frobenius norm. The N3 regularizer penalizes higher-order norms (i.e., the cube of the absolute value of the embeddings), while the Frobenius norm ensures that the embeddings retain a certain level of structure in the model. Furthermore, Wasserstein regularization leverages the Sinkhorn distance, a form of optimal transport, to smooth the embeddings. This regularization technique enforces a smoother embedding space by considering the distance between embeddings from different modalities. As a result, it improves the fusion stability when the model processes multimodal data, ensuring that textual and visual data are effectively integrated. By combining these regularization techniques, the model achieves enhanced stability and smoother fusion between modalities, leading to better generalization and improved performance in tasks such as link prediction and analogy reasoning. We conduct experiments to evaluate the effectiveness of this approach, demonstrating its superior performance in MKGC tasks compared to standard regularization techniques.

The training process follows a mini-batch stochastic gradient descent approach, where entity and relation embeddings are initialized randomly from a uniform distribution. At each iteration, a mini-batch S_{batch} is sampled from the training set S. For each valid triple S_{batch}, a corrupted triple is generated by randomly substituting either the head or the tail entity:

$$S'(h, r, t) = \{(h', r, t) | h' \in \mathcal{E}\} \cup \{(h, r, t') | t' \in \mathcal{E}\} \tag{39}$$

This corruption strategy forces the model to differentiate valid triples from invalid ones. The set of valid triples S_{batch} and their corrupted counterparts are combined into a training batch T_{batch}, from which the embeddings are updated via the Adagrad optimizer. Details of the optimization procedure are presented in Algorithm 1.

Algorithm 1. Learning algorithm of the BiCore model

Input: Training set $S = (h, r, t)$, entities set $\mathcal{E}$, relations set $\mathcal{R}$, and hyperparameters

Output: Entity and relation embeddings

Initialize linguistic embeddings e_l, visual embeddings e_v, and structural embeddings e_s for entities, and relation embeddings e_r.
repeat
  $S_{batch} \leftarrow sample(S, b)$ // sample a mini-batch of size b from S
  $T_{batch} \leftarrow \emptyset$
  for $(h, r, t) \in S_{batch}$ do
    $(h', r, t') \leftarrow sample(S')$ //negative sampling by corrupting the triple
    $T_{batch} \leftarrow T_{batch} \cup \{(h, r, t), (h', r, t')\}$
  end for
  for each entity in T_{batch}:
    $e_{l_{norm}} = \frac{e_l}{\|e_l\|}$, $e_{v_{norm}} = \frac{e_v}{\|e_v\|}$ //normalize embeddings
    $d_{Fisher} = \arccos(\langle e_{v_{norm}}, e_{l_{norm}} \rangle)$ //compute Fisher-Rao distance
    $e_{lv} = \exp(d_{Fisher}) \cdot e_l + (1 - \exp(d_{Fisher})) \cdot e_v$ //fuse visual and linguistic embeddings
    $curve = \frac{e_v - e_l}{\|e_v - e_l\|}$
    $e_{clv} = \sigma(curve) \cdot e_{lv} + (1 - \sigma(curve)) \cdot e_v$ //compute curvature-aware fusion
    $e_F = \alpha \cdot e_s + (1 - \alpha) \cdot e_{clv}$ //combine with structural embeddings
    $e_{TF} = \text{Transformers}(e_F)$
    $e_{h|t} = \text{split}(e_{TF})$
  end for
  Update embeddings by minimizing the loss function w.r.t.
    $\nabla \left[\sum_{(h, r, t) \in T_{batch}} \log(1 + \exp(y \cdot f(h, r, t))) + \lambda \left(\|e_h^i\|_3^3 + \|e_r^i\|_3^3 + \|e_t^i\|_3^3 \right) \right]$
until convergence or reaching maximum training time.

5. Experiments

In this section, we outline the methodological process used to evaluate the proposed BiCore model. We begin by introducing the datasets selected for evaluation and provide justification for their use. Next, we describe the baseline models chosen for comparison, ensuring that our approach is assessed against state-of-the-art methodologies. Following this, we define the evaluation metrics and experimental settings to ensure a rigorous and reproducible experimental framework. Finally, we present and analyze the results, with a particular focus on model convergence, generalization, and performance gains.

5.1. Datasets

To assess the effectiveness of BiCore in multimodal analogical reasoning and link prediction, we conduct experiments on multiple benchmark datasets. Our evaluation spans two primary tasks: multimodal analogical reasoning and multimodal knowledge graph completion.

BiCore is designed to enhance multimodal analogical reasoning by capturing intricate relationships across textual and visual modalities. To evaluate its performance in this task, we utilize the MarKG and MARS datasets.

- MarKG (N. Zhang et al., 2023) is a background multimodal knowledge graph consisting of 34,420 triplets extracted from Wikidata. This dataset provides prior knowledge of analogous entities and relations, forming the foundation for training and evaluating our model.
- MARS (N. Zhang et al., 2023) is a specialized dataset designed explicitly for multimodal analogical reasoning. It contains 10,685 training questions, 1,228 validation questions, and 1,415 test questions, enabling a comprehensive assessment of BiCore's ability to infer missing analogies across different modalities.

Next, to evaluate BiCore in the link prediction task, we conduct experiments on three widely used MKGC datasets: DB15K, MKG-W, and MKG-Y. These datasets incorporate textual and visual information to enrich the representation of entities and relations.

- DB15K (Liu et al., 2019): This dataset is a cross-lingual knowledge graph alignment dataset constructed from DBpedia. It consists of four language-specific KGs: English (En), Chinese (Zh), French (Fr), and Japanese (Ja), with each language containing between 65,000 and 106,000 entities. The dataset is primarily used for entity alignment tasks, where models aim to match identical entities across different languages. DB15K also facilitates cross-lingual reasoning and knowledge graph completion across multiple linguistic contexts. The evaluation data for DB15K are derived from official DBpedia releases.
- MKG-W (Xu et al., 2022): This dataset is a MKG specifically designed to support MKGC tasks. It integrates both textual and visual knowledge, including images associated with entities and relations. MKG-W enables the development and evaluation of models that perform reasoning over both text and images, making it ideal for link prediction and multimodal knowledge representation. The dataset is constructed from official Wikidata releases and includes images sourced from publicly available databases.
- MKG-Y (Xu et al., 2022): This dataset is derived from YAGO3 (Mahdisoltani et al., 2015), another large-scale MKG. Similar to MKG-W, MKG-Y incorporates textual and visual entity representations to facilitate multimodal reasoning and KGC. However, MKG-Y provides a more heterogeneous and diverse dataset, offering a broader spectrum of entities and relationships sourced from multiple data providers. The dataset is used to benchmark models on their ability to integrate and reason over multimodal information, improving missing link prediction in KGs.

5.2. Evaluation protocol

To evaluate the performance of BiCore in multimodal analogical reasoning and link prediction, we formulate the task as a link prediction problem, where the model directly predicts the answer entity. To assess the quality of these predictions, we employ two widely used ranking-based evaluation metrics: Mean Reciprocal Rank (MRR) and Hit at k (Hits@k, H@k).

MRR is a widely used metric for evaluating ranked retrieval systems, particularly in tasks where the goal is to identify the first correct prediction in an ordered list of results. Given a set of queries, MRR computes the average reciprocal rank of the first correctly predicted entity. Mathematically, it is defined as:

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i} \quad (40)$$

where $|Q|$ is the total number of queries and $rank_i$ represents the position of the first correct entity in the ranked list for query i. A higher MRR value indicates that the model is better at ranking the correct entity near the top of the predicted list, leading to more accurate retrieval.

H@k metric measures the percentage of test queries for which the correct entity appears within the top-k ranked predictions. It provides an intuitive measure of how often the model successfully ranks the correct answer among the top candidates. Formally, H@k is computed as:

$$H@k = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\mathbb{I}(rank_i \leq k) \tag{41}$$

where $\mathbb{I}(\cdot)$ is an indicator function that returns 1 if the correct entity appears within the top-k predictions and 0 otherwise. Commonly used values for k include 1, 3, and 10, providing insights into both strict and relaxed ranking performance.

5.3. Baselines

In this study, we comprehensively evaluate BiCore against a diverse set of baseline models across three datasets: DB15K, MKG-W, and MKG-Y. These baselines span multiple methodological categories, including traditional KGE methods, deep learning-based models, MKGE approaches, and hybrid KGC models, ensuring a fair and meaningful comparison.

For the link prediction task, we compare BiCore with classical KGE models, including TransE, DistMult, ComplEx, RotatE, and TuckER. These foundational methods employ translation-based operations, bilinear interactions, or tensor decomposition to model entity relationships. Since BiCore introduces bicomplex embeddings and curvature-weighted fusion, comparing it against these traditional models allows us to assess the advantages of its advanced geometric representations. Additionally, we incorporate deep learning-based KGE models such as TBKGC (Mousselly-Sergieh et al., 2018), MACO (Y. Zhang et al., 2023b), QEB (X. Wang et al., 2023), VISTA (J. Lee et al., 2023), AdaMF (Y. Zhang et al., 2024), and MANS (Y. Zhang et al., 2023a), which leverage GNNs and Transformer architectures to enhance relational reasoning. Given that BiCore integrates Transformer-based architectures, this comparison highlights its ability to generalize across large-scale KGs while maintaining computational efficiency.

Since BiCore is designed for MKGC, we further assess its performance against leading MKGE models, including IKRL (Xie et al., 2017), TransAE (Z. Wang et al., 2019), MMKR (Zheng et al., 2023), RSME (M. Wang et al., 2021), VBKGC (Y. Zhang & Zhang, 2022), and OTKGE (Cao et al., 2022). These methods integrate multimodal data—such as images and textual descriptions—to enrich representation learning. However, many of them rely on static fusion techniques, limiting their ability to dynamically capture cross-modal dependencies. By incorporating bicomplex embeddings, Riemannian fusion, and adaptive curvature mechanisms, BiCore aims to address these limitations, offering a more flexible and expressive approach to multimodal reasoning.

To further establish strong baselines, we also consider advanced relational reasoning models such as IMF (X. Li et al., 2023) and MMRNS (Xu et al., 2022), which incorporate metric learning, multi-hop reasoning, and Riemannian embeddings to improve link prediction accuracy. Since BiCore employs geometric representations and Fisher-Rao distance, evaluating its performance against these models helps determine its ability to capture complex, non-linear relationships in KGs.

5.4. Experiment settings and hyperparameters

BiCore model is implemented using PyTorch 2.4.1 + cu117, running on an Ubuntu 18.04.6 system. The experiments are conducted on an Nvidia Tesla A100 GPU with 40GB VRAM, ensuring sufficient computational power for large-scale MKG tasks. To optimize hyperparameters, we employ a grid search strategy, selecting from a range of configurations to identify the best-performing model. The key hyperparameters explored include batch size, learning rate, regularization techniques, and optimization methods. The grid search space is designed to cover a diverse range of values, ensuring robustness in model training and evaluation. The detailed hyperparameter settings are summarized in Table 2.

Table 2
Hyperparameter settings and model configurations used in BiCore during training and evaluation.

Hyperparameter	Search Space/Value	Type
Rank (embedding size)	500	Fixed
Regularization weight	1e-2	Fixed
Regularizer	{N3, N3FW, L2}	Choice
Batch size	{1000, 2000, 4000, 5000, 10000}	Choice
Learning rate (α)	{1e-1,5e-3, 5e-4}	Choice
Optimizer	Adam, Adagard, AdamW	Choice
Decay rate β_1 (Adam)	0.9	Fixed
Decay rate β_2 (Adam)	0.98	Fixed
Training epochs	{300, 500, 1000}	Choice

5.5. Results and discussion

5.5.1. Prediction performance

To evaluate the effectiveness of BiCore in MKGC, we compare its link prediction performance against baseline models across benchmark datasets. The results are reported in Table 3 with the best-performing results for each dataset are highlighted in bold, while the second-best scores are underlined. On DB15K, BiCore demonstrates better performance over competing approaches, achieving an MRR of 0.399, which marks a 17.92% improvement over the second-best model, TuckER (0.338). Furthermore, BiCore attains a H@1 of 0.323, surpassing TuckER's 0.253 by a significant 27.5% margin, highlighting its strength in ranking the correct entity at the top position. These results suggest that BiCore's BiComplex embeddings provide a more expressive space for modeling relational dependencies, outperforming traditional embedding techniques such as TransE, ComplEx, and RotatE, which struggle with more complex relation types. The improvements in H@3 (0.435) and H@10 (0.546) further demonstrate that BiCore effectively balances precision and recall, ensuring robust link prediction performance across different ranking thresholds.

Table 3
Link prediction results on DB15K, MKG-W, and MKG-Y datasets. The best scores are highlighted in bold, and the second-best scores are underlined.

Model	DB15K				MKG-W				MKG-Y			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE	.248	.127	.314	.470	.291	.210	.332	.442	.307	.234	.351	.433
DistMult	.230	.147	.262	.395	.209	.159	.222	.308	.250	.193	.278	.359
ComplEx	.274	.183	.315	.453	.249	.190	.266	.363	.287	.222	.321	.409
RotatE	.292	.178	.361	.496	.336	.268	.366	.467	.349	.291	.383	.453
TuckER	<u>.338</u>	<u>.253</u>	.379	.503	.303	.244	.329	.412	.370	<u>.345</u>	.384	.414
IKRL	.268	.140	.349	.490	.323	.261	.347	.440	.332	.303	.342	.382
TBKGC	.284	.156	.370	.498	.314	.253	.339	.432	.339	.304	.352	.400

TransAE	.290	.212	.311	.411	.300	.212	.349	.447	.281	.253	.291	.330
MMKRL	.268	.138	.350	.493	.301	.221	.340	.446	.368	.316	.397	.453
RSME	.297	.241	.321	.402	.292	.233	.319	.404	.344	.317	.360	.390
VBKGC	.306	.197	.371	.494	.306	.249	.330	.408	.370	.337	.387	.423
OTKGE	.238	.184	.258	.342	.343	.288	.362	.448	.355	.319	.371	.413
MACO	.274	.146	.355	.500	.317	.252	.342	.443	.349	.315	.366	.405
IMF	.322	.242	.360	.481	.345	.287	.366	.454	.357	.329	.371	.406
QEB	.281	.148	.366	.515	.323	.254	.350	.453	.343	.294	.369	.423
VISTA	.304	.224	.335	.459	.329	.261	.353	.456	.304	.248	.323	.415
AdaMF	.325	.213	.396	.516	.342	.272	.378	.472	**.380**	.334	**.404**	**.454**
MANS	.288	.168	.365	.492	.308	.248	.336	.417	.290	.252	.313	.344
MMRNS	.326	.230	.378	.510	.350	.285	.374	.474	.359	.305	.390	.454
BiCore	**.399**	**.323**	**.435**	**.546**	**.369**	**.313**	**.393**	**.477**	.376	**.348**	.391	.427
Improvement (%)	17.92	27.5	9.73	5.65	5.42	8.56	3.99	0.49	-	0.69	-	-

On MKG-W, BiCore again achieves the highest MRR of 0.369, outperforming the second-best model, MMRNS, which records an MRR of 0.35, representing a 5.42% gain. The H@1 score of 0.313, which surpasses OTKGE's 0.288, further reflects BiCore's advantage in precisely ranking correct triples at the top of the prediction list, marking an 8.56% improvement. The gains in H@3 (0.393) and H@10 (0.477) emphasize BiCore's effectiveness in handling multimodal information, particularly due to its Riemannian fusion mechanism, which ensures that textual and visual embeddings are aligned in a geometrically structured space. Compared to MMKRL and VBKGC, which employ basic concatenation-based fusion strategies, BiCore's Fisher-Rao fusion dynamically adjusts the contributions of different modalities based on their geometric alignment, preventing modality dominance and ensuring a more balanced integration of multimodal information.

On MKG-Y, however, BiCore, while still among the top-performing models, does not achieve the highest scores in all ranking metrics. BiCore records a H@1 of 0.348, which is the best performance among all models, indicating its strength in ranking the correct triple at the first position. However, its MRR of 0.376 is slightly lower than that of AdaMF (0.38), and its H@10 of 0.427 falls behind MMRNS (0.454) and AdaMF (0.454). These results suggest that while BiCore excels in precision-based ranking (as evidenced by H@1 and H@3 scores), it does not fully optimize recall performance, leading to slightly lower H@10 scores compared to models like AdaMF and MMRNS. The reasons behind BiCore's slightly lower performance on MKG-Y can be attributed to several factors. First, MKG-Y, derived from YAGO3, has a more diverse and heterogeneous data structure, where entity relations span multiple domains with varying degrees of textual and visual alignment. While BiCore's Fisher-Rao fusion is effective in aligning multimodal information, it is designed to prioritize structural consistency over adaptive weighting, unlike AdaMF, which utilizes adaptive multimodal fusion strategies to dynamically adjust the weighting of text and images for each query. As a result, AdaMF is better suited for datasets like MKG-Y, where different entity types may require different fusion strategies. Second, models like TuckER and MMRNS, which outperform BiCore in H@10, use tensor-based decomposition and multi-relational learning mechanisms that may provide better high-dimensional factorization for complex entity relations in MKG-Y. In contrast, BiCore's BiComplex embeddings, while effective at modeling hierarchical and asymmetric relations, may not fully exploit the broader contextual dependencies that tensor-based models like TuckER capture more effectively.

Table 4
Event prediction results of BiCore and competitive multimodal KGE baselines on the MARS dataset.

Baselines	Backbone	H@1	H@3	H@10	MRR
IKRL	TransE	.254	.285	.304	.274
TransAE	TransE	.203	.233	.253	.223
RSME	ComplEx	.255	.274	.291	.268
IKRL	ANALOGY	.266	.294	.310	.283
TransAE	ANALOGY	.261	.285	.293	.276
RSME	ANALOGY	.266	.298	**.311**	.285
IKRL	BiCore	.273	.296	.302	.284
TransAE	BiCore	.268	.289	.295	.281
RSME	BiCore	**.280**	**.299**	.305	**.290**

Table 4 presents a comparative analysis of event prediction performance for BiCore, our proposed model, against other leading MKGE models on the MARS dataset. The results illustrate how different backbone architectures, including TransE, ComplEx, and ANALOGY, influence the performance of IKRL, TransAE, and RSME, providing insights into BiCore's effectiveness in event prediction tasks.

Each baseline model is evaluated using a distinct backbone architecture, which determines how entity and relation representations are learned. TransE-based models, such as IKRL and TransAE, represent entity relationships as translations in vector space. While TransE is effective in handling simple relationships, its linear nature limits its ability to model complex multimodal interactions, which is reflected in the weaker performance of TransAE across all H@k metrics. In contrast, ComplEx-based models, such as RSME, introduce complex-valued embeddings that better capture asymmetric relationships—a key requirement in event prediction. This ability to represent more nuanced relational structures explains why RSME consistently outperforms TransE-based models, particularly in H@1 (0.28) and MRR (0.29).

The ANALOGY-based models, including IKRL, TransAE, and RSME, employ a higher-order tensor factorization technique that enables deeper reasoning over event structures. While these models generally improve upon TransE-based approaches, they still do not surpass ComplEx or BiCore in most ranking metrics. BiCore, our proposed model, extends traditional embedding methods by integrating BiComplex embeddings with Riemannian fusion, enabling a more expressive multimodal reasoning process. The strong performance of RSME with BiCore in H@1 (0.28), H@3 (0.299), and MRR (0.29) highlights BiCore's ability to rank relevant triples more accurately than standard embedding methods.

Among all BiCore-backed models, RSME with BiCore emerges as the best performer, consistently achieving the highest scores across all evaluation metrics. It records a H@1 of 0.28, surpassing both IKRL-BiCore (0.273) and TransAE-BiCore (0.268), indicating that BiCore's embedding mechanism is particularly effective when combined with ComplEx-based models. Additionally, RSME-BiCore achieves the highest H@3 (0.299) and MRR (0.290), solidifying its ability to retrieve correct event triples with high accuracy. The H@10 score of 0.305 suggests that BiCore, when paired with RSME, also excels in recall-based retrieval, ranking relevant triples within the top 10 results.

For the IKRL model with BiCore as the backbone, the results indicate strong performance in mid-to-high-ranked predictions. It achieves a H@1 score of 0.273, placing it above the TransAE-BiCore model (0.268) but slightly behind RSME-BiCore (0.280). However, IKRL-BiCore outperforms TransAE-BiCore in every ranking metric, particularly in H@3 (0.296) and H@5 (0.3), demonstrating its effectiveness in retrieving correct entities within top-k ranks. The MRR score of 0.284 is also higher than TransAE-BiCore (0.281), reinforcing that IKRL benefits more from BiCore's multimodal integration than TransAE does. The TransAE model with BiCore underperforms compared to IKRL-BiCore and RSME-BiCore, indicating that TransAE may not fully leverage BiCore's advanced embedding space. Its H@1 (0.268) is the lowest among the BiCore-backed models, reflecting potential limitations in how TransAE interacts with BiComplex embeddings.

Similarly, H@3 (0.289), H@5 (0.292), and H@10 (0.295) remain competitive but consistently trail behind RSME-BiCore and IKRL-BiCore, suggesting that TransAE may require additional architectural modifications to maximize BiCore's potential.

The results strongly indicate that BiCore enhances event prediction across all backbone architectures, but its impact varies depending on the base model. The RSME backbone benefits the most, achieving state-of-the-art results in H@1, H@3, and MRR, proving that BiCore's BiComplex embeddings and Riemannian fusion significantly improve event ranking accuracy. The success of RSME-BiCore in H@10 also highlights BiCore's ability to balance precision and recall, ensuring that important event triples are retrieved effectively. On the other hand, IKRL with BiCore exhibits strong mid-range ranking performance, particularly in H@3 and H@5. However, its H@1 score suggests that while BiCore enhances IKRL's ability to retrieve correct triples within top-k, it may not fully optimize exact match retrieval in top-1 scenarios. This could indicate a need for fine-tuning BiCore's interaction with IKRL's relational representations.

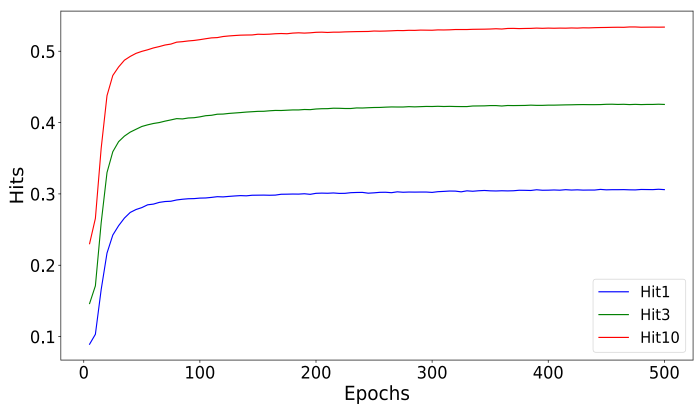

Fig. 5. Training dynamics of BiCore on the DB15K dataset, showing the evolution of H@K metrics over 500 epochs.

Fig. 5 illustrates the growth of H@K metrics performance over training epochs for the DB15K dataset, providing insights into BiCore's learning behavior and convergence properties. The progression of H@K provides further insights into how BiCore refines its ranking predictions over time. H@1, represented by the blue line, increases steeply in the initial epochs, indicating that BiCore rapidly learns to predict the most probable entity correctly at rank 1. This suggests that BiCore prioritizes precise predictions early in training, aligning well with the objective of improving top-1 ranking accuracy. Following this, H@3 and H@10 exhibit a more gradual rise, reflecting improvements in the model's ability to rank relevant entities within the top-3 and top-10 positions. The stable growth of H@K confirms that BiCore maintains strong ranking performance across multiple recall levels.

5.5.2. Influence of regularization

In this section, we analyze the effect of different regularization techniques on BiCore's performance in link prediction tasks across three benchmark datasets: DB15K, MKG-W, and MKG-Y. Specifically, we compare N3 regularization and its enhanced version, N3FW, to evaluate their impact on model accuracy, reasoning capability, and robustness in MKGC. N3 regularization applies a cubic L2 norm penalty to the model's parameters, effectively controlling embedding magnitudes to mitigate overfitting. This regularization is widely used in KGE models, ensuring that BiCore does not rely excessively on any single feature or modality. By applying this constraint, N3 enhances link prediction performance by improving embedding generalization across unseen entity-relation triples. Meanwhile, N3FW regularization, an extension of N3, introduces additional constraints by incorporating the Frobenius norm (which penalizes large values in embedding matrices) and the Wasserstein distance, computed via the Sinkhorn-Knopp algorithm. This modification is particularly useful for MKGC, where textual and visual information require effective integration in a unified space. The added constraints ensure that BiCore maintains stable fusion across different modalities, reducing representational biases. Fig. 6 illustrates a comparison of H@1 scores for BiCore trained with N3 and N3FW regularizations.

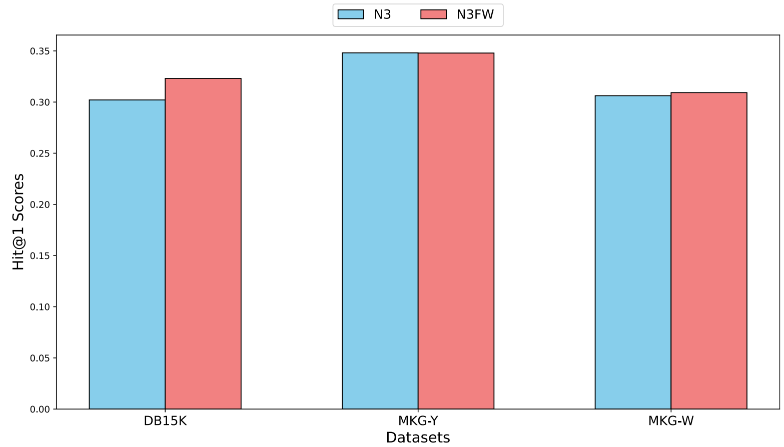

Fig. 6. Comparison of H@1 scores under N3 and N3FW regularization schemes across three benchmark datasets. Results show consistent performance gains with N3FW, highlighting its effectiveness in enhancing link prediction accuracy.

For DB15K, BiCore achieves higher H@1 scores with N3 regularization, suggesting that the standard cubic penalty efficiently regulates embeddings in simpler knowledge graph environments. The results indicate that N3 excels in structured, text-heavy datasets, where multimodal fusion is less complex and knowledge completion depends more on structured entity-relation patterns. Conversely, for MKG-W and MKG-Y, BiCore performs slightly better with N3FW regularization, although the improvement is marginal. The integration of Wasserstein-based constraints and Frobenius normalization likely improves stability in multimodal fusion, ensuring more effective reasoning when textual and visual embeddings must be jointly interpreted. These results suggest that while N3 is a strong baseline regularization, N3FW is better suited for more complex multimodal settings, where additional structural constraints enhance cross-modal interactions.

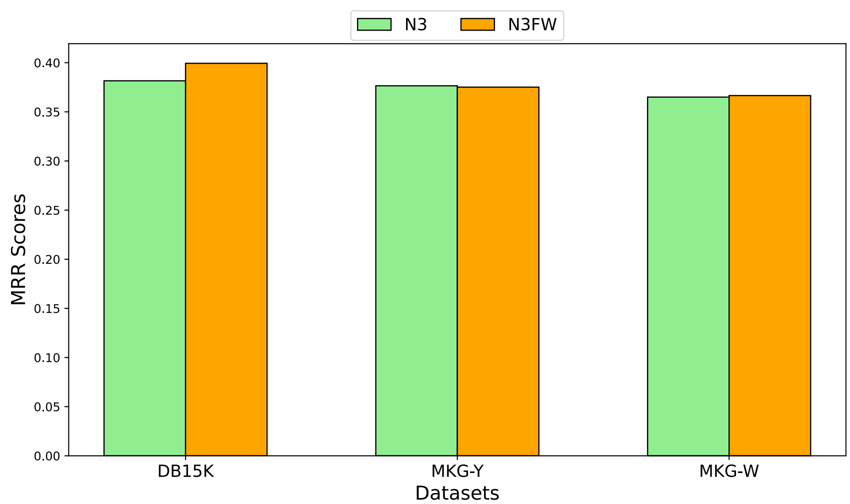

Fig. 7. MRR comparison between N3 and N3FW regularization strategies across three datasets. The results indicate that N3FW maintains or slightly improves MRR scores, reinforcing its potential for enhancing ranking-based performance.

To further evaluate the influence of N3 and N3FW regularizations on BiCore's ranking effectiveness, we analyze their impact on MRR scores. Fig. 7 presents a comparative visualization of MRR scores achieved using both regularization strategies, highlighting how different regularization techniques affect BiCore's ability to rank the correct entity within the top positions. The results indicate that N3FW consistently outperforms N3 across all datasets, as evidenced by the higher orange bars compared to the green bars. On DB15K, the improvement is most pronounced, with N3FW leading to a clear increase in MRR compared to N3. This suggests that the enhanced constraints introduced by N3FW, including Wasserstein-based and Frobenius norm penalties, improve BiCore's ability to generalize and accurately rank correct entities in simpler structured KGs. The impact is less pronounced in MKG-Y and MKG-W, where the difference between N3 and N3FW is smaller, but N3FW still achieves marginally higher scores.

These findings reinforce the hypothesis that N3FW is particularly beneficial in multimodal KGC tasks, where the need for effective cross-modal alignment is greater. The incorporation of Wasserstein distance minimization and Frobenius normalization ensures that BiCore's embeddings remain well-structured, reducing overfitting to single-modality biases and improving the overall ranking performance. The consistent MRR improvement across datasets indicates that N3FW is a more effective regularization strategy for BiCore, particularly when working with heterogeneous MKG.

5.5.3. Influence of embedding dimensions

The choice of embedding dimension significantly impacts the effectiveness of KGC models, particularly in the context of MKGs, where entities and relations require expressive representations to capture both structural and multimodal dependencies. While higher embedding dimensions provide a richer representation space that can model complex relational structures and multimodal interactions more effectively, they also come with increased computational costs and a higher risk of overfitting. On the other hand, lower-dimensional embeddings offer improved efficiency but may struggle to retain sufficient information for accurate reasoning.

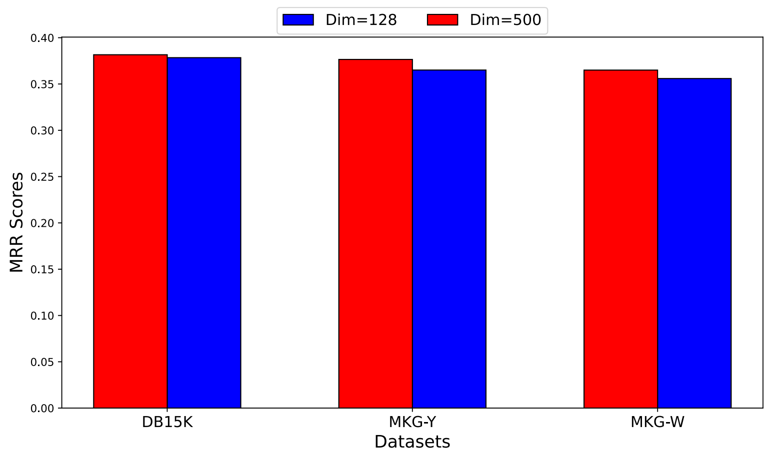

Fig. 8. Impact of embedding dimensionality on MRR scores across three benchmark datasets: DB15K, MKG-Y, and MKG-W. The comparison highlights the stability of BiCore across varying embedding sizes, with minimal performance degradation observed at lower dimensions.

To investigate the trade-off between model expressiveness and computational efficiency, we experimented with different embedding dimensions, focusing on 128-dimensional and 500-dimensional embeddings as representative cases. While it is impractical to test every possible dimensional setting within this range due to computational constraints, these two configurations allow us to analyze performance trends across low, moderate, and high-dimensional settings. Through our evaluations, we aim to understand how increasing the embedding size impacts the model's ability to generalize across datasets and whether a higher-dimensional space consistently leads to better performance or if diminishing returns are observed. This analysis provides insights into selecting an optimal embedding size that balances expressiveness and computational feasibility.

The improvement can be attributed to the nature of BiComplex embeddings, which are divided into four components: real, imaginary, extra-real, and extra-imaginary. When using 128-dimensional embeddings, each component is constrained to only 32 dimensions, limiting the model's ability to capture complex relational structures such as symmetry, anti-symmetry, inversion, and composition. In contrast, with 500-dimensional embeddings, each component has 125 dimensions, allowing for richer feature representations and more effective reasoning over knowledge graph relationships.

Among the datasets, DB15K exhibits the most substantial performance gain, suggesting that it benefits significantly from a higher embedding size due to its diverse relational structures. In contrast, MKG-Y and MKG-W show a more modest improvement, indicating that their relational complexity might already be sufficiently captured even with a lower embedding dimension.

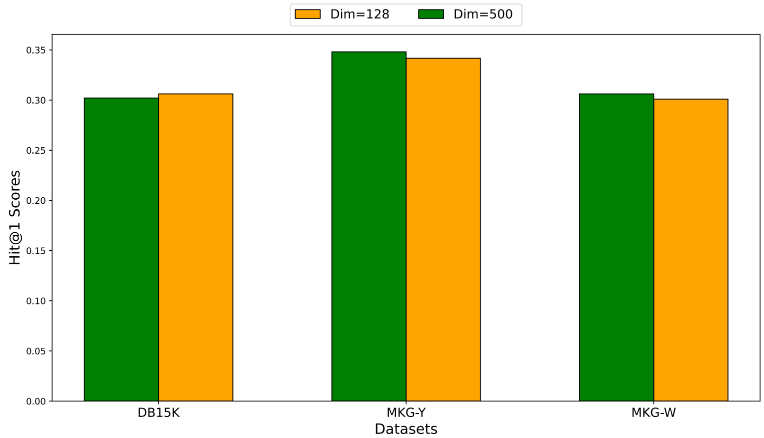

Fig. 9. Effect of embedding dimensionality on H@1 scores across DB15K, MKG-Y, and MKG-W datasets. The results indicate that BiCore maintains competitive performance across different embedding sizes, with only marginal variation in top-1 accuracy.

Our experiments on DB15K, MKG-Y, and MKG-W datasets reveal that increasing the embedding dimension from 128 to 500 has a positive but dataset-dependent impact on model performance. Specifically, MRR improves across all datasets, indicating that a higher embedding dimension enhances the model's ability to capture complex relational structures. However, the improvement is marginal, suggesting that beyond a certain threshold, increasing the embedding size does not yield significant gains.

For Hit@1, the results are mixed. While MKG-Y and MKG-W exhibit slight improvements, DB15K shows a minor decline in performance. This suggests that although larger embeddings provide richer representations, they may introduce additional complexity that hinders the model's ability to make precise predictions at rank 1. This is particularly relevant in BiComplex models, where the embedding dimension is split into multiple components (real, imaginary, extra real, extra imaginary). A higher embedding dimension should ideally allow better representation learning, but it also increases the risk of overfitting or difficulty in optimization, especially for entities with sparse relational connections.

While a higher embedding dimension generally benefits MRR, its impact on Hit@1 is less consistent. The findings suggest that increasing the embedding dimension alone is not always the best strategy for improving model performance, especially when precise ranking is required. Instead, a balance between embedding size and model regularization is crucial. Future work could explore adaptive embedding strategies or dynamic dimension allocation to optimize performance across different datasets and task-specific needs.

5.6. Ablation study

The BiCore model represents a significant advancement in MKG reasoning, effectively integrating heterogeneous data sources such as text, images, and structured relations. By extending traditional complex embeddings with additional mathematical dimensions, BiCore provides a more expressive representation, capable of capturing intricate relational dependencies. At the core of its architecture lie several key mechanisms, including Fisher-Rao distance fusion, curvature adjustment, and Transformer-based reasoning, all working in synergy to enhance the model's predictive and reasoning capabilities.

To assess the individual contribution of these components, we conduct an ablation study by systematically omitting each key mechanism while retaining BiCore's core structure. This controlled experimental setup allows us to examine how each architectural modification impacts model performance. The evaluation is carried out using MRR and H@K on the validation set, providing insight into convergence trends and stability.

One variant of BiCore under investigation in this study excludes the external knowledge graph MarKG during training. The original BiCore framework incorporates MarKG to provide structured semantic relationships across multimodal data sources, reinforcing the model's ability to learn meaningful relational dependencies. By removing this component, we train BiCore solely on multimodal embeddings, without leveraging external structured knowledge.

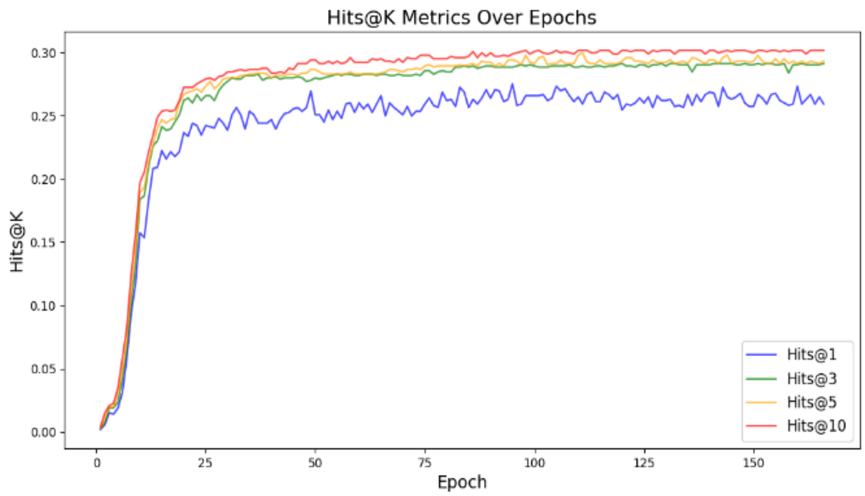

Fig. 10. H@K performance over training epochs for the ablated model without the MarKG component. The results demonstrate slower convergence and lower final accuracy compared to the full BiCore model, highlighting the importance of MarKG in enhancing top-K prediction performance.

Fig. 10 visualizes the training dynamics of this variant of BiCore, tracking H@K performance over epochs. In the H@K results, a steep rise in accuracy is observed within the first 50 epochs, reflecting the model's ability to rapidly learn from multimodal embeddings. However, after this initial phase, performance plateaus, particularly in H@1, suggesting that the model struggles to rank the most probable entity at the top position without MarKG. While H@3, H@5, and H@10 continue to exhibit some level of improvement, their gains are notably less pronounced compared to BiCore's full version, confirming the role of external knowledge supervision in refining ranking accuracy.

To further investigate the contribution of different architectural components in BiCore, we conduct a series of ablation experiments on the DB15K, MKG-W, and MKG-Y datasets for link prediction tasks. Table 5 presents the results of these controlled experiments, where we systematically remove or modify key components of BiCore, including Fisher-Rao distance fusion, curvature adjustment, Transformer-based representation learning, and BiComplex embeddings. These ablation versions are compared against the full BiCore model, allowing us to quantify the impact of each component on overall performance.

Table 5
Link prediction performance of BiCore and its model variants on DB15K, MKG-W, and MKG-Y datasets.

Variants	DB15K				MKG-W				MKG-Y			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
BiCore w/o Fisher-Rao	.385	.306	.425	.533	<u>.367</u>	<u>.312</u>	.375	.451	<u>.375</u>	<u>.347</u>	.389	<u>.425</u>
BiCore w/o curvature	<u>.397</u>	<u>.321</u>	<u>.433</u>	<u>.542</u>	.363	.306	.387	.469	.370	.345	.381	.416
BiCore w/o Transformers	.396	.321	.430	.541	.366	.309	<u>.390</u>	<u>.472</u>	.375	.346	<u>.390</u>	.417
BiCore w/ ling	.396	.320	.433	.540	.365	.307	.387	.471	.370	.342	.385	.416
BiCore w/ Complex	.381	.304	.430	.521	.364	.305	.389	.470	.373	.346	.387	.421
BiCore	**.399**	**.323**	**.435**	**.546**	**.369**	**.313**	**.393**	**.477**	**.376**	**.348**	**.391**	**.427**

The first ablation examines the impact of removing Fisher-Rao distance fusion (BiCore w/o Fisher-Rao). Fisher-Rao distance plays a critical role in aligning multimodal embeddings by capturing the geometric similarity between linguistic and visual features. The results show a notable performance drop across all datasets, particularly in MRR (0.385 → 0.399 on DB15K, 0.367 → 0.369 on MKG-W, 0.375 → 0.376 on MKG-Y). The largest degradation is observed in H@1, where BiCore w/o Fisher-Rao scores 0.306 on DB15K, significantly lower than the 0.323 achieved by the full BiCore model. This suggests that removing Fisher-Rao distance weakens BiCore's ability to accurately rank the most probable entities, confirming its essential role in multimodal representation fusion.

Next, we analyze the effect of removing curvature adjustment (BiCore w/o curvature). Curvature plays a role in adapting entity embeddings based on the structural characteristics of the multimodal space, ensuring that fused representations remain geometrically meaningful. While the overall performance drop is less pronounced than in the Fisher-Rao ablation, there is still a clear degradation in H@1 and H@3 across all datasets. The DB15K dataset sees H@1 decrease from 0.323 to 0.321, indicating that curvature adjustment refines BiCore's ranking accuracy. The drop is even more evident in MKG-W (0.313 → 0.306 in H@1) and MKG-Y (0.348 → 0.345 in H@1), reinforcing the hypothesis that curvature is crucial for effectively modeling entity relations in MKGs.

Another crucial component in BiCore is the Transformer-based fusion mechanism, which enhances multimodal reasoning by learning hierarchical interactions between entities and relations. The BiCore w/o Transformers ablation demonstrates that while removing Transformers does not entirely disrupt BiCore's performance, it results in a marginal drop in MRR and H@K scores across datasets. Specifically, the DB15K dataset records an MRR decrease from 0.399 to 0.396, and H@1 falls from 0.323 to 0.321. These results indicate that while BiCore remains effective without Transformers, their inclusion provides additional benefits in capturing fine-grained entity relationships.

We further assess the impact of modifying Fisher-Rao distance fusion to focus primarily on linguistic embeddings (BiCore w/ ling). This variation forces the Fisher-Rao metric to be more aligned with text-based features, reducing the influence of visual modalities. The results reveal a consistent performance decline, particularly in H@1 and H@3, suggesting that visual information plays a crucial role in BiCore's reasoning process. The H@1 score drops from 0.323 to 0.320 on DB15K, confirming that an imbalanced fusion strategy limits BiCore's ability to leverage the full potential of MKGs.

Finally, we investigate the effect of replacing BiComplex embeddings with standard Complex embeddings (BiCore w/ Complex). The BiComplex embedding space enables richer representations by extending entity embeddings into four-dimensional space, capturing both symmetric and asymmetric relations more effectively. The results clearly demonstrate that switching to Complex embeddings leads to the most significant drop in performance among all ablation experiments. On DB15K, MRR falls from 0.399 to 0.381, and H@1 drops from 0.323 to 0.304, indicating a substantial degradation in predictive accuracy. Similar trends appear in MKG-W (MRR: 0.369 → 0.364) and MKG-Y (MRR: 0.376 → 0.373), reinforcing that BiComplex embeddings are a fundamental component in BiCore's success.

6. Case study

We also conduct a case study to evaluate the BiCore model's accuracy in predicting entity relationships across three scenarios, comparing its performance against the RSME model. Table 6 and Table 7 show these scenarios.

Table 6
Case study comparing RSME and BiCore models on selected samples from the DB15K dataset. Note that only structural triples are used in this evaluation, without incorporating multimodal information such as images or textual descriptions.

Query	Ground Truth	RSME	BiCore
(*China, largest city, ?*)	Shanghai	Beijing (rank 1) **Shanghai** (rank 2)	**Shanghai** (rank 1)
(*New York metropolitan area, is part of, ?*)	Waterbury, Connecticut	Frederick, Maryland (rank 1) Michigan, Indiana (rank 2) **Waterbury, Connecticut** (rank 3)	**Waterbury, Connecticut** (rank 1)
(*Northeastern University, is located in, ?*)	Massachusetts	Florida (rank 1) … **Massachusetts** (rank 5)	**Massachusetts** (rank 1)
(*Richard Marx, use an*	Guitar	bassoon (rank 1)	electric bass (rank 1)

instrument, ?)		… **guitar** (rank 7)	drum kit (rank 2) **guitar** (rank 3)
(*Treasure Planet, is distributed, ?*)	Walt Disney Studios Motion Pictures	Walt Disney Studios Home Entertainment (rank 1) … **Walt Disney Studios Motion Pictures** (rank 10)	Paramount Pictures (rank 1) **Walt Disney Studios Motion Pictures** (rank 2)

The first scenario examines China's largest city. RSME makes an inaccurate prediction by ranking Beijing as the biggest city while placing the correct answer, Shanghai, in second place. In contrast, BiCore demonstrates greater precision by correctly identifying Shanghai as the largest city. The second test evaluates which city includes the New York metropolitan area. RSME incorrectly ranks Frederick first, Michigan second, and Waterbury third, failing to match the actual data. Meanwhile, BiCore successfully aligns with the ground truth by correctly identifying Waterbury as part of the New York metropolitan area. The third scenario involves determining the location of Northeastern University. RSME misidentifies the state in which the university is located, ranking the correct answer, Massachusetts, only in fifth place. On the other hand, BiCore once again demonstrates accuracy by correctly identifying Massachusetts as the right answer. The fourth example focuses on the musical instrument played by Richard Marx. RSME ranks the correct answer, guitar, in seventh place, while BiCore also makes an error by listing guitar in third place instead of first. The final scenario examines the distributor of the film Treasure Planet. While the correct answer is Walt Disney Studios Motion Pictures, both RSME and BiCore fail to provide a fully accurate prediction. RSME ranks the correct answer in tenth place, while BiCore, though closer, still places it second instead of first.

Table 7
Case study comparing RSME and BiCore models on selected samples from the DB15K dataset, using both structural triples and multimodal information (e.g., images and textual descriptions).

Query	Ground Truth	RSME	BiCore
(*China, largest city, ?*)	Shanghai	Beijing (rank 1) **Shanghai** (rank 2)	**Shanghai** (rank 1)
(*New York metropolitan area, is part of, ?*)	Waterbury, Connecticut	Frederick, Maryland (rank 1) Michigan, Indiana (rank 2) **Waterbury, Connecticut** (rank 3)	**Waterbury, Connecticut** (rank 1)
(*Northeastern University, is located in, ?*)	Massachusetts	Florida (rank 1) **Massachusetts** (rank 2)	**Massachusetts** (rank 1)
(*Richard Marx, use an instrument, ?*)	Guitar	bassoon (rank 1) … **guitar** (rank 5)	**guitar** (rank 1)
(*Treasure Planet, is distributed, ?*)	Walt Disney Studios Motion Pictures	Walt Disney Studios Home Entertainment (rank 1) **Walt Disney Studios Motion Pictures** (rank 2)	**Walt Disney Studios Motion Pictures** (rank 1)

When utilizing multimodal data, we can observe that model improvements enhance the ability to predict correct answers. We focus on the two last examples where BiCore and RSME have the incorrect ranking answers on Table 6. The fourth example answers the question on the musical instrument played by Richard Marx. RSME ranks the correct answer, guitar, in fifth place, while BiCore provides the correct response by listing guitar in first place. With the image of Richard Marx playing guitar combined with text description information he is a pop rock singer-songwriter, finally our model can reason the correct answer is guitar in first rank. The final scenario evaluates the distributor of the film Treasure Planet. While the correct answer is Walt Disney Studios Motion Pictures, RSME does not deliver a fully accurate prediction, ranking it in second place. In contrast, BiCore correctly identifies it as the first choice. Similarly, the film Treasure Planet has the text description that it was produced by Walt Disney Feature Animation and distributed by Walt Disney Pictures, which helped BiCore correctly identify Walt Disney Studios Motion Pictures as the first place.

7. Conclusion

In this paper, we introduced BiCore, a novel MKGC model that leverages BiComplex embeddings combined with Fisher-Rao distance fusion and curvature-aware representations, enhancing relational reasoning in multimodal environments. Our experiments on the DB15K, MKG-W, MKG-Y, and MARS datasets demonstrate that BiCore consistently outperforms state-of-the-art models, achieving up to 17.92% MRR improvement in link prediction, and showing significant performance in analogy reasoning over traditional models. BiCore not only improves performance in combining textual and visual data but also provides a more robust model for handling complex relationships in KGs, enabling better understanding of the intricate interactions between entities and relations within the knowledge space. The use of BiComplex embeddings allows the model to represent abstract relationships between entities, improving its ability to handle non-linear data. While BiCore shows strong performance, several directions remain open for future work. First, scaling to large graphs could be improved via hyperbolic embedding fusion, which would better capture hierarchical structures common in domains like biomedicine or chemistry. Second, implementing adaptive embedding allocation such as dynamically adjusting embedding sizes based on the importance of entities and relations could improve resource utilization and scalability.

Data availability

Data will be made available on request.

References

Balažević, I., Allen, C., & Hospedales, T. (2019). Multi-relational poincaré graph embeddings. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 4463–4473.

Balazevic, I., Allen, C., & Hospedales, T. (2019). TuckER: Tensor Factorization for Knowledge Graph Completion. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 5184–5193. https://doi.org/10.18653/v1/D19-1522

Bellomarini, L., Benedetto, D., Gottlob, G., & Sallinger, E. (2022). Vadalog: A modern architecture for automated reasoning with large knowledge graphs. *Information Systems*, *105*, 101528. https://doi.org/10.1016/j.is.2020.101528

Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, 2787–2795.

Cao, Z., Xu, Q., Yang, Z., He, Y., Cao, X., & Huang, Q. (2022). OTKGE: Multi-modal Knowledge Graph Embeddings via Optimal Transport. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in Neural Information Processing Systems* (Vol. 35, pp. 39090–39102). Curran Associates, Inc.

Chen, T., Wang, T., Zhang, H., & Xu, J. (2025). M2KGRL: A semantic-matching based framework for multimodal knowledge graph representation learning. *Expert Systems with Applications*, *269*, 126388. https://doi.org/10.1016/j.eswa.2025.126388

Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2d knowledge graph embeddings. *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, *32*, 1811–1818. https://doi.org/10.1609/aaai.v32i1.11573

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, *12*(61), Article 61.

Kaiser, M., Roy, R. S., & Weikum, G. (2021). Reinforcement Learning from Reformulations in Conversational Question Answering over Knowledge Graphs. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 459–469. https://doi.org/10.1145/3404835.3462859

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *CoRR*, *abs/1412.6980*. https://api.semanticscholar.org/CorpusID:6628106

Lao, N., Mitchell, T., & Cohen, W. W. (2011). Random Walk Inference and Learning in A Large Scale Knowledge Base. In R. Barzilay & M. Johnson (Eds.), *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 529–539). Association for Computational Linguistics. https://aclanthology.org/D11-1049

Lee, J., Chung, C., Lee, H., Jo, S., & Whang, J. (2023). VISTA: Visual-Textual Knowledge Graph Representation Learning. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 7314–7328). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.488

Lee, J. M. (1997). Curvature. In J. M. Lee, *Riemannian Manifolds* (Vol. 176, pp. 115–129). Springer New York. https://doi.org/10.1007/0-387-22726-1_7

Li, X., Zhao, X., Xu, J., Zhang, Y., & Xing, C. (2023). IMF: Interactive Multimodal Fusion Model for Link Prediction. *Proceedings of the ACM Web Conference 2023*, 2572–2580. https://doi.org/10.1145/3543507.3583554

Li, Y. (2017). Research and Analysis of Semantic Search Technology Based on Knowledge Graph. *22017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 887–890. https://doi.org/10.1109/CSE-EUC.2017.179

Liu, Y., Li, H., Garcia-Duran, A., Niepert, M., Onoro-Rubio, D., & Rosenblum, D. S. (2019). MMKG: Multi-modal Knowledge Graphs. In P. Hitzler, M. Fernández, K. Janowicz, A. Zaveri, A. J. G. Gray, V. Lopez, A. Haller, & K. Hammar (Eds.), *The Semantic Web* (pp. 459–474). Springer International Publishing. https://doi.org/10.1007/978-3-030-21348-0_30

Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 13–23.

Mahdisoltani, F., Biega, J., & Suchanek, F. (2015). YAGO3: A knowledge base from multilingual wikipedias. *Proceedings of the 7th Biennial Conference on Innovative Data Systems Research*.

Miyamoto, H. K., Meneghetti, F. C. C., Pinele, J., & Costa, S. I. R. (2024). On closed-form expressions for the Fisher–Rao distance. *Information Geometry*, *7*(2), Article 2. https://doi.org/10.1007/s41884-024-00143-2

Mousselly-Sergieh, H., Botschen, T., Gurevych, I., & Roth, S. (2018). A Multimodal Translation-Based Approach for Knowledge Graph Representation Learning. In M. Nissim, J. Berant, & A. Lenci (Eds.), *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics* (pp. 225–234). Association for Computational Linguistics. https://doi.org/10.18653/v1/S18-2027

Niu, G., Li, B., Zhang, Y., & Pu, S. (2022). *CAKE: A Scalable Commonsense-Aware Framework For Multi-View Knowledge Graph Completion* (No. arXiv:2202.13785; Issue arXiv:2202.13785). arXiv. https://doi.org/10.48550/arXiv.2202.13785

Opdahl, A. L., Al-Moslmi, T., Dang-Nguyen, D.-T., Gallofré Ocaña, M., Tessem, B., & Veres, C. (2022). Semantic Knowledge Graphs for the News: A Review. *ACM Comput. Surv.*, *55*(7), 140:1-140:38. https://doi.org/10.1145/3543508

Pan, Z., & Wang, P. (2021). Hyperbolic Hierarchy-Aware Knowledge Graph Embedding for Link Prediction. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2941–2948. https://doi.org/10.18653/v1/2021.findings-emnlp.251

Qin, S., & Chow, K. P. (2019). Automatic Analysis and Reasoning Based on Vulnerability Knowledge Graph. In H. Ning (Ed.), *Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health* (Vol. 1137, pp. 3–19). Springer Singapore. https://doi.org/10.1007/978-981-15-1922-2_1

Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling Relational Data with Graph Convolutional Networks. In A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, & M. Alam (Eds.), *The Semantic Web* (pp. 593–607). Springer International Publishing. https://doi.org/10.1007/978-3-319-93417-4_38

Sun, Z., Deng, Z.-H., Nie, J.-Y., & Tang, J. (2019). RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. *Proceedings of 7th International Conference on Learning Representations*.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., & Bouchard, G. (2016). Complex embeddings for simple link prediction. *International Conference on Machine Learning*, *48*, 2071–2080.

Tu, L. W. (2011). *An Introduction to Manifolds*. Springer New York. https://doi.org/10.1007/978-1-4419-7400-6

Vu, M., Nebgen, B., Skau, E., Zollicoffer, G., Castorena, J., Rasmussen, K., Alexandrov, B., & Bhattarai, M. (2024). *LaFA: Latent Feature Attacks on Non-negative Matrix Factorization* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2408.03909

Wang, J., Xie, H., Zhang, S., Qin, S. J., Tao, X., Wang, F. L., & Xu, X. (2025). Multimodal fusion framework based on knowledge graph for personalized recommendation. *Expert Systems with Applications*, *268*, 126308. https://doi.org/10.1016/j.eswa.2024.126308

Wang, M., Wang, S., Yang, H., Zhang, Z., Chen, X., & Qi, G. (2021). Is Visual Context Really Helpful for Knowledge Graph? A Representation Learning Perspective. *Proceedings of the 29th ACM International Conference on Multimedia*, 2735–2743. https://doi.org/10.1145/3474085.3475470

Wang, X., Meng, B., Chen, H., Meng, Y., Lv, K., & Zhu, W. (2023). TIVA-KG: A Multimodal Knowledge Graph with Text, Image, Video and Audio. *Proceedings of the 31st ACM International Conference on Multimedia*, 2391–2399. https://doi.org/10.1145/3581783.3612266

Wang, Z., Li, L., Li, Q., & Zeng, D. (2019). Multimodal Data Enhanced Representation Learning for Knowledge Graphs. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN.2019.8852079

Wei, Y., Huang, Q., Zhang, Y., & Kwok, J. (2023). KICGPT: Large Language Model with Knowledge in Context for Knowledge Graph Completion. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 8667–8683). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.580

Wu, Y., Xu, H., & Zhang, J. (2024). Multimodal Representation Learning by Hybrid Transformer with Multi-level Fusion. *2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, 860–865. https://doi.org/10.1109/ICBASE63199.2024.10762462

Xian, Y., Fu, Z., Muthukrishnan, S., de Melo, G., & Zhang, Y. (2019). Reinforcement Knowledge Graph Reasoning for Explainable Recommendation. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 285–294. https://doi.org/10.1145/3331184.3331203

Xie, R., Liu, Z., Luan, H., & Sun, M. (2017). Image-embodied knowledge representation learning. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 3140–3146.

Xu, D., Xu, T., Wu, S., Zhou, J., & Chen, E. (2022). Relation-enhanced Negative Sampling for Multimodal Knowledge Graph Completion. *Proceedings of the 30th ACM International Conference on Multimedia*, 3857–3866. https://doi.org/10.1145/3503161.3548388

Yang, B., Yih, W., He, X., Gao, J., & Deng, L. (2015). Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In Y. Bengio & Y. LeCun (Eds.), *Proceedings of the 3rd International Conference on Learning Representations*.

Yang, J., Yang, S., Gao, Y., Yang, J., & Yang, L. T. (2024). Multimodal Contextual Interactions of Entities: A Modality Circular Fusion Approach for Link Prediction. *Proceedings of the 32nd ACM International Conference on Multimedia*, 8374–8382. https://doi.org/10.1145/3664647.3681696

Yao, L., Mao, C., & Luo, Y. (2019). KG-BERT: BERT for Knowledge Graph Completion. *CoRR*, *abs/1909.03193*. http://arxiv.org/abs/1909.03193

Yu, J., Yin, H., Li, J., Gao, M., Huang, Z., & Cui, L. (2022). Enhancing Social Recommendation With Adversarial Graph Convolutional Networks. *IEEE Transactions on Knowledge and Data Engineering*, *34*(8), 3727–3739. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2020.3033673

Yu, M., Bai, C., Yu, J., Zhao, M., Xu, T., Liu, H., Li, X., & Yu, R. (2022). Translation-Based Embeddings with Octonion for Knowledge Graph Completion. *Applied Sciences*, *12*(8), Article 8. https://doi.org/10.3390/app12083935

Zhang, N., Li, L., Chen, X., Liang, X., Deng, S., & Chen, H. (2023). *Multimodal Analogical Reasoning over Knowledge Graphs*. *arXiv:2210.00312*.

Zhang, S., Tay, Y., Yao, L., & Liu, Q. (2019). Quaternion knowledge graph embeddings. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2735–2745.

Zhang, Y., Chen, M., & Zhang, W. (2023a). Modality-Aware Negative Sampling for Multi-modal Knowledge Graph Embedding. *2023 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN54540.2023.10191314

Zhang, Y., Chen, Z., Liang, L., Chen, H., & Zhang, W. (2024). Unleashing the Power of Imbalanced Modality Information for Multi-modal Knowledge Graph Completion. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 17120–17130). ELRA and ICCL. https://aclanthology.org/2024.lrec-main.1487/

Zhang, Y., Chen, Z., & Zhang, W. (2023b). MACO: A Modality Adversarial and Contrastive Framework for Modality-Missing Multi-modal Knowledge Graph Completion. In F. Liu, N. Duan, Q. Xu, & Y. Hong (Eds.), *Natural Language Processing and Chinese Computing* (pp. 123–134). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-44693-1_10

Zhang, Y., & Zhang, W. (2022). Knowledge Graph Completion with Pre-trained Multimodal Transformer and Twins Negative Sampling. *CoRR*, *abs/2209.07084*.

Zhao, K., Zhang, Y., Yin, H., Wang, J., Zheng, K., Zhou, X., & Xing, C. (2020). Discovering Subsequence Patterns for Next POI Recommendation. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 3216–3222. https://doi.org/10.24963/ijcai.2020/445

Zheng, S., Wang, W., Qu, J., Yin, H., Chen, W., & Zhao, L. (2023). MMKGR: Multi-hop Multi-modal Knowledge Graph Reasoning. *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 96–109. https://doi.org/10.1109/ICDE55515.2023.00015

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

Zhou, P., Feng, J., Ma, C., Xiong, C., Hoi, S., & Weinan, E. (2020). Towards theoretically understanding why SGD generalizes better than ADAM in deep learning. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 21285–21296.

Zhou, Z., Conia, S., Lee, D., Li, M., Huang, S., Minhas, U. F., Potdar, S., Xiao, H., & Li, Y. (2025). KG-TRICK: Unifying Textual and Relational Information Completion of Knowledge for Multilingual Knowledge Graphs. In O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, & S. Schockaert (Eds.), *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 9096–9111). Association for Computational Linguistics. https://aclanthology.org/2025.coling-main.611/