

# Beyond Geometric Blindness: Leveraging Ollivier-Ricci Curvature for Effective Knowledge Graph Completion

Journal Title  
XX(X):1–10  
©The Author(s) 2016  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/ToBeAssigned  
www.sagepub.com/

SAGE

Donglin Zhang<sup>1</sup>, Haotian Li<sup>1</sup>, Rui Zhang<sup>1</sup>, Lingzhi Wang<sup>1</sup>, Bailing Wang<sup>1</sup>, Yang Liu<sup>1</sup>, Yuliang Wei<sup>1</sup>

## Abstract

Graph Neural Networks (GNNs) are widely used for Knowledge Graph Completion (KGC) but often suffer from “**geometric blindness**”, resulting in redundant message propagation and limiting their ability to perform effective long-range reasoning. We introduce **ORCA-GCN** (**O**llivier-**R**icci **C**urvature-**A**ware **G**raph **C**onvolutional **N**etwork), a novel geometry-aware GNN that integrates **Ollivier-Ricci Curvature (ORC)** into its message-passing mechanism. While prior curvature-guided methods for node classification often enhance high-curvature intra-cluster links, which can be detrimental for KGC, ORCA-GCN strategically **down-weights these redundant high-curvature connections and amplifies low-curvature bridge edges** to improve information flow. It also features a **layer-wise evolutionary framework**, transitioning from geometric priors in shallow layers to learned semantic similarity in deeper layers. Extensive experiments on FB15k-237 and WN18RR demonstrate ORCA-GCN consistently outperforms strong baselines in link prediction. Our analysis confirms that high-curvature intra-cluster edges tend to be semantically redundant, thereby offering limited utility in KGC tasks and underscoring the importance of structural geometry in guiding discriminative representation learning.

## Keywords

Knowledge Graph Completion, Graph Neural Networks, Topology-aware Aggregation, Ollivier-Ricci Curvature

## Introduction

Knowledge Graphs (KGs) represent structured factual information and support a wide range of applications, including semantic search [Xiong et al. \(2017\)](#), recommender systems [Wang et al. \(2019\)](#), and question answering [West et al. \(2014\)](#). Despite their utility, real-world KGs are often incomplete, making automated Knowledge Graph Completion (KGC), especially in the form of link prediction, a critical task.

Graph Neural Networks (GNNs) have become a leading paradigm for KGC by propagating information through multi-hop message passing, as exemplified by models like R-GCN [Schlichtkrull et al. \(2018\)](#) and CompGCN [Vashishth et al. \(2020\)](#). Their strength lies in aggregating neighborhood signals to enrich node representations. However, recent studies have challenged the universal value of message passing. Interestingly, even simple Multi-Layer Perceptrons (MLPs), which entirely forgo structural aggregation, have achieved performance comparable to GNN-based approaches in recent benchmarks [Li et al. \(2023\)](#). This suggests that neighborhood aggregation may not always contribute meaningfully to performance.

We revisit the common GNN assumption of treating all neighboring nodes equally. Such uniform aggregation leads to a phenomenon we term *geometric blindness*—characterized by redundant message passing within dense clusters and insufficient utilization of topologically critical *bridge edges*, as illustrated in Figure 1. This indiscriminate propagation exacerbates two well-known issues in GNNs: *over-smoothing* [Li et al. \(2018\)](#), where node representations

become indistinguishably similar, and *over-squashing* [Alon and Yahav \(2021\)](#), where information from distant nodes is excessively compressed through limited paths. Both effects are detrimental to long-range reasoning in knowledge graphs (KGs).

To address these issues, we incorporate **Ollivier-Ricci Curvature (ORC)** [Ollivier \(2009\)](#), a theoretically grounded and unsupervised metric that quantifies the structural role of each edge in the graph. ORC effectively distinguishes between high-curvature intra-cluster edges (typically redundant) and low-curvature inter-cluster *bridge edges* that facilitate information flow. This distinction is particularly valuable in KGC, where oversaturation with redundant edges can obscure meaningful signals. Our empirical analysis on FB15k-237 shows that removing up to 60% of the highest-curvature edges during aggregation has negligible impact on performance, indicating that many such connections contribute little to KGC.

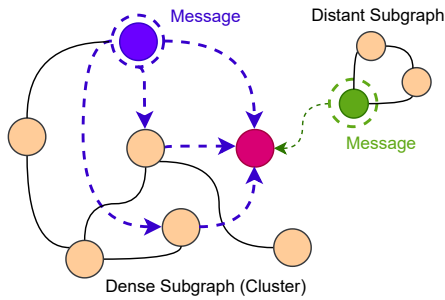
Building on this insight, we propose **ORCA-GCN** (**O**llivier-**R**icci **C**urvature-**A**ware **G**raph **C**onvolutional **N**etwork), a novel geometry-aware GNN that reweights edge importance according to Ollivier-Ricci curvature. Unlike prior curvature-based methods, typically designed for node

<sup>1</sup>Harbin Institute of Technology (Weihai) Qingdao Research Institute

## Corresponding author:

Yuliang Wei, Harbin Institute of Technology (Weihai) Qingdao Research Institute

Email: wei.yl@hit.edu.cn



**Figure 1.** Illustration of redundant message passing in a locally dense subgraph. Nodes within the cluster repeatedly exchange messages (purple dashed arrows). The green dashed edge connects a distant subgraph, but due to uniform treatment of edges, traditional GNNs often underutilize such structurally important connections.

classification and favoring high-curvature connections, ORCA-GCN down-weights structurally redundant edges while amplifying low-curvature bridges to enhance long-range information flow for link prediction. In addition, we introduce a **layer-wise evolutionary strategy**: early layers rely on geometric priors when semantic representations are weak, while deeper layers transition toward learned semantic similarity for refined aggregation. This gradual shift enables ORCA-GCN to adaptively balance structure and semantics across layers.

We validate ORCA-GCN on standard KGC benchmarks including FB15k-237 and WN18RR, where it consistently outperforms strong baselines.

The main contributions of this paper are summarized as follows:

- We identify and characterize the phenomenon of “geometric blindness” of message passing in standard GNNs for Knowledge Graph Completion (KGC) and empirically show that high-curvature intra-cluster edges are often redundant or even detrimental to link prediction performance.
- We propose a geometry-aware aggregation framework that leverages Ollivier-Ricci Curvature to down-weight high-curvature edges and emphasize low-curvature inter-cluster links—contrasting with prior approaches in node classification that favor high-curvature connectivity.
- We introduce ORCA-GCN, which integrates curvature-guided aggregation with a layer-wise transition from geometric to semantic signals, and demonstrate its consistent gains over strong baselines on benchmark KGC datasets.

The remainder of this paper is organized as follows: Section 2 discusses related work. Section 3 details our proposed ORCA-GCN model. Section 4 presents our experimental results and analysis. Finally, Section 5 concludes this paper.

## 2. Related Work

**Knowledge Graph Completion with Graph Neural Networks.** Knowledge graph completion (KGC) aims to infer missing links between entities, typically framed as

a link prediction task. Early methods such as TransE, DistMult, and ComplEx learn entity and relation embeddings based on predefined scoring functions over triples. However, these approaches struggle with multi-hop reasoning and global structure integration. To address this, graph neural networks (GNNs) have emerged as a powerful alternative due to their message-passing capability.

Relational GCN (R-GCN) extends the GCN framework to multi-relational graphs by using relation-specific transformations during aggregation. CompGCN further improves this by jointly embedding entities and relations in a unified space and composing them through various operators (e.g., addition, multiplication, circular correlation). These models achieve strong performance on benchmark KGC datasets. However, they typically assign equal or learned weights to all neighbors without explicit awareness of their topological roles, potentially limiting their ability to distinguish redundant from informative connections.

**Curvature-Guided Graph Neural Networks.** Ricci curvature, a concept from differential geometry, quantifies the local deviation of a space from being flat. Among its extensions to discrete spaces, Ollivier-Ricci Curvature (ORC) has gained significant attention in graph learning for its geometrically meaningful interpretation based on optimal transport theory [Li et al. \(2022\)](#). Intuitively, ORC characterizes the connectivity of neighborhoods: As conceptually illustrated in Figure 2, edges with high positive curvature typically reside within densely connected communities (intra-cluster links), whereas edges with low or negative curvature often act as “bridges” between different communities (inter-cluster links). A detailed formulation of ORC is provided in Section 3.1.

Leveraging this property, several studies have incorporated ORC as a structural prior to guide the message-passing process in GNNs. However, the strategies for utilizing curvature vary significantly, often tailored to the specific downstream task.

For **node classification**, prevailing methods operate on the homophily assumption, where strengthening intra-community signals is paramount. These approaches often treat inter-community bridge connections as noise. Following this line of thought:

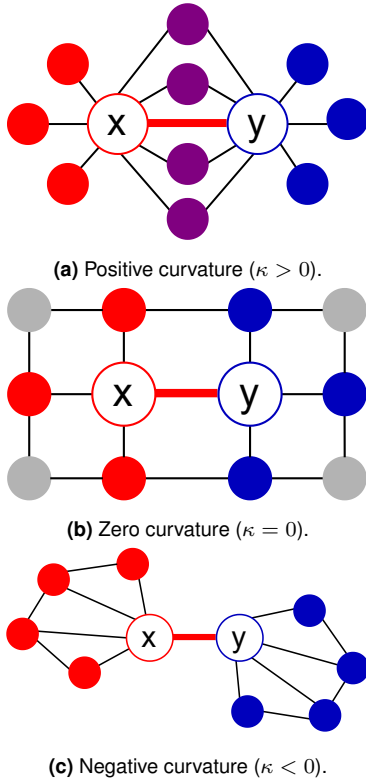
- **CGNN** [Li et al. \(2022\)](#) proposes an explicit method that transforms ORC into aggregation weights through specialized processing and normalization modules, thereby assigning greater weights to edges with high curvature to reinforce intra-community connections.
- Employing a similar rationale through a stochastic lens, **CurvDrop** introduces a dropout strategy that is more inclined to remove edges with lower curvature during training [Liu et al. \(2023\)](#). This approach aims to better isolate communities for node classification by pruning sparse connections between them.

Other methods have adopted data-driven approaches. The pioneering **CurvGN** utilized a multi-layer perceptron (MLP) to learn a mapping from curvature values to message weights [Ye et al. \(2019\)](#). However, subsequent work suggests its success may be largely attributable to the learning capacity of the MLP, with curvature potentially acting as a mere numerical perturbation rather than a source of meaningful

topological information Li et al. (2022). The CD-GCN also follows this MLP-based paradigm, applying it to the task of link prediction Guo et al. (2023).

A recent theoretical work offers a unified geometric framework for understanding key GNN limitations, rigorously demonstrating that **over-smoothing is linked to positive curvature**, while **over-squashing is associated with negative curvature** Nguyen et al. (2023). Based on this duality, the **BORF** algorithm was proposed as a graph rewiring method that mitigates these issues by removing high-curvature edges and adding new connections around low-curvature ones before training begins Nguyen et al. (2023).

The aforementioned methods provide valuable insights but either focus on node classification, employ implicit learning schemes, or alter the graph’s fundamental topology. In contrast, our work targets **knowledge graph completion (KGC)**, a task that often demands effective multi-hop and long-range reasoning. We hypothesize that for KGC, high-curvature intra-cluster links can be information-redundant, whereas low-curvature bridge edges are vital for global information flow. Therefore, our proposed **ORCA-GCN** takes a distinct approach by intentionally down-weighting high-curvature edges while amplifying the influence of low-curvature ones, a strategy motivated by the unique demands of the KGC task.



**Figure 2. Conceptual Illustrations of Ollivier-Ricci Curvature (ORC) on Graphs.** This figure demonstrates how ORC reflects the local topology surrounding an edge. (a) Positive curvature ( $\kappa > 0$ ) typically signifies an edge within a dense community. (b) Zero curvature ( $\kappa = 0$ ) corresponds to regular, grid-like structures. (c) Negative curvature ( $\kappa < 0$ ) identifies a “bridge” edge connecting two distinct communities.

### 3. Methodology

This section introduces the proposed ORCA-GCN model. We begin by adapting Ollivier-Ricci curvature (ORC) Ollivier (2009) to accommodate the directed, multi-relational, and self-loop-enriched characteristics of knowledge graphs. Then, we propose a Geometry-Aware Aggregation module that leverages curvature-based weights to prioritize structurally significant (i.e., negatively curved) edges while attenuating redundant intra-cluster connections. Subsequently, a Semantic-Aware Aggregation mechanism refines message propagation in deeper layers by exploiting cosine dissimilarity between node embeddings to capture semantic divergence. Together, these components enhance both structural sensitivity and feature-level expressiveness. The overall architecture of ORCA-GCN is depicted in Figure 3.

#### 3.1 Ollivier-Ricci Curvature

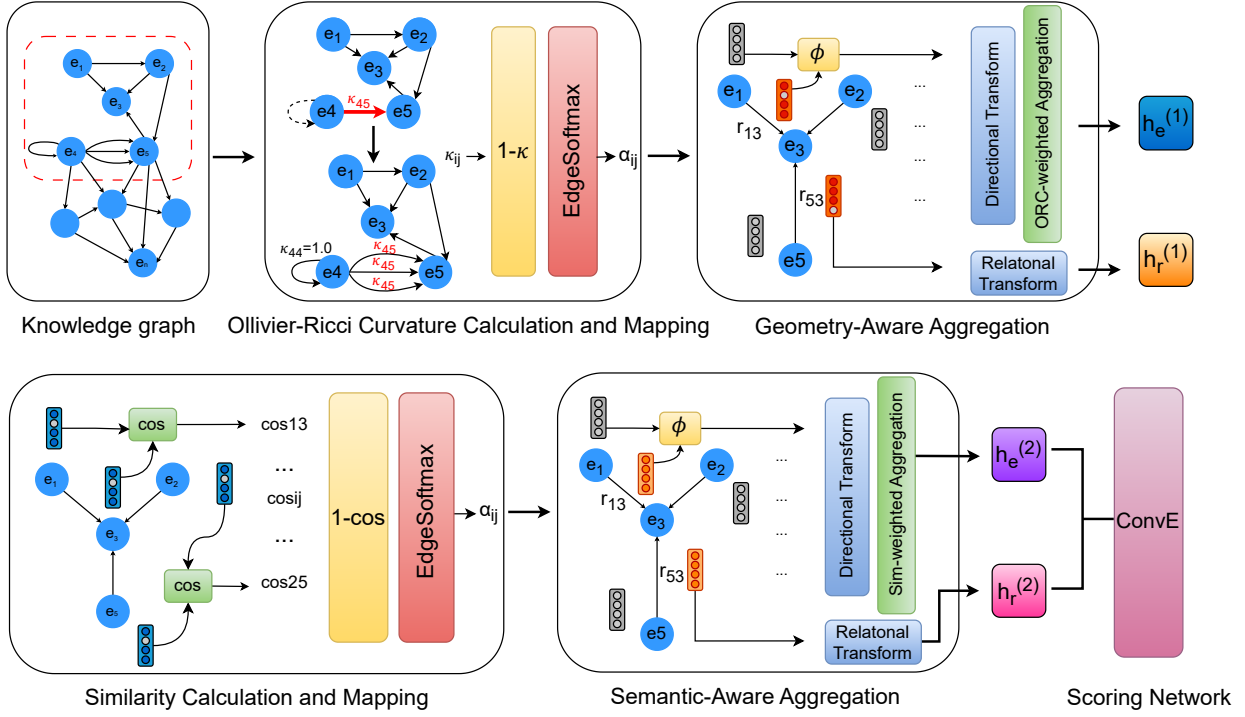
Ricci curvature originates from differential geometry, where it quantifies how much a space deviates from being locally Euclidean by measuring the degree to which geodesics converge or diverge. In positively curved regions, geodesics tend to converge; in negatively curved regions, they diverge do Carmo (1992). Ollivier-Ricci curvature (ORC) Ollivier (2009) extends this concept to metric measure spaces, including graphs, by comparing the optimal transport distance between local probability distributions. In this framework, each node is considered the center of a probability distribution over its neighborhood. ORC then quantifies the transportation cost between these local distributions, capturing the graph’s intrinsic geometric structure. Specifically, for a pair of nodes  $i$  and  $j$ , let  $m_i$  and  $m_j$  denote the probability distributions over their respective neighborhoods. In unweighted graphs, these are uniform; in weighted graphs, they are typically proportional to edge weights. The ORC along edge  $(i, j)$  is defined as:

$$\kappa(i, j) = 1 - \frac{W_1(m_i, m_j)}{d(i, j)} \quad (1)$$

where  $W_1(\cdot, \cdot)$  is the Wasserstein-1 distance (also known as Earth Mover’s Distance), computed using the graph’s shortest-path metric, and  $d(i, j)$  denotes the distance between nodes  $i$  and  $j$  (typically 1 for adjacent nodes). Intuitively, if two nodes share highly overlapping neighborhoods,  $W_1(m_i, m_j)$  is small, and the curvature is positive; if their neighborhoods are far apart, the curvature becomes negative.

The standard definition of ORC applies to simple graphs, where at most one edge exists between any pair of nodes Ollivier (2009). However, knowledge graphs are typically modeled as multigraphs, in which multiple directed edges (corresponding to distinct relations) may connect the same pair of entities. To apply ORC in this context, we propose a two-step adaptation process that involves aggregating parallel edges and mapping the resulting curvature values back to the original multigraph. In particular, self-loop edges are excluded from standard ORC computation, since the ORC formula involves division by edge length and is therefore undefined in the zero-distance case.

*Edge Aggregation via Parallel Resistance Analogy.* To prepare the multigraph  $G_M = (V, E_M)$  for ORC



**Figure 3.** The overall architecture of ORCA-GCN. The model first computes Ollivier-Ricci Curvature for the input knowledge graph. The first layer performs Geometry-Aware Aggregation using ORC values to re-weight messages. The second layer transitions to Semantic-Aware Aggregation, using cosine dissimilarity between node embeddings. Final embeddings are passed to a scoring network for link prediction.

computation, we first convert it into a weighted simple digraph  $G_S = (V, E_S)$ . Our approach is inspired by the **parallel resistance analogy** from circuit theory [Alexander & Sadiku \(2017\)](#). We interpret edge weights as a form of **resistance**, where a higher value signifies a weaker connection. Accordingly, since edges in standard knowledge graphs are typically unweighted, we assign a default **unit resistance** of  $w_M(e_i) = 1$  to all non-loop edges.

Just as multiple resistors in parallel decrease the total resistance, multiple edges between two nodes signify a stronger effective connection. We formalize this by computing an **equivalent weight (resistance)** for the single edge in  $G_S$  that replaces the parallel edges in  $G_M$ . For a node pair  $(u, v)$  with  $k$  parallel edges, this is:

$$w_S(u, v) = \left( \sum_{i=1}^k \frac{1}{w_M(e_i)} \right)^{-1} \quad (2)$$

where  $w_M(e_i)$  is the resistance of an individual parallel edge. This method ensures the resulting graph  $G_S$  preserves the crucial distance-based properties required for accurate ORC computation. With our unit resistance assumption, the formula elegantly simplifies to  $w_S(u, v) = 1/k$ .

**Uniform Curvature Mapping.** Once the ORC values  $\kappa_S(u, v)$  are computed on the aggregated graph  $G_S$ , we propagate them back to the corresponding edges in  $G_M$  using a uniform assignment scheme. This choice is motivated by the fact that ORC quantifies the geometric coupling between the neighborhoods of two nodes, independent of the number of edges connecting them. Thus, for each original edge  $e \in$

$E_{uv}(M)$ , the mapped curvature is:

$$\kappa_M(e) = \kappa_S(u, v), \quad \forall e \in E_{uv}(M) \quad (3)$$

This method guarantees that all parallel edges, which represent a single underlying structural link, share a unified geometric interpretation.

**Curvature Assignment for Self-Loops.** The final adaptation concerns self-loops that naturally occur in the original knowledge graph  $G_M$ , such as self-referential triples  $(e, r, e)$ . These should be distinguished from the artificial self-loops introduced in many GNN architectures solely to incorporate a node’s own features during message aggregation. For naturally occurring self-loops in a KG, the ORC formula is undefined because the distance  $d(i, i) = 0$  appears in the denominator. Following **Curvature Graph Neural Networks (CGNN)** [Li et al. \(2022\)](#), we assign a fixed curvature value of  $+1.0$  to all such edges:

$$\kappa_M(e_{ii}) = 1.0 \quad \text{for any self-loop edge } e_{ii} \quad (4)$$

Other conventions exist; for example, **CurvGN** sets self-loop curvature to 0, analogous to an edge in a flat grid [Ye et al. \(2019\)](#). We adopt the  $+1.0$  assignment to reflect the notion of QuatE, treating these self-referential relations as strong stabilizing forces in the graph’s geometry.

### 3.2 Geometry-Aware Aggregation Module

The first layer of ORCA-GCN employs a **Geometry-Aware Aggregation** module, designed to inject a strong structural inductive bias while node representations are

still developing. The core mechanism is to **modulate the influence of each neighbor** based on the Ollivier-Ricci Curvature of the connecting edge.

This is achieved in two steps. First, we compute normalized attention scores. For an edge  $(j \rightarrow i)$ , we define a raw geometric weight,  $w_{ji}^{(1)} = 1 - \kappa(j, i)$ , which directly implements our central hypothesis: it amplifies the influence of low-curvature bridges (where  $\kappa < 0$ ) and attenuates that of high-curvature intra-cluster links (where  $\kappa > 0$ ). These raw weights are then processed through a softmax operator over all of node  $i$ 's neighbors to produce normalized attention coefficients  $\alpha_{ji}^{(1)}$ :

$$\alpha_{ji}^{(1)} = \frac{\exp(w_{ji}^{(1)})}{\sum_{k \in \mathcal{N}(i)} \exp(w_{ki}^{(1)})} \quad (5)$$

Second, these attention scores are used to modulate the messages before aggregation. Each message  $\mathbf{m}_{ji}^{(1)}$  is constructed by applying the attention score  $\alpha_{ji}^{(1)}$  to the transformed composition of the source node and relation embeddings:

$$\mathbf{m}_{ji}^{(1)} = \alpha_{ji}^{(1)} \mathbf{W}_{\lambda(r_{ji})} \phi(h_j^{(0)}, r_{ji}) \quad (6)$$

where  $\phi(\cdot, \cdot)$  is a relational composition function and  $\mathbf{W}_{\lambda(r_{ji})}$  is a direction-specific learnable matrix. The final first-layer representation  $h_i^{(1)}$  is then obtained by aggregating all incoming modulated messages, along with a transformed self-loop message, and passing the result through a non-linear activation function  $\sigma(\cdot)$ :

$$h_i^{(1)} = \sigma \left( \mathbf{W}_S \phi(h_i^{(0)}, r_{loop}) + \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ji}^{(1)} \right) \quad (7)$$

This geometry-guided mechanism enables the model to selectively prioritize signals from structurally important neighbors, thereby building a more topologically-informed representation in its initial layer.

### 3.3 Semantic-Aware Aggregation Module

While the geometry-aware module leverages structural priors, it does not capture the semantic relationships embedded in the learned node representations. To address this, the second layer of ORCA-GCN employs a semantic-aware aggregation mechanism that modulates message passing based on the semantic dissimilarity between first-layer node embeddings. We adopt a non-parametric dissimilarity measure to directly evaluate the hypothesis that semantic divergence is an informative aggregation signal, while avoiding additional trainable parameters that could increase overfitting risk.

Given the first-layer representations  $h_j^{(1)}$  and  $h_i^{(1)}$  for a source node  $j$  and a target node  $i$ , we compute their semantic dissimilarity using the cosine distance:

$$d_{ji}^{\cos} = 1 - \cos(h_j^{(1)}, h_i^{(1)}) = 1 - \frac{h_j^{(1)} \cdot h_i^{(1)}}{\|h_j^{(1)}\| \cdot \|h_i^{(1)}\|} \quad (8)$$

This value quantifies how semantically distinct the two nodes are. As in the first layer, we convert dissimilarity scores into

normalized attention coefficients using a softmax over the neighborhood of each target node:

$$\alpha_{ji}^{(2)} = \frac{\exp(d_{ji}^{\cos})}{\sum_{k \in \mathcal{N}(i)} \exp(d_{ki}^{\cos})} \quad (9)$$

The semantic-aware message  $\mathbf{m}_{ji}^{(2)}$  is then obtained by weighting the composed message with the semantic attention score:

$$\mathbf{m}_{ji}^{(2)} = \alpha_{ji}^{(2)} \mathbf{W}_{\lambda(r_{ji})} \phi(h_j^{(1)}, r_{ji}) \quad (10)$$

The second-layer node update aggregates these messages together with a self-loop message:

$$h_i^{(2)} = \sigma \left( \mathbf{W}_S \phi(h_i^{(1)}, r_{loop}) + \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ji}^{(2)} \right) \quad (11)$$

By introducing semantic-aware aggregation in the second layer, ORCA-GCN transitions from reliance on structural priors to exploitation of learned semantic information, enabling a balanced integration of topological and semantic cues for improved link prediction performance.

## 4 Experiments

In this section, we evaluate ORCA-GCN on standard knowledge graph completion benchmarks to assess the effectiveness of integrating structural and semantic priors into link prediction. We focus on two key aspects: (i) How do the geometry-aware and semantic-aware modules, both independently and in synergy, contribute to ORCA-GCN's overall performance? (ii) Can Ollivier-Ricci Curvature effectively reveal structural properties like redundancy or bottlenecks in knowledge graphs, thereby providing insights for optimizing graph connectivity and message propagation? The remainder of this section is organized as follows: Section 4.1 describes the experimental setup, Section 4.2 presents the quantitative comparisons, Section 4.3 details the ablation studies, and Section 4.4 provides a curvature-based structural analysis.

### 4.1 Experimental Setup

**4.1.1 Datasets** We evaluate ORCA-GCN on two widely-used knowledge graph completion benchmark datasets: FB15k-237 and WN18RR. Both datasets are derived from larger knowledge bases and are commonly used for link prediction tasks.

- **FB15k-237:** A subset of Freebase [Bollacker et al. \(2008\)](#), FB15k-237 is characterized by its diverse set of entities and relations, making it a challenging benchmark for KGC. It contains a significant number of entities and relations, often exhibiting complex, multi-relational patterns, as well as dense clusters and critical bridge connections between them. The dataset is split into training, validation, and test sets.
- **WN18RR:** Derived from WordNet [Miller \(1995\)](#), WN18RR is a more sparse and hierarchical dataset compared to FB15k-237. It primarily captures lexical relations between words, resulting in a more tree-like structure. This distinct topological characteristic

**Table 1.** Model performance is reported on the FB15k-237 and WN18RR test sets. Best results are bolded. Results marked with † were reproduced using the provided code; Results marked with ‡ are as reported in Li et al. (2022); all other results are from their respective published papers.

| Models                                         | FB15k-237   |             |             |             | WN18RR      |             |             |             |
|------------------------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                                                | MRR         | H@1         | H@3         | H@10        | MRR         | H@1         | H@3         | H@10        |
| <b>Translational Distance</b>                  |             |             |             |             |             |             |             |             |
| TransE <sup>†</sup> Bordes et al. (2013)       | .330        | .231        | .369        | .528        | .223        | -           | .401        | .529        |
| RotatE Sun et al. (2019)                       | .338        | .241        | .375        | .533        | .476        | .428        | .492        | <b>.571</b> |
| <b>Semantic Matching</b>                       |             |             |             |             |             |             |             |             |
| DistMult <sup>†</sup> Yang et al. (2015)       | .308        | .219        | .336        | .485        | .439        | .394        | .452        | .533        |
| ComplEx <sup>†</sup> Trouillon et al. (2016)   | .323        | .229        | .353        | .513        | .468        | .427        | .485        | .554        |
| ConvE Dettmers et al. (2018)                   | .325        | .237        | .356        | .501        | .430        | .400        | .440        | .520        |
| TuckER Balazevic et al. (2019)                 | .358        | .266        | .394        | .544        | .470        | .443        | .482        | .526        |
| ProcrustEs Peng et al. (2021)                  | .345        | .249        | .379        | .541        | .474        | .421        | <b>.502</b> | .569        |
| <b>MLP-based</b>                               |             |             |             |             |             |             |             |             |
| MLP-DistMult Li et al. (2023)                  | .334        | .245        | .366        | .511        | .433        | .399        | .446        | .507        |
| MLP-ConvE Li et al. (2023)                     | .355        | .264        | .389        | .537        | .473        | .437        | .488        | .544        |
| <b>GNN-based</b>                               |             |             |             |             |             |             |             |             |
| R-GCN <sup>†</sup> Schlichtkrull et al. (2018) | .330        | .239        | .365        | .505        | .445        | .406        | .458        | .524        |
| SACN Shang et al. (2019)                       | .310        | .260        | .390        | .540        | .470        | .430        | .480        | .540        |
| KBGAT <sup>‡</sup> Nathani et al. (2019)       | .336        | .244        | .373        | .520        | .438        | .401        | .453        | .511        |
| CompGCN <sup>†</sup> Vashishth et al. (2020)   | .355        | .264        | .390        | .535        | .475        | .443        | .494        | .546        |
| <b>ORCA-GCN (ours)</b>                         | <b>.363</b> | <b>.272</b> | <b>.398</b> | <b>.544</b> | <b>.483</b> | <b>.446</b> | .498        | .544        |

provides a valuable contrast for evaluating the generalizability of our curvature-aware approach. The dataset is also partitioned into training, validation, and test sets.

For both datasets, we adhere to the standard evaluation protocol for link prediction, using filtered Mean Reciprocal Rank (MRR) and Hits@k (H@1, H@3, H@10) as our primary metrics. MRR is the average of the reciprocal ranks of all correct entities, while Hits@k measures the proportion of correct entities ranked within the top  $k$  candidates.

**4.1.2 Baselines** To comprehensively assess the performance of ORCA-GCN, we compare it against a wide range of strong baselines, categorized as follows:

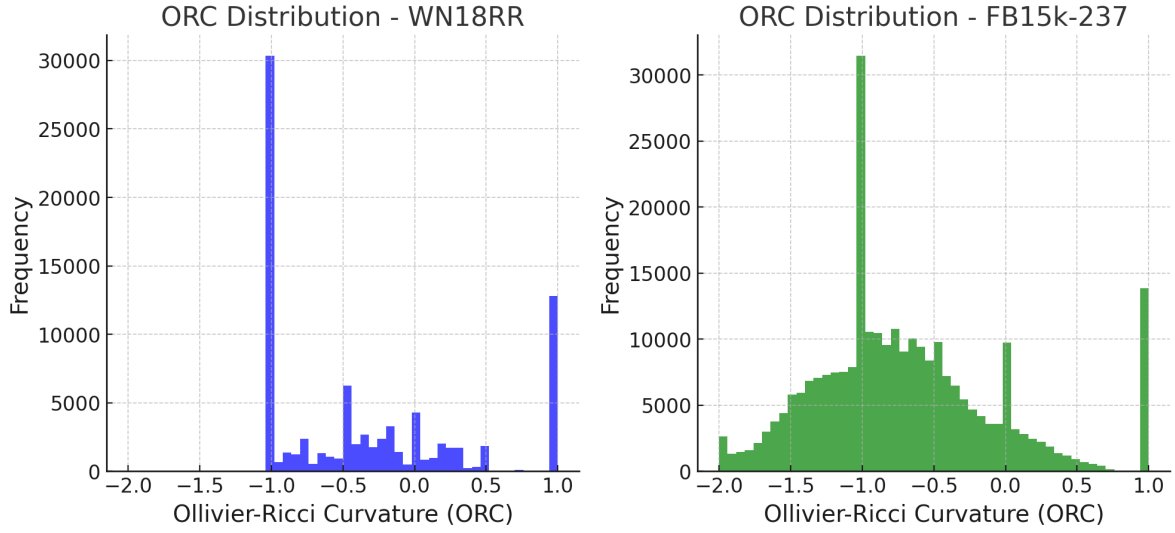
- **Translational Distance Models:** These models learn entity and relation embeddings such that relations are modeled as translations in the embedding space. We include TransE and RotatE.
- **Semantic Matching Models:** These approaches define a scoring function to measure the plausibility of a triple based on semantic matching. This category includes DistMult, ComplEx, TuckER, ConvE, InteractE, and ProcrustEs.
- **MLP-based Models:** Recent work has shown that simple Multi-Layer Perceptrons (MLPs) can achieve competitive results on KGC. We include **MLP-DistMult** and **MLP-ConvE** to provide a baseline that explicitly omits message passing, allowing for a clearer understanding of the benefits of graph-aware mechanisms.
- **GNN-based Models:** These models leverage message passing to aggregate information from neighboring nodes and relations. We compare against

prominent GNN architectures for KGC, including R-GCN, SACN, KBGAT and CompGCN. Our proposed **ORCA-GCN** is built upon the **CompGCN** architecture, augmenting its message-passing capabilities with curvature and semantic awareness.

## 4.2 Main Results

This section presents the primary experimental results of ORCA-GCN on the knowledge graph completion task, comparing its performance against a comprehensive set of baselines on the FB15k-237 and WN18RR datasets. The complete results, evaluated using standard link prediction metrics (Mean Reciprocal Rank (MRR), and Hits@k (H@1, H@3, H@10) as introduced in Section 4.1.1), are summarized in Table 1.

As shown in Table 1, ORCA-GCN consistently outperforms all strong baselines across both FB15k-237 and WN18RR datasets. These results represent a notable improvement over its base architecture, CompGCN Vashishth et al. (2020). This demonstrates the effectiveness of integrating Ollivier-Ricci Curvature and semantic awareness into the message-passing mechanism for enhancing knowledge graph completion. On the WN18RR dataset, ORCA-GCN also achieves the best overall performance. While the relative improvement over CompGCN is slightly less pronounced compared to FB15k-237, it still confirms the robustness and generalizability of ORCA-GCN across different knowledge graph structures. The differing degrees of improvement across datasets can be attributed to their distinct structural properties, which we will further analyze in Section 4.4.



**Figure 4.** Distribution of Ollivier-Ricci Curvature (ORC) on the WN18RR (left) and FB15k-237 (right) datasets. FB15k-237 shows a wide distribution with many negative-curvature edges, while WN18RR’s distribution is more polarized, reflecting their distinct topological properties.

### 4.3 Ablation Study

To comprehensively isolate and quantify the contribution of each component within ORCA-GCN, we conduct a detailed ablation study by evaluating three key model variants. Our ORCA-GCN is built upon the CompGCN architecture. The results are summarized in Table 2.

- **w/o Geometry-Aware Module:** In this variant, we disable the geometry-aware aggregation in the initial layer by replacing the curvature-based weights ( $1 - \kappa$ ) with uniform weighting for all incoming messages. This effectively reverts the first layer’s aggregation to a standard, geometry-agnostic approach, similar to a conventional GCN. This change leads to a noticeable degradation in performance on FB15k-237, highlighting the crucial role of Ollivier-Ricci Curvature (ORC) in suppressing redundant intra-cluster signals and amplifying informative bridge-like connections for KGC.
- **w/o Semantic-Aware Module:** For this variant, we remove the semantic-aware aggregation from the second layer, reverting its message passing to using uniform weights, similar to the first layer’s behavior in the “w/o Geometry-Aware Module” setting. The subsequent performance drop, indicates that leveraging embedding-level dissimilarity becomes increasingly important for refining representations in deeper layers, where initial structural cues may become less discriminative and semantic alignment becomes crucial.
- **w/ Random Curvature:** To rigorously verify that the performance gain arises from meaningful geometric information embedded in ORC, rather than a mere numerical perturbation, we replace the true ORC values with random noise. This noise is uniformly sampled from the observed range of ORC values in our datasets, approximately  $[-2, 1]$ . The resulting model exhibits significantly degraded performance, which underscores that ORC provides a potent and

indispensable topological signal for effective message aggregation.

Taken together, these findings robustly validate the complementary and distinct roles of our structural (geometry-aware) and semantic (semantic-aware) priors in creating a more effective message aggregation framework. The degradation observed in each ablated scenario confirms the necessity and efficacy of each proposed module within the ORCA-GCN architecture, particularly for the challenging task of Knowledge Graph Completion.

**Table 2.** This table evaluates the contribution of ORCA-GCN’s components. “w/o” indicates the removal of a specific module, while “w/” signifies the inclusion of an alternative component or setting.

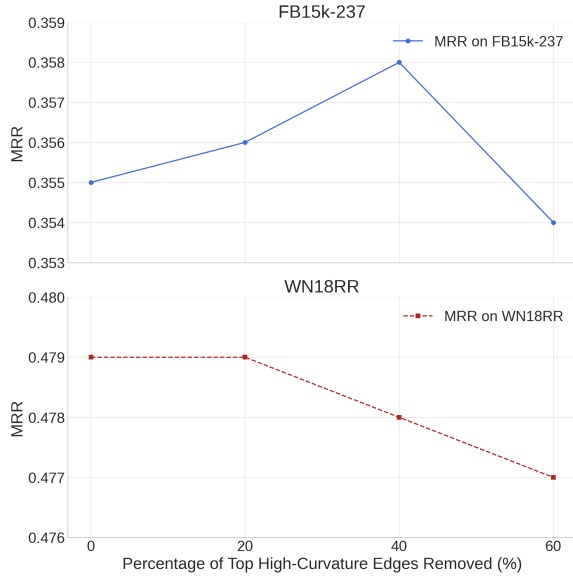
| Models              | FB15k-237   |             |             |             |
|---------------------|-------------|-------------|-------------|-------------|
|                     | MRR         | H@1         | H@3         | H@10        |
| <b>ORCA-GCN</b>     | <b>.363</b> | <b>.272</b> | <b>.398</b> | <b>.544</b> |
| w/o Geometry-Aware  | .359        | .268        | .395        | .541        |
| w/o Semantic-Aware  | .360        | .269        | .392        | .543        |
| w/ Random Curvature | .358        | .266        | .394        | .540        |

### 4.4 Structural Analysis

Beyond performance metrics, we conduct a **structural analysis** of FB15k-237 and WN18RR to elucidate the differing effectiveness of ORCA-GCN. As shown in Figure 4, the Ollivier-Ricci Curvature (ORC) distributions reveal distinct topological patterns in the two datasets.

FB15k-237 exhibits a broad curvature spectrum, with a pronounced concentration of negatively curved edges—indicative of bridge connections linking densely interconnected clusters. This supports the view that FB15k-237 comprises tightly knit communities connected by structural bridges. By down-weighting redundant intra-cluster edges and amplifying such bridges, ORCA-GCN facilitates richer long-range message propagation across communities.

In contrast, WN18RR displays a highly polarized curvature distribution, with most edges clustered at  $\kappa = +1$  (including self-loops) or  $\kappa = -1$ , and very few in the intermediate range. This pattern reflects a sparse, hierarchical structure aligned with WordNet’s lexical taxonomy. In such tree-like graphs, neighbor connectivity is more uniform and depth-oriented, offering fewer exploitable bridges for ORC-based reweighting.



**Figure 5.** MRR performance on FB15k-237 and WN18RR when removing the top- $k$ % of edges with the highest curvature. On FB15k-237, removing up to 40% of high-curvature edges slightly improves performance, highlighting their redundancy. Performance on the WN18RR is less affected.

Edge-pruning experiments further corroborate these insights. Figure 5 and Table 3 show that, for FB15k-237, removing up to 40% of the highest-curvature edges—primarily intra-cluster links—slightly improves performance, suggesting that excessive local redundancy can oversaturate aggregation and obscure informative signals. Pruning reduces this redundancy, enabling the model to focus on structurally critical long-range connections. In contrast, pruning in WN18RR yields negligible impact at all levels, consistent with its sparse structure and lack of redundant intra-cluster connectivity.

Overall, these results demonstrate that ORC-based edge weighting is most beneficial in clustered, bridge-rich graphs such as FB15k-237, where differentiating between redundant local links and structurally vital long-range edges enhances message propagation. Conversely, in tree-like, sparsely connected graphs such as WN18RR, where structural pathways are already streamlined, curvature-guided strategies provide limited additional gains.

These findings motivate a broader discussion of the method’s practical implications, limitations, and potential extensions, as outlined in the conclusion.

## 5. Conclusion

In this paper, we proposed **ORCA-GCN** (Ollivier–Ricci Curvature–Aware Graph Convolutional Network), a geometry-aware GNN designed to address the “geometric

**Table 3.** Impact of removing the top- $k$ % edges (ranked by ORC value) during the aggregation stage in CompGCN. Performance is reported on FB15k-237 and WN18RR.

| $k$              | MRR   | Hits@1 | Hits@3 | Hits@10 |
|------------------|-------|--------|--------|---------|
| <b>FB15k-237</b> |       |        |        |         |
| 0                | 0.355 | 0.264  | 0.390  | 0.535   |
| 20               | 0.356 | 0.268  | 0.388  | 0.534   |
| 40               | 0.358 | 0.268  | 0.392  | 0.539   |
| 60               | 0.354 | 0.264  | 0.387  | 0.534   |
| <b>WN18RR</b>    |       |        |        |         |
| 0                | 0.479 | 0.443  | 0.494  | 0.546   |
| 20               | 0.479 | 0.440  | 0.492  | 0.553   |
| 40               | 0.478 | 0.438  | 0.494  | 0.554   |
| 60               | 0.477 | 0.435  | 0.494  | 0.551   |

**blindness”** of conventional message passing in Knowledge Graph Completion (KGC). By integrating Ollivier–Ricci Curvature (ORC) into a layer-wise evolutionary aggregation framework, ORCA-GCN selectively attenuates redundant high-curvature intra-cluster edges and amplifies informative low-curvature bridges. This design enables more effective long-range reasoning, particularly in clustered graphs where structural bottlenecks hinder message flow. Experiments on FB15k-237 and WN18RR demonstrate consistent improvements over strong baselines, validating the benefit of combining structural priors with semantic dissimilarity signals.

**Limitations.** ORCA-GCN requires computing ORC values for all edges, which can be computationally expensive for very large-scale graphs. Although parallel computation and sampling strategies can mitigate this, the method may still face scalability challenges in web-scale knowledge bases. Moreover, the current work focuses on static graphs; handling dynamic or streaming RDF graphs remains an open issue.

**Future directions.** Extending ORCA-GCN to multi-modal knowledge graphs that integrate textual, visual, or geospatial data is a promising avenue. Another direction is combining curvature-aware aggregation with symbolic reasoning, enabling tighter integration of statistical learning and logical inference in semantic web systems. Finally, evaluating the model on large, real-world semantic web datasets such as DBpedia and YAGO would further confirm its applicability and benefit to the community.

In summary, this work highlights the value of incorporating discrete geometric signals into KGC, offering a principled way to enhance semantic web knowledge bases through improved structural and semantic sensitivity in graph neural networks.

## Acknowledgements

This class file was developed by Sunrise Setting Ltd, Brixham, Devon, UK.

Website: <http://www.sunrise-setting.co.uk>

## References

- Xiong, C., Power, R., & Callan, J. (2017, April). Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th International Conference on World*

- Wide Web* (pp. 1271–1279). International World Wide Web Conferences Steering Committee.
- Wang, H., Zhao, M., Xie, X., Li, W., & Guo, M. (2019, May). Knowledge graph convolutional networks for recommender systems. In *The World Wide Web Conference* (pp. 3307–3313). Association for Computing Machinery.
- West, R., Gabrilovich, E., Murphy, K., Sun, S., Gupta, R., & Lin, D. (2014, April). Knowledge base completion via search-based question answering. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 515–526). Association for Computing Machinery.
- Li, Q., Han, Z., & Wu, X. M. (2018, April). Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).
- Alon, U., & Yahav, E. (2021, January). On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations (ICLR)*. OpenReview.
- Li, J., Shomer, H., Ding, J., Wang, Y., Ma, Y., Shah, N., Tang, J., & Yin, D. (2023, July). Are Message Passing Neural Networks Really Helpful for Knowledge Graph Completion?. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 10696–10711). Association for Computational Linguistics.
- Li, H., Cao, J., Zhu, J., Liu, Y., Zhu, Q., & Wu, G. (2022). Curvature graph neural network. *Information Sciences*, 592, 50–66. <https://doi.org/10.1016/j.ins.2021.12.077>
- Liu, Y., Zhou, C., Pan, S., Wu, J., Li, Z., Chen, H., & Zhang, P. (2023, April). CurvDrop: A Ricci curvature based approach to prevent graph neural networks from over-smoothing and over-squashing. In *Proceedings of the ACM Web Conference 2023* (pp. 221–230). Association for Computing Machinery.
- Ye, Z., Liu, K. S., Ma, T., Gao, J., & Chen, C. (2019, September). Curvature graph network. In *International Conference on Learning Representations (ICLR)*.
- Guo, D., Su, M., Cao, C., Yuan, F., Zhang, X., Liu, Y., & Fu, J. (2023, May). Curvature-driven knowledge graph embedding for link prediction. In *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 1226–1231). IEEE.
- Nguyen, K., Hieu, N. M., Nguyen, V. D., Ho, N., Osher, S., & Nguyen, T. M. (2023, July). Revisiting over-smoothing and over-squashing using Ollivier-Ricci curvature. In *International Conference on Machine Learning (ICML)* (pp. 25956–25979). Proceedings of Machine Learning Research (PMLR).
- Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems (NeurIPS)* (pp. 2787–2795).
- Zhang, S., Tay, Y., Yao, L., & Liu, Q. (2019). Quaternion knowledge graph embeddings. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)* (pp. 8659–8666). Association for the Advancement of Artificial Intelligence.
- Chao, L., He, J., Wang, T., & Chu, W. (2021, August). PairRE: Knowledge Graph Embeddings via Paired Relation Vectors. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on NLP (ACL-IJCNLP 2021)* (Volume 1: Long Papers, pp. 4360–4369). Association for Computational Linguistics.
- Yang, B., Yih, W.-t., He, X., Gao, J., & Deng, L. (2015). Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, co-located with ICLR 2015.
- Trouillon, T., Welbl, J., Riedel, S., Gaussier, E., & Bouchard, G. (2016). Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*. JMLR Workshop and Conference Proceedings.
- Balazevic, I., Allen, C., & Hospedales, T. M. (2019). TuckER: Tensor Factorization for Knowledge Graph Completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP) or ICML Adaptive Multitask Learning Workshop*.
- Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2D Knowledge Graph Embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*.
- Vashishth, S., Sanyal, S., Nitin, V., Agrawal, N., & Talukdar, P. (2020). InteractE: Improving Convolution-based Knowledge Graph Embeddings by Increasing Feature Interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2020)*.
- Peng, X., Chen, G., Lin, C., & Stevenson, M. (2021, June). Highly Efficient Knowledge Graph Embedding Learning with Orthogonal Procrustes Analysis (Procrustes). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 2364–2375). Association for Computational Linguistics.
- Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling relational data with graph convolutional networks. In *Proceedings of the European Semantic Web Conference (ESWC)*.
- Shang, C., Tang, Y., Huang, J., Bi, J., He, X., & Zhou, B. (2019). End-to-End Structure-Aware Convolutional Networks for Knowledge Base Completion. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1712–1722).
- Nathani, D., Chauhan, J., Sharma, C., & Kaul, M. (2019, July). Learning attention-based embeddings for relation prediction in knowledge graphs. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 4710–4723).
- Vashishth, S., Sanyal, S., Nitin, V., & Talukdar, P. (2020, April). Composition-Based Multi-Relational Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.
- Sun, Z., Deng, Z.-H., Nie, J.-Y., & Tang, J. (2019). RotatE: Knowledge graph embedding by relational rotation in complex space. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*.
- do Carmo, M. P. (1992). *Riemannian Geometry*. Birkhäuser.
- Ollivier, Y. (2009). Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3), 810–864.
- Alexander, C. K., & Sadiku, M. N. O. (2017). *Fundamentals of Electric Circuits* (6th ed.). McGraw-Hill Education.
- Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008, June). Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM*

- SIGMOD International Conference on Management of Data* (pp. 1247–1250).
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39–41.
- Li, R., Cao, Y., Zhu, Q., Bi, G., Fang, F., Liu, Y., & Li, Q. (2022, June). How does knowledge graph embedding extrapolate to unseen data: a semantic evidence view. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 5, pp. 5781–5791). AAAI Press.