LEARNING TRANSFERABLE TOPOLOGICAL META-KNOWLEDGE VIA SUBGRAPH EMBEDDING FOR INDUCTIVE LINK PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Inductive link prediction over knowledge graphs aims to generalize relational reasoning to entities that are entirely unseen during training. Existing methods alleviate the identity-binding limitation of conventional knowledge graph models, but remain challenged by cross-graph structural variation and the propagation of query-irrelevant information. To address these issues, we propose TOME (TOpological Meta-knowledge Extractor), a topology-based framework for inductive link prediction. TOME couples two complementary mechanisms: episodic meta-training over localized subgraphs, which aligns training with entity-disjoint graph transfer, and query-conditioned in-situ message passing, which incorporates the query relation into both message formulation and attention before neighborhood aggregation. This coupling enables TOME to learn transferable structural reasoning while suppressing irrelevant signals during message construction, without relying on target-entity-specific embeddings. Extensive experiments on WN18RR, FB15k-237, and NELL-995 demonstrate that TOME achieves strong overall performance in both inductive and transductive link prediction, outperforming topology-based baselines on most inductive and all transductive dataset–metric combinations. Ablation studies show that episodic meta-training and query-conditioned message passing make complementary contributions, with message-level query conditioning providing a substantial improvement. Further sparsity analyses support the robustness of TOME, while a qualitative case study illustrates its interpretable structural reasoning. These results demonstrate that coupling transfer-oriented meta-training with query-conditioned in-situ reasoning is an effective approach to entity-independent link prediction over unseen knowledge graphs.

1 INTRODUCTION

Knowledge graphs (KGs) represent real-world knowledge as multi-relational graphs, where facts are encoded as triples of entities and relations Krötzsch et al. (2018); Hogan et al. (2021). Knowledge graph link prediction estimates missing facts and ranks candidate entities for incomplete queries such as $(e_h, r, ?)$ or $(?, r, e_t)$, and is fundamental to knowledge graph completion Rossi et al. (2022); Chen et al. (2020). However, real-world KGs are incomplete and continuously evolving, with new entities and facts emerging over time Ji et al. (2021); Wang et al. (2017). This setting makes it important for a model to generalize beyond the entities observed during training.

Most conventional link prediction methods are developed under the transductive setting, where test entities are also observed during training Nickel et al. (2011). KGE and GNN-based models learn or propagate entity-specific representations, which creates an identity-binding bottleneck: when a target KG contains unseen entities, their representations are unavailable. These models therefore require retraining or additional adaptation, even when the source and target graphs share similar relational structures. Inductive link prediction addresses this limitation by requiring generalization to entirely unseen entities Chen et al. (2023); Galkin et al. (2024a). Methods based on auxiliary information, such as entity descriptions Shah et al. (2019), can construct representations for new entities, but such information is often unavailable or inconsistent. This motivates purely topological approaches that reason from relational structures without relying on entity identities or external attributes.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

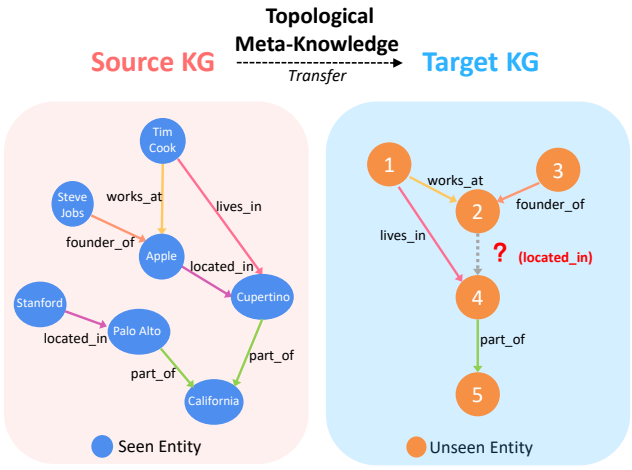

Figure 1: Inductive link prediction via topological meta-knowledge transfer. Unlike traditional KGEs that depend on entity-specific embeddings and cannot directly represent unseen entities (right), TOME infers the missing *located_in* relation by matching the neighborhood patterns of unseen nodes with analogous topological patterns in the source KG (left).

Purely topological methods reason from paths, rules, or localized subgraphs. Path- and rule-based approaches are interpretable but provide limited coverage of complex neighborhoods Xiong et al. (2017); Das et al. (2018); Meilicke et al. (2018; 2020). Subgraph models, including GraIL Teru et al. (2020), NBFNet Zhu et al. (2021), and RED-GNN Zhang & Yao (2022), improve structural coverage, while ULTRA Galkin et al. (2024b) and InGram Lee et al. (2023) explore relation-aware topological representations. Yet two challenges remain: source-only training is sensitive to structural shifts, and generic messages may propagate irrelevant information before filtering. AdaProp Zhang et al. (2023) and Context Pooling (CP) Su et al. (2025) reduce such propagation through adaptive exploration or query-specific pooling, but introduce additional selection operations.

We propose TOME (TOpological Meta-knowledge Extractor), a framework that jointly learns transferable structural reasoning and query-relevant propagation. TOME constructs re-indexed episodic tasks from localized source subgraphs, exposing the model to entity-independent transfer during training. Within each task, the query relation is injected into message formulation and attention before aggregation, allowing the model to determine which structural evidence should be propagated for the current query. The two mechanisms are deliberately coupled: episodic training learns a transferable propagation function, while query conditioning determines the evidence that this function should transfer. Thus, TOME does not merely combine independent modules; it aligns the training objective and inference operator around query-relevant topological reasoning.

In summary, our main contributions are:

- **Coupled Transferable and Query-Relevant Reasoning.** We propose TOME, which integrates episodic meta-training with query-conditioned in-situ message passing to transfer topological reasoning across entity-disjoint graphs.
- **Pre-Aggregation Query Conditioning.** We inject the query relation into both message construction and attention before aggregation, suppressing irrelevant structural branches at the propagation stage rather than filtering them only after message formation.
- **Comprehensive Evaluation.** Experiments on WN18RR, FB15k-237, and NELL-995 demonstrate TOME's effectiveness across 12 inductive splits and in the transductive setting. Ablation and sparsity analyses examine contributions and robustness of the coupled design.

2 Related Work

We review prior work on inductive link prediction from four perspectives: transductive GNNs, topology-based reasoning, relation-aware representation learning, and meta-learning.

Transductive GNN-based Methods. R-GCN Schlichtkrull et al. (2018), CompGCN Vashishth et al. (2020), and KE-GCN Yu et al. (2021) use relation-aware neighborhood aggregation to model multi-relational graphs. However, these models learn representations tied to entity identities observed during training, which limits their ability to generalize to graphs containing unseen entities.

Topology-based Methods. Path-based methods such as DeepPath Xiong et al. (2017), MINERVA Das et al. (2018), and MultiHopKG Lin et al. (2018) learn multi-hop reasoning paths, while rule-based methods such as RuleN Meilicke et al. (2018), AnyBURL Meilicke et al. (2020), Neural-LP Yang et al. (2017), and DRUM Sadeghian et al. (2019) extract symbolic or differentiable relational rules. Although interpretable, these methods may have limited coverage of complex local structures and can be sensitive to graph sparsity. Subgraph-based methods address this limitation by encoding query-centered neighborhoods. GraIL Teru et al. (2020), NBFNet Zhu et al. (2021), and RED-GNN Zhang & Yao (2022) learn identity-independent representations through localized relational propagation. AdaProp Zhang et al. (2023) reduces irrelevant exploration through query-guided adaptive propagation, whereas Context Pooling Su et al. (2025) selects query-specific neighborhoods after candidate representations are produced. These methods improve the selectivity of propagation, but they mainly control which nodes or paths are explored or retained, rather than conditioning message content before aggregation.

Relation-aware Inductive Representation Methods. ULTRA Galkin et al. (2024b) learns relation-aware representations through a pretrained graph reasoning model, while InGram Lee et al. (2023) models dependencies among relations using a relation graph. These methods provide important alternatives for inductive link prediction, but their predictions primarily rely on relation-level representation learning or embedding generation. In contrast, TOME directly learns an entity-independent scorer through query-centered topological message passing.

Meta-Learning-based Methods. MetaR Chen et al. (2019) transfers relation-specific knowledge for few-shot link prediction, while MorsE Chen et al. (2022b) and MaKEr Chen et al. (2022a) extend meta-learning to unseen entities or elements. MorsE encodes local topology into structural priors for initializing entity representations, and MaKEr aggregates contextual evidence to construct representations for newly emerged elements. However, these approaches primarily use meta-learning to initialize representations, while the final prediction is performed by a static KGE scoring function. Thus, the meta-learning objective is not directly coupled with the subsequent topological message-passing process.

Motivation for TOME. Existing methods have separately investigated local subgraph reasoning, adaptive propagation, relation-aware representation learning, and meta-learning-based transfer. TOME bridges these directions by performing episodic training on re-indexed subgraph tasks while injecting the query relation into message construction before aggregation. This coupling enables the model to learn transferable topological reasoning and suppress irrelevant evidence during propagation, yielding an entity-independent scorer for unseen entities.

3 PRELIMINARIES

We briefly define knowledge graphs, inductive link prediction, and the episodic tasks used by TOME.

3.1 BASIC DEFINITIONS

A knowledge graph is defined as $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$, where $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{T}$ denote the sets of entities, relations, and factual triplets, respectively. Given an incomplete query $(e_h, r, ?)$ or $(?, r, e_t)$, link prediction ranks candidate entities according to the plausibility of the resulting triplets.

A relational digraph (r-digraph) is a hierarchical directed s-t graph anchored by the query entity e_q and a candidate answer e_a. Its edges are directed from layer ℓ to layer $\ell + 1$, and entities within the same layer are distinct. Thus, each directed path from e_q to e_a represents a relational path in the KG.

3.2 INDUCTIVE LINK PREDICTION

Given a training KG $\mathcal{G}_{\text{train}}$ and a testing KG $\mathcal{G}_{\text{test}}$, where $\mathcal{E}_{\text{train}} \cap \mathcal{E}_{\text{test}} = \emptyset$ and $\mathcal{R}_{\text{test}} \subseteq \mathcal{R}_{\text{train}}$, inductive link prediction learns a scoring function $\text{score}(h, r, t)$ from $\mathcal{G}_{\text{train}}$ and applies it to unseen entities in

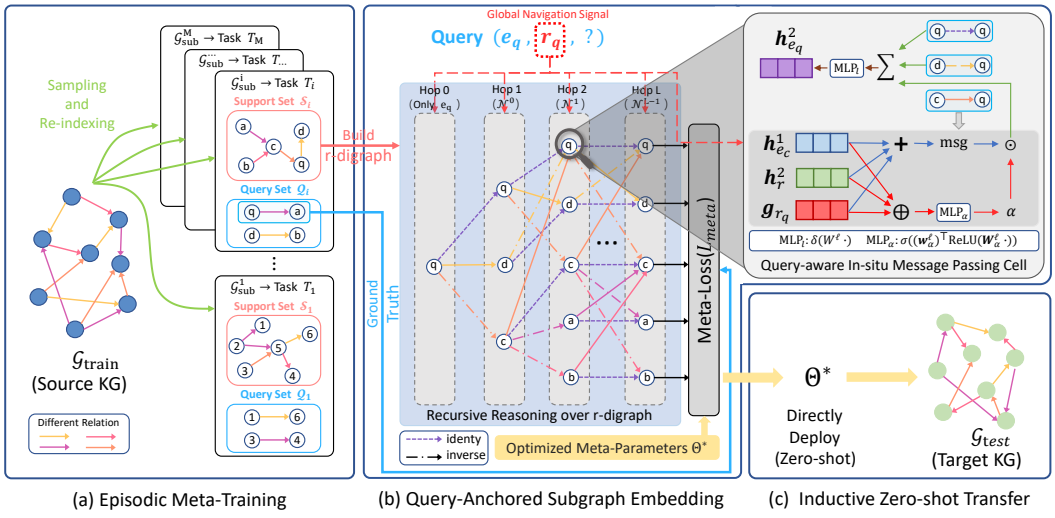

Figure 2: The architecture of TOME. (a) Episodic Meta-Training: Meta-tasks are generated via sampling and re-indexing. The support set builds the r-digraph, while the query set provides ground truth to compute the meta-loss $\mathcal{L}_{\text{meta}}$. (b) Query-Anchored Subgraph Embedding: Driven by the query-aware in-situ message passing cell, the query relation acts as a global navigation signal to suppress topological noise at the formulation stage and perform recursive reasoning. (c) Inductive Zero-Shot Transfer: The optimized meta-parameters Θ^* are directly deployed to the target KG for zero-shot prediction without retraining.

$\mathcal{G}_{\text{test}}$. Unlike transductive models, the scoring function cannot rely on entity-specific embeddings and must instead exploit transferable topological patterns.

3.3 Episodic Meta-Training Task for KGs

Given $\mathcal{G}_{\text{train}}$, we construct independent meta-tasks $\mathfrak{T}_{\text{train}} = \{T_1, \ldots, T_M\}$, where each task $T_i = (\mathcal{S}_i, \mathcal{Q}_i)$ contains a support set $\mathcal{S}_i$ describing a local topology and a query set $\mathcal{Q}_i$ containing incomplete triplets. The model conditions its predictions on the support graph and is optimized across tasks:

$$\mathcal{L}_{\text{meta}} = \sum_{T_i \in \mathfrak{T}_{\text{train}}} \sum_{(h,r,t) \in \mathcal{Q}_i} \ell\big(f_\theta(h, r, t \mid \mathcal{S}_i)\big), \tag{1}$$

where $f_\theta(h, r, t \mid \mathcal{S}_i)$ denotes the plausibility score conditioned on the support graph. Re-indexing entities within each task prevents the model from relying on source-specific identities and encourages transferable topological reasoning.

4 Methodology

We present TOME, a framework for entity-independent inductive link prediction. TOME couples episodic meta-training with query-aware in-situ message passing to learn transferable topological reasoning and suppress irrelevant evidence during propagation. The framework consists of three stages: episodic meta-training, query-anchored subgraph embedding, and inductive zero-shot transfer. We first describe the overall framework, followed by the query-anchored embedding process, meta-training objective, and zero-shot deployment.

4.1 Framework Overview

TOME learns transferable topological reasoning through three coupled stages (Figure 2).

Episodic Meta-Training. Instead of training on the source graph, TOME samples and re-indexes localized subgraphs to form meta-tasks (Figure 2a). Each task includes a support set for the local

r-digraph and a query set for computing the meta-loss $\mathcal{L}_{\text{meta}}$. Re-indexing removes entity identities and promotes transferable structural patterns.

Query-Anchored Subgraph Embedding. TOME recursively encodes the query-anchored r-digraph (Figure 2b). It injects the query relation r_q into message formulation and attention, propagating relevant evidence while suppressing irrelevant paths before aggregation.

Inductive Zero-Shot Transfer. After meta-training, the optimized parameters Θ^* are directly applied to the unseen target KG (Figure 2c). Because TOME relies on relation semantics and local topology rather than entity identities, it predicts links for unseen entities without retraining or fine-tuning.

4.2 Query-Anchored Subgraph Embedding

Conventional subgraph methods (e.g., GraIL Teru et al. (2020)) typically extract broad neighborhoods before learning representations, making iterative evaluation across all candidate entities computationally prohibitive. To address this, inspired by RED-GNN Zhang & Yao (2022), we adopt a recursive subgraph encoding strategy that computes embeddings for all candidate entities simultaneously. Building upon this foundation, we introduce query-aware in-situ message passing to structurally guide the encoding process.

4.2.1 Recursive Subgraph Encoding

Given a query $(e_q, r_q, ?)$, potential answers are naturally constrained within the L-hop neighborhood of e_q. Instead of building independent subgraphs, we construct a unified query-anchored subgraph. Entity states are derived through layer-wise updates. We define $\mathbf{h}^{\ell}_{e_o}(e_q, r_q)$ as the representation of entity e_o at layer ℓ conditioned on the query. This explicit query-dependence captures topological evidence along all length-ℓ relational paths from e_q to e_o. For brevity, we denote it as $\mathbf{h}^{\ell}_{e_o}$.

4.2.2 Query-Aware In-Situ Message Passing

To overcome topological noise and prevent the combinatorial explosion of paths during the query-anchored encoding, we design a query-aware in-situ message passing cell. Rather than executing late-stage filtering on generic messages, this mechanism treats the query relation as a global navigation signal, aligning every message with the target semantics upon formulation.

First, the message passing mechanism ensures that propagated evidence is structurally conditioned on the query. We directly inject the query embedding $\mathbf{g}_{r_q}$ into the initial message formulation. The update at layer ℓ is defined as:

$$\mathbf{h}^{\ell}_{e_o} = \delta\Bigg(\mathbf{W}^{\ell} \sum_{(e_s,r,e_o)\in\mathcal{T}^{\ell}} \alpha^{\ell}_{e_s,r,e_o|r_q}\big(\mathbf{h}^{\ell-1}_{e_s} + \mathbf{h}^{\ell}_{r} + \mathbf{g}_{r_q}\big)\Bigg), \tag{2}$$

where δ is the activation function, d is the base embedding dimension, $\mathbf{W}^{\ell} \in \mathbb{R}^{d\times d}$ is the layer-specific weight matrix, and $\mathbf{h}^{\ell}_{r}$ is the edge relation embedding. Injecting $\mathbf{g}_{r_q}$ directly into the summation conditions each formulated message on the query relation, encouraging the propagation of query-relevant evidence. Consequently, entity states evolve exclusively within a query-dependent semantic space, avoiding the blind aggregation of noisy local topologies.

Second, integrated into this message-passing process, we formulate a query-aware attention mechanism for fine-grained routing. The attention weight $\alpha^{\ell}_{e_s,r,e_o|r_q}$ is computed by:

$$\alpha^{\ell}_{e_s,r,e_o|r_q} = \sigma\Big((\mathbf{w}^{\ell}_{\alpha})^{\top}\mathrm{ReLU}\big(\mathbf{W}^{\ell}_{\alpha}(\mathbf{h}^{\ell-1}_{e_s} \oplus \mathbf{h}^{\ell}_{r} \oplus \mathbf{g}_{r_q})\big)\Big), \tag{3}$$

where σ is the Sigmoid activation function. The weight matrix $\mathbf{W}^{\ell}_{\alpha} \in \mathbb{R}^{d_a\times 3d}$ transforms the concatenated features into an attention hidden space of dimension d_a, $\mathbf{w}^{\ell}_{\alpha} \in \mathbb{R}^{d_a}$ projects this hidden representation into a scalar score, and $\oplus$ denotes concatenation. Since the messages are already query-aligned, this attention mechanism acts as a precise filter. It can assign lower weights to edges that are less relevant to the current query, achieving early noise suppression at the formulation stage.

The complete recursive procedure is provided in Algorithm 1 in Appendix A. Anchored at e_q, the model expands its receptive field layer by layer. In step ℓ, $\mathcal{T}^{\ell}$ denotes the set of active triplets originating from reached entities. $\mathcal{N}^{\ell}$ denotes the active entities discovered at the current hop.

Specifically, message passing is restricted to active triplets $\mathcal{T}^\ell$. Attention routing then suppresses irrelevant paths, and the model updates active entity states via target-aligned messages. Entities unreachable within L hops receive zero vectors. This strategy confines computation to the relevant local r-digraph, accelerating inductive reasoning and improving robustness against topological noise.

After L layers of recursive aggregation, the final representation $\mathbf{h}_{e_o}^L$ encapsulates refined topological evidence. To evaluate candidate e_a, we apply a linear scoring function:

$$f(e_q, r_q, e_a) = \mathbf{w}^\top \mathbf{h}_{e_a}^L, \tag{4}$$

where $\mathbf{w} \in \mathbb{R}^d$ is a learnable weight vector. This score quantifies the topological compatibility between the query $(e_q, r_q, ?)$ and candidate e_a. These scores and representations are subsequently utilized within the episodic meta-training framework (Section 4.3) for task-level optimization.

4.3 Episodic Meta-Training

4.3.1 Meta-Training Setting

To enable generalization to unseen entities, we formulate the training process within an episodic meta-training paradigm. By confining the reasoning process to local contexts, we simulate the emergence of novel graphs. Each meta-task T_i is constructed from an independently sampled subgraph, rather than the monolithic source graph $\mathcal{G}_{\text{train}}$.

Task generation follows a sampling and local re-indexing pipeline. First, we sample $\mathcal{G}_{\text{train}}$ to extract a collection of subgraphs $\{\mathcal{G}_{\text{sub}}^1, \ldots, \mathcal{G}_{\text{sub}}^M\}$. Each subgraph $\mathcal{G}_{\text{sub}}^i$, consisting of entities $\mathcal{E}_{\text{sub}}^i$, relations $\mathcal{R}_{\text{sub}}^i$, and triplets $\mathcal{T}_{\text{sub}}^i$, represents a localized topological region. Second, we instantiate the meta-task $T_i = \{\mathcal{S}_i, \mathcal{Q}_i\}$ by partitioning the triplet set $\mathcal{T}_{\text{sub}}^i \subset \mathcal{T}_{\text{train}}$ into a support set $\mathcal{S}_i$ and a query set $\mathcal{Q}_i$. This partition satisfies:

$$\mathcal{S}_i \cup \mathcal{Q}_i = \mathcal{T}_{\text{sub}}^i, \quad \mathcal{S}_i \cap \mathcal{Q}_i = \emptyset. \tag{5}$$

The support set $\mathcal{S}_i$ constructs the r-digraph, providing the topological neighborhood required to extract the active triplet set for Algorithm 1. Concurrently, the query set $\mathcal{Q}_i$ provides the ground-truth target triplets to evaluate representations and compute the loss. Restricting data to $\mathcal{G}_{\text{sub}}^i$ prevents the model from exploiting global path shortcuts during inference.

To emulate the inductive setting, we apply a local re-indexing mechanism. For each task T_i, we map the global entity IDs in $\mathcal{E}_{\text{sub}}^i$ to anonymous indices. Discarding absolute identities forces the model to rely entirely on topological meta-knowledge and recursive patterns for robust reasoning.

4.3.2 Meta-Training Regime

During meta-training, the query-anchored subgraph encoder (Algorithm 1) operates on the support set $\mathcal{S}_i$ to generate state representations. The predictions for $\mathcal{Q}_i$ are compared against the ground truth to compute the task-specific meta-loss $\mathcal{L}_{\text{meta}}$. The objective is to minimize the accumulated loss across all M independently sampled tasks:

$$\min_{\Theta} \sum_{i=1}^{M} \mathcal{L}_{\text{meta}}(\mathcal{Q}_i) = \min_{\Theta} \sum_{i=1}^{M} \sum_{(e_q, r_q, e_a) \in \mathcal{Q}_i} \left(-f(e_q, r_q, e_a) + \log \sum_{e \in \mathcal{E}_{\text{sub}}^i} \exp(f(e_q, r_q, e)) \right), \tag{6}$$

where $\Theta = \{\{\mathbf{W}^\ell\}, \{\mathbf{w}_\alpha^\ell\}, \{\mathbf{W}_\alpha^\ell\}, \{\mathbf{h}_r^\ell\}, \{\mathbf{g}_r\}, \mathbf{w}\}$ denotes the learnable parameters defined in Section 4.2. The candidate entities e in the denominator are confined to the task set $\mathcal{E}_{\text{sub}}^i$.

Optimizing this objective forces the parameters to capture entity-independent topological rules from the source graph, yielding the optimized parameters Θ^*. Upon concluding meta-training, these parameters are deployed to the unseen target graph $\mathcal{G}_{\text{test}}$ for zero-shot inductive link prediction.

4.4 Inductive Zero-Shot Transfer

Applying the model to a target knowledge graph $\mathcal{G}_{\text{test}}$ initiates the inductive transfer phase. Unlike conventional methods that bind reasoning to fixed entity identities, our query-anchored encoding

relies on transferable topological patterns and relation semantics. Consequently, the optimized meta-parameters Θ^* learned during episodic meta-training are entity-agnostic.

This entity-independent property enables generalization to newly emerged graphs. To evaluate target queries, we directly deploy the meta-trained model to $\mathcal{G}_{\text{test}}$, completely obviating the need for retraining or fine-tuning on the new entities. Driven by the query-aware in-situ message passing mechanism, the encoder evaluates unseen entities using transferable topological patterns learned from the source graph. This inference paradigm performs inductive link prediction in a zero-shot manner, demonstrating robust topological transferability.

4.5 Inference Complexity

TOME performs query-anchored reasoning only on the active local neighborhood, avoiding full-graph propagation. The detailed complexity analysis is provided in Appendix B.

5 Experiments

Our experiments evaluate TOME from three perspectives: predictive performance, robustness under incomplete graphs, and the contributions of its key components. We compare TOME with representative baselines in inductive and transductive settings, and provide additional statistical tests, sparsity analyses, ablations, and a qualitative case study in the appendix.

5.1 Datasets

Following the standard inductive link prediction protocol Teru et al. (2020), we evaluate TOME on WN18RR Dettmers et al. (2018), FB15k-237 Toutanova et al. (2015), and NELL-995 Xiong et al. (2017). Each dataset contains four versions, V1–V4, with different graph scales and structural characteristics. Each version consists of a source KG and a target KG with disjoint entity sets. For inductive evaluation, models are trained on the source KG and tested on the corresponding target KG. Complete dataset statistics are provided in Appendix C.

5.2 Baselines

We compare TOME with representative rule-based methods (AnyBURL Meilicke et al. (2020), Neural-LP Yang et al. (2017), and DRUM Sadeghian et al. (2019)); subgraph-based GNNs (GraIL Teru et al. (2020), NBFNet Zhu et al. (2021), and RED-GNN Zhang & Yao (2022)); query-guided propagation and pooling methods (AdaProp Zhang et al. (2023), NBFNet+CP, and RED-GNN+CP Su et al. (2025)); and relation-aware methods (ULTRA Galkin et al. (2024b) and InGram Lee et al. (2023)). We also include MorsE Chen et al. (2022b), a meta-learning-based KGE method that initializes representations for unseen entities and uses KGE-style scoring. Additional baseline descriptions and result sources are provided in Appendix D.

5.3 Evaluation Metrics

We use the standard filtered ranking protocol, removing other known true triplets from the candidate set. We report MRR and Hits@1 for the main results, and additionally report Hits@5 and Hits@10 for the ablation study. Higher scores indicate better performance.

5.4 Implementation Details

We select hyperparameters by grid search using validation MRR. TOME is implemented in PyTorch and trained with Adam Kingma & Ba (2015) on a single NVIDIA RTX 3090 GPU for at most 50 epochs, with early stopping based on validation MRR (patience 5). The full hyperparameter search space is provided in Appendix E.

Table 1: Inductive link prediction results on WN18RR, FB15k-237, and NELL-995. Best and second-best results are **bolded** and underlined, respectively. Baselines marked with * are sourced from Su et al. (2025).

Models	WN18RR				FB15k-237				NELL-995			
	V1	V2	V3	V4	V1	V2	V3	V4	V1	V2	V3	V4
Metric: MRR												
AnyBURL*	0.391	0.548	0.340	0.550	0.355	0.437	0.338	0.318	0.477	0.431	0.391	0.332
Neural-LP*	0.649	0.635	0.361	0.628	0.325	0.389	0.400	0.396	0.610	0.361	0.367	0.261
DRUM*	0.666	0.646	0.380	0.627	0.333	0.395	0.402	0.410	0.628	0.365	0.375	0.273
GraIL*	0.627	0.625	0.323	0.553	0.279	0.276	0.251	0.227	0.481	0.297	0.322	0.262
NBFNet*	0.684	0.652	0.425	0.604	0.307	0.369	0.331	0.305	0.544	0.410	0.425	0.287
RED-GNN*	0.693	0.690	0.422	0.651	0.318	0.452	0.407	0.422	0.618	0.381	0.393	0.342
AdaProp	0.733	0.715	0.474	0.662	0.310	0.471	0.471	0.454	0.644	0.452	0.435	0.366
NBFNet+CP*	0.702	0.658	0.393	0.605	0.387	0.423	0.400	0.377	0.560	0.442	0.457	0.372
RED-GNN+CP*	0.708	0.694	0.425	0.648	0.383	0.463	0.447	0.427	0.634	0.430	0.426	0.382
ULTRA	0.685	0.679	0.411	0.614	0.509	0.524	0.504	0.496	**0.757**	0.575	0.563	**0.469**
InGram	0.311	0.253	0.134	0.228	0.201	0.212	0.188	0.151	0.604	0.232	0.265	0.197
MorsE	0.641	0.658	0.512	0.641	**0.611**	**0.751**	**0.738**	**0.745**	0.356	0.561	0.615	0.453
TOME (Ours)	**0.836**	**0.811**	**0.550**	**0.809**	0.572	0.591	0.579	0.591	0.695	**0.637**	**0.629**	0.328
Metric: Hits@1 (%)												
AnyBURL*	23.3	41.7	26.4	45.8	30.1	33.7	25.5	23.3	38.6	33.0	30.7	23.6
Neural-LP*	59.2	57.5	30.4	58.3	24.3	28.6	30.9	28.9	50.0	24.9	26.7	13.7
DRUM*	61.3	59.5	33.0	58.6	24.7	28.4	30.8	30.9	50.0	27.1	26.2	16.3
GraIL*	55.4	54.2	27.8	44.3	20.5	20.2	16.5	14.3	42.5	19.9	22.4	15.3
NBFNet*	59.2	57.5	30.4	57.4	19.0	22.9	20.6	18.5	50.0	27.1	26.2	23.3
RED-GNN*	64.6	63.3	36.4	60.6	24.7	35.6	32.2	32.7	50.0	28.8	30.0	23.8
AdaProp	66.8	64.2	39.6	61.1	19.1	37.2	37.7	35.3	52.2	34.4	33.7	24.7
NBFNet+CP*	62.2	57.8	32.9	53.7	28.5	28.3	26.9	24.4	52.0	31.2	35.0	25.7
RED-GNN+CP*	66.0	63.7	36.5	60.3	31.6	36.8	35.8	33.0	51.7	32.9	32.4	27.8
ULTRA	61.5	58.7	33.5	58.7	32.2	39.9	40.5	37.2	50.7	35.8	36.4	**28.8**
InGram	18.6	15.0	7.1	10.4	12.2	11.9	10.9	8.3	49.0	13.8	18.2	11.5
MorsE	57.5	59.2	42.1	57.5	**49.8**	**64.4**	**63.1**	**63.3**	27.3	41.6	47.0	27.8
TOME (Ours)	**78.7**	**75.5**	**46.5**	**76.0**	48.3	46.2	47.6	46.6	**55.7**	**52.5**	**50.7**	21.0

Table 2: Transductive link prediction results of the best baseline and TOME across 12 dataset splits. Each entry reports MRR / Hits@1 (%). Bold values indicate the better result for each metric.

Models	WN18RR				FB15k-237				NELL-995			
	V1	V2	V3	V4	V1	V2	V3	V4	V1	V2	V3	V4
Best Baseline	0.641 / 57.1	0.658 / 59.9	0.610 / 53.9	0.638 / 56.8	**0.767 / 66.6**	0.641 / 51.0	0.601 / 48.6	0.582 / 44.0	0.685 / 57.7	0.651 / 52.3	0.602 / 48.7	0.688 / 57.9
TOME	**0.755 / 70.3**	**0.773 / 72.8**	**0.769 / 72.4**	**0.786 / 74.0**	0.710 / 62.6	**0.671 / 57.4**	**0.659 / 55.1**	**0.679 / 56.9**	**0.750 / 65.6**	**0.730 / 62.8**	**0.678 / 57.0**	**0.720 / 59.0**

5.5 Main Results

5.5.1 Inductive Link Prediction

In the inductive setting, models are trained on source graphs and evaluated on corresponding target graphs with unseen entities. Table 1 reports MRR and Hits@1 across the four splits of each dataset. We treat methods other than MorsE as topology-based, as they primarily rely on graph structure or relational patterns.

TOME outperforms the topology-based baselines on 21 of 24 dataset–metric combinations, ranking first on all WN18RR splits and leading this group on every FB15k-237 split. MorsE, which combines neighborhood-based initialization with KGE scoring, performs better than TOME on all FB15k-237 splits. On NELL-995, TOME ranks first on both metrics for V2 and V3, and achieves the best Hits@1 on V1, although ULTRA has higher MRR. These results are consistent with TOME's query-aware message construction, which conditions evidence during propagation rather than relying only on post-

message selection. For example, TOME improves Hits@1 over RED-GNN+CP by 19.6 percentage points on NELL-995 V2 (52.5% vs. 32.9%) and 16.7 points on FB15k-237 V1 (48.3% vs. 31.6%).

NELL-995 V4 is the main exception: TOME obtains 0.328 MRR and 21.0% Hits@1, below ULTRA (0.469, 28.8%) and MorsE (0.453, 27.8%). In a separate diagnostic experiment that uses part of the target graph to construct re-indexed meta-training tasks, TOME reaches 0.708 MRR and 58.1% Hits@1, suggesting that additional target-side structural evidence can help on this split; this result is not part of the zero-shot evaluation. In the transductive setting, TOME performs strongly on NELL-995 V4. Across all baselines, TOME ranks first on 13 of 24 inductive combinations, while V4 remains challenging for zero-shot transfer.

5.5.2 Transductive Link Prediction

In the transductive setting, test entities are observed during training, and the evaluation measures link ranking within the known graph. Table 2 summarizes the results across the 12 dataset splits, where each entry reports MRR / Hits@1 and the best baseline is selected separately for each metric.

TOME achieves the best result on 22 of the 24 dataset–metric combinations. It outperforms the best baseline on all WN18RR and NELL-995 splits, as well as on FB15k-237 V2–V4. The only exception is FB15k-237 V1, where the KGE-based MorsE obtains higher MRR and Hits@1. These results show that TOME provides stable transductive performance, while its query-aware in-situ message passing remains effective even when the observed graph contains many query-irrelevant structural connections. Full split-level comparisons and additional analysis are provided in Appendix F.

5.6 Additional Analyses

We provide additional analyses to evaluate the reliability, robustness, and internal mechanisms of TOME. Paired t-tests show that TOME achieves higher scores than the compared baselines in 138 out of 144 evaluated settings, with statistically significant improvements for both MRR and Hits@1. Experiments with incomplete source and target graphs further demonstrate that TOME remains robust under graph sparsity. The ablation results show that episodic meta-training and query-conditioned in-situ message passing make complementary contributions, with message-level query conditioning providing the primary gain and attention-level conditioning offering an additional refinement. Finally, a qualitative case study illustrates how TOME preserves query-relevant evidence while suppressing less informative structural branches. Detailed results are provided in Appendices G, H, I, J, and K.

6 Conclusion

We presented TOME, a framework for inductive link prediction that couples episodic meta-training with query-conditioned in-situ message passing. Episodic tasks expose the model to entity-disjoint local structures, while the query relation shapes message construction and attention before aggregation. This coupling aligns transferable training with query-relevant inference, enabling TOME to reason over unseen entities without entity-specific embeddings. Ablations support the complementary roles of these components.

Experiments on WN18RR, FB15k-237, and NELL-995 show that TOME outperforms topology-based baselines on 21 of 24 inductive and all 24 transductive dataset–metric combinations. Among all evaluated baselines, it ranks first on 13 inductive and 22 transductive combinations. These results demonstrate the effectiveness of the proposed approach across both settings, while the weaker zero-shot performance on inductive NELL-995 V4 highlights a remaining challenge in transferring to structurally difficult target graphs. Future work will explore extensions to dynamic and temporal knowledge graphs.

AI use statement

The authors acknowledge the use of Gemini (Google) solely for language editing and grammar refinement. The authors confirm that no AI tool was used to generate scientific ideas, design methods, conduct experiments, analyze results, or draw conclusions. The authors have reviewed and verified all content and take full responsibility for the manuscript.

REFERENCES

Mingyang Chen, Wen Zhang, Wei Zhang, Qiang Chen, and Huajun Chen. Meta relational learning for few-shot link prediction in knowledge graphs. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pp. 4217–4226, 2019.

Mingyang Chen, Wen Zhang, Zhen Yao, Xiangnan Chen, Mengxiao Ding, Fei Huang, and Huajun Chen. Meta-learning based knowledge extrapolation for knowledge graphs in the federated setting. In Lud De Raedt (ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 1966–1972. International Joint Conferences on Artificial Intelligence Organization, 7 2022a. doi: 10.24963/ijcai.2022/273. URL https://doi.org/10.24963/ijcai.2022/273. Main Track.

Mingyang Chen, Wen Zhang, Yushan Zhu, Hongting Zhou, Zonggang Yuan, Changliang Xu, and Huajun Chen. Meta-knowledge transfer for inductive knowledge graph embedding. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pp. 927–937, 2022b.

Mingyang Chen, Wen Zhang, Yuxia Geng, Zezhong Xu, Jeff Z. Pan, and Huajun Chen. Generalizing to unseen elements: A survey on knowledge extrapolation for knowledge graphs. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China*, pp. 6574–6582. ijcai.org, 2023. doi: 10.24963/IJCAI.2023/737. URL https://doi.org/10.24963/ijcai.2023/737.

Xiaojun Chen, Shengbin Jia, and Yang Xiang. A review: Knowledge reasoning over knowledge graph. *Expert systems with applications*, 141:112948, 2020.

Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alex Smola, and Andrew McCallum. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL https://openreview.net/forum?id=Syg-YfWCW.

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

Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards foundation models for knowledge graph reasoning. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024a. URL https://openreview.net/forum?id=jVEoydFOl9.

Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards foundation models for knowledge graph reasoning. In *International Conference on Learning Representations*, volume 2024, pp. 31598–31619, 2024b.

Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514, 2021.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

Markus Krötzsch, Maximilian Marx, Ana Ozaki, and Veronika Thost. Attributed description logics: Reasoning on knowledge graphs. In *IJCAI*, pp. 5309–5313, 2018.

Jaejun Lee, Chanyoung Chung, and Joyce Jiyoung Whang. Ingram: Inductive knowledge graph embedding via relation graphs. In *International conference on machine learning*, pp. 18796–18809. PMLR, 2023.

Xi Victoria Lin, Richard Socher, and Caiming Xiong. Multi-hop knowledge graph reasoning with reward shaping. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 3243–3253, 2018.

Christian Meilicke, Manuel Fink, Yanjie Wang, Daniel Ruffinelli, Rainer Gemulla, and Heiner Stuckenschmidt. Fine-grained evaluation of rule-and embedding-based systems for knowledge graph completion. In *International semantic web conference*, pp. 3–20. Springer, 2018.

Christian Meilicke, Melisachew Wudage Chekol, Manuel Fink, and Heiner Stuckenschmidt. Reinforced anytime bottom up rule learning for knowledge graph completion. *arXiv preprint arXiv:2004.04412*, 2020.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In Lise Getoor and Tobias Scheffer (eds.), *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pp. 809–816. Omnipress, 2011. URL https://icml.cc/2011/papers/438_icmlpaper.pdf.

Andrea Rossi, Donatella Firmani, Paolo Merialdo, and Tommaso Teofili. Kelpie: an explainability framework for embedding-based link prediction models. *Proc. VLDB Endow.*, 15(12):3566–3569, 2022. doi: 10.14778/3554821.3554845. URL https://www.vldb.org/pvldb/vol15/p3566-rossi.pdf.

Ali Sadeghian, Mohammadreza Armandpour, Patrick Ding, and Daisy Zhe Wang. Drum: End-to-end differentiable rule mining on knowledge graphs. *Advances in neural information processing systems*, 32, 2019.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pp. 593–607. Springer, 2018.

Haseeb Shah, Johannes Villmow, Adrian Ulges, Ulrich Schwanecke, and Faisal Shafait. An open-world extension to knowledge graph completion models. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 3044–3051, 2019.

Zhixiang Su, Di Wang, and Chunyan Miao. Context pooling: Query-specific graph pooling for generic inductive link prediction in knowledge graphs. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 2680–2689, 2025.

Komal Teru, Etienne Denis, and Will Hamilton. Inductive relation prediction by subgraph reasoning. In *International conference on machine learning*, pp. 9448–9457. PMLR, 2020.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. Representing text for joint embedding of text and knowledge bases. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1499–1509, 2015.

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha P. Talukdar. Composition-based multi-relational graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=BylA_C4tPr.

Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE transactions on knowledge and data engineering*, 29(12): 2724–2743, 2017.

Wenhan Xiong, Thien Hoang, and William Yang Wang. Deeppath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pp. 564–573, 2017.

Fan Yang, Zhilin Yang, and William W Cohen. Differentiable learning of logical rules for knowledge base reasoning. *Advances in neural information processing systems*, 30, 2017.

Donghan Yu, Yiming Yang, Ruohong Zhang, and Yuexin Wu. Knowledge embedding based graph convolutional network. In *Proceedings of the web conference 2021*, pp. 1619–1628, 2021.

Yongqi Zhang and Quanming Yao. Knowledge graph reasoning with relational digraph. In *Proceedings of the ACM web conference 2022*, pp. 912–924, 2022.

Yongqi Zhang, Zhanke Zhou, Quanming Yao, Xiaowen Chu, and Bo Han. Adaprop: Learning adaptive propagation for graph neural network based knowledge graph reasoning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, pp. 3446–3457, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701030. doi: 10.1145/3580305.3599404. URL https://doi.org/10.1145/3580305.3599404.

Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Xhonneux, and Jian Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. *Advances in neural information processing systems*, 34:29476–29490, 2021.

A Recursive r-Digraph Encoding

Algorithm 1 summarizes the recursive query-anchored encoding procedure used by TOME.

Algorithm 1 Recursive r-digraph Encoding for $(e_q, r_q, ?)$

Require: Query entity e_q, query relation r_q, knowledge graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$, layers L.
Ensure: Final state representations $\{\mathbf{h}_{e_a}^L \mid e_a \in \mathcal{E}\}$.

Initialize: $\mathbf{h}_{e_q}^0 \leftarrow \mathbf{0}, \mathcal{N}^0 \leftarrow \{e_q\}$;
for $\ell = 1$ to L **do**
  $\mathcal{T}^\ell \leftarrow \{(e_s, r, e_o) \in \mathcal{T} \mid e_s \in \mathcal{N}^{\ell-1}\}$;
  $\mathcal{N}^\ell \leftarrow \{e_o \mid (e_s, r, e_o) \in \mathcal{T}^\ell\}$;
  Calculate attention weights $\alpha^\ell_{e_s,r,e_o|r_q}$ via Eq. equation 3 to filter $\mathcal{T}^\ell$;
  Update entity representations $\mathbf{h}^\ell_{e_o}$ via Eq. equation 2 for all $e_o \in \mathcal{N}^\ell$;
end for
Assign $\mathbf{h}^L_{e_a} = \mathbf{0}$ for all $e_a \notin \mathcal{N}^L$;
return $\{\mathbf{h}^L_{e_a} \mid e_a \in \mathcal{E}\}$;

B Inference Complexity

For a query $(e_q, r_q, ?)$, the computational complexity of TOME depends on the L-layer query-anchored encoding (Algorithm 1). Let d and d_a denote the embedding and attention dimensions. At each layer ℓ, processing the $|\mathcal{T}^\ell|$ active triplets incurs $\mathcal{O}(|\mathcal{T}^\ell| d d_a)$ for projecting features into the attention hidden space and $\mathcal{O}(|\mathcal{T}^\ell| d)$ for scaling and aggregating the weighted messages. Meanwhile, updating the $|\mathcal{N}^\ell|$ active entities requires $\mathcal{O}(|\mathcal{N}^\ell| d^2)$. Since the attention dimension satisfies $d_a \geq 1$, the overall computational complexity is bounded by $\mathcal{O}\big(\sum_{\ell=1}^{L}(|\mathcal{T}^\ell| d d_a + |\mathcal{N}^\ell| d^2)\big)$.

TOME localizes the reasoning process. By restricting exploration to query-specific topological neighborhoods, the operational space remains strictly independent of the global graph scale, satisfying $|\mathcal{N}^\ell| \ll |\mathcal{E}|$ and $|\mathcal{T}^\ell| \ll |\mathcal{T}|$. This mechanism bypasses the full-graph computational bottleneck of standard transductive GNNs (which scale as $\mathcal{O}(|\mathcal{T}| d + |\mathcal{E}| d^2)$), enabling TOME to score all candidate entities simultaneously in a single efficient forward pass.

C Dataset Statistics

Table 3 reports the statistics of the source and target KGs used in our experiments. Here, $|\mathcal{R}|$, $|\mathcal{E}|$, and $|\mathcal{T}|$ denote the numbers of relations, entities, and triples, respectively.

Table 3: Statistics of the inductive datasets. $|\mathcal{R}|$, $|\mathcal{E}|$, and $|\mathcal{T}|$ denote the number of relations, entities, and triples.

Dataset	Ver.	Source KG			Target KG		
		$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{T}\rvert$
WN18RR	V1	9	2,746	6,678	8	922	1,618
	V2	10	6,954	18,968	10	2,757	4,011
	V3	11	12,078	32,150	11	5,084	6,327
	V4	9	3,861	9,842	9	7,084	12,334
FB15k-237	V1	180	1,594	5,226	142	1,093	1,993
	V2	200	2,608	12,085	172	1,660	4,145
	V3	215	3,668	22,394	183	2,501	7,406
	V4	219	4,707	33,916	200	3,051	11,714
NELL-995	V1	14	3,103	5,540	14	225	833
	V2	88	2,564	10,109	79	2,086	4,586
	V3	142	4,647	20,117	122	3,566	8,048
	V4	76	2,092	9,289	61	2,795	7,073

D Baseline Details

We provide additional details about the baselines evaluated in this work. Rule-based methods include AnyBURL Meilicke et al. (2020), which extracts symbolic rules from observed triples, Neural-LP Yang et al. (2017), which learns differentiable rule compositions, and DRUM Sadeghian et al. (2019), which learns relation-specific rule structures. Subgraph-based GNNs reason over local graph structures: GraIL Teru et al. (2020) encodes query-centered subgraphs, NBFNet Zhu et al. (2021) performs generalized Bellman-Ford-style propagation, and RED-GNN Zhang & Yao (2022) uses dynamic programming with query-dependent attention.

AdaProp Zhang et al. (2023) uses query-guided adaptive propagation to limit the entities considered during reasoning. Context Pooling (CP) Su et al. (2025) selects query-relevant neighborhoods based on candidate representations; we evaluate CP with the NBFNet and RED-GNN backbones. ULTRA Galkin et al. (2024b) learns relation-aware representations with a pretrained graph reasoning model, while InGram Lee et al. (2023) models relation dependencies using a relation graph and generates entity and relation representations through graph aggregation.

MorsE Chen et al. (2022b) is a meta-learning-based KGE method. It uses a meta-GNN to derive structural initializations for unseen entities, followed by KGE-style scoring. Unlike the topology-based methods above, its final prediction relies on entity representations and a KGE scoring function.

For result provenance, baseline results marked with * are taken from Su et al. Su et al. (2025). All other baseline results are reproduced using the corresponding publicly released implementations, including AdaProp, ULTRA, and InGram. For inductive evaluation, models are trained on each source KG and tested on its corresponding target KG. The full split-level results are reported in the corresponding main and appendix tables.

E Implementation Details

We select hyperparameters by grid search using validation MRR. The learning rate and L2 weight decay are searched within $[10^{-5}, 10^{-2}]$, and the dropout rate within $[0, 0.3]$. For the r-digraph architecture, the embedding dimension is selected from $\{16, 32, 48, 64\}$, the attention projection

Table 4: Transductive link prediction results on WN18RR, FB15k-237, and NELL-995. Best and second-best results are **bolded** and underlined, respectively. Baselines marked with * are sourced from Su et al. (2025).

Models	WN18RR				FB15k-237				NELL-995			
	V1	V2	V3	V4	V1	V2	V3	V4	V1	V2	V3	V4
Metric: MRR												
AnyBURL*	0.185	0.234	0.183	0.195	0.248	0.253	0.248	0.248	0.190	0.215	0.196	0.223
Neural-LP*	0.549	0.580	0.518	0.550	0.304	0.364	0.347	0.328	0.374	0.302	0.259	0.268
DRUM*	0.584	0.585	0.519	0.574	0.317	0.377	0.357	0.333	0.413	0.320	0.274	0.292
GraIL*	0.455	0.486	0.212	0.448	0.252	0.253	0.218	0.226	0.346	0.363	0.323	0.296
NBFNet*	0.418	0.393	0.515	0.266	0.495	0.473	0.438	0.415	0.503	0.422	0.393	0.400
RED-GNN*	0.574	0.597	0.561	0.581	0.506	0.505	0.474	0.456	0.516	0.503	0.466	0.475
AdaProp	0.604	0.620	0.583	0.592	0.455	0.511	0.501	0.474	0.520	0.531	0.470	0.487
NBFNet+CP*	0.435	0.439	0.510	0.331	0.513	0.480	0.447	0.421	0.515	0.449	0.407	0.420
RED-GNN+CP*	0.594	0.612	0.573	0.593	0.512	0.508	0.478	0.450	0.524	0.516	0.468	0.484
ULTRA	0.573	0.592	0.534	0.579	0.593	0.545	0.563	0.521	0.633	0.651	0.602	0.681
InGram	0.291	0.240	0.137	0.237	0.236	0.206	0.150	0.142	0.533	0.202	0.240	0.185
MorsE	0.641	0.658	0.610	0.638	**0.767**	0.641	0.601	0.582	0.685	0.620	0.580	0.688
TOME (Ours)	**0.755**	**0.773**	**0.769**	**0.786**	0.710	**0.671**	**0.659**	**0.679**	**0.750**	**0.730**	**0.678**	**0.720**
Metric: Hits@1 (%)												
AnyBURL*	16.1	19.5	15.8	16.6	22.3	19.2	18.7	18.6	16.2	15.1	13.2	16.8
Neural-LP*	47.0	52.1	45.4	47.5	23.9	28.7	27.0	25.3	28.9	22.9	18.7	20.2
DRUM*	52.1	52.5	45.5	51.3	26.0	30.5	28.3	26.0	33.8	25.4	20.2	22.8
GraIL*	37.5	40.8	10.8	35.8	39.5	39.5	34.3	31.1	24.6	25.0	23.2	20.5
NBFNet*	31.9	29.3	43.4	18.8	40.4	37.5	33.9	31.5	41.6	32.2	29.1	30.1
RED-GNN*	53.1	55.0	50.9	53.2	43.7	41.1	37.5	35.5	42.3	40.5	36.6	38.2
AdaProp	55.0	57.3	53.9	54.2	36.8	43.9	42.1	38.1	43.6	43.4	37.2	39.2
NBFNet+CP*	34.1	35.0	43.6	20.2	44.6	38.1	34.5	31.7	42.3	34.5	29.8	32.0
RED-GNN+CP*	54.8	56.5	52.0	54.4	44.6	41.9	37.6	35.1	43.8	41.8	36.4	38.5
ULTRA	53.6	55.3	50.6	52.5	49.1	45.3	46.2	43.7	54.5	52.3	48.7	57.4
InGram	15.7	13.9	8.4	9.5	14.3	10.6	8.7	6.7	41.6	11.5	15.5	11.6
MorsE	57.1	59.9	51.3	56.8	**66.6**	51.0	48.6	44.0	57.7	49.8	44.7	57.9
TOME (Ours)	**70.3**	**72.8**	**72.4**	**74.0**	62.6	**57.4**	**55.1**	**56.9**	**65.6**	**62.8**	**57.0**	**59.0**

dimension from $\{3, 5\}$, the number of reasoning layers from $\{3, 4, 5\}$, and the activation function from $\{\text{identity}, \tanh, \text{ReLU}\}$. The meta-training batch size is selected from $\{32, 64\}$.

F Detailed Transductive Results

Table 4 reports the complete transductive results for all compared methods across the 12 dataset splits. TOME outperforms the strongest baseline on 22 of the 24 dataset–metric combinations. It ranks first on all WN18RR and NELL-995 splits and on FB15k-237 V2–V4. On FB15k-237 V1, MorsE remains stronger in both MRR and Hits@1, indicating that its KGE-based entity representation is particularly effective on this split.

The overall advantage of TOME is consistent with its query-aware in-situ message passing mechanism. Because the query relation is incorporated into message construction and attention before aggregation, irrelevant structural connections can be down-weighted during propagation. This complements episodic meta-training, which encourages the model to learn transferable structural patterns across localized tasks. The transductive results therefore support the effectiveness of the coupled design beyond the zero-shot inductive setting.

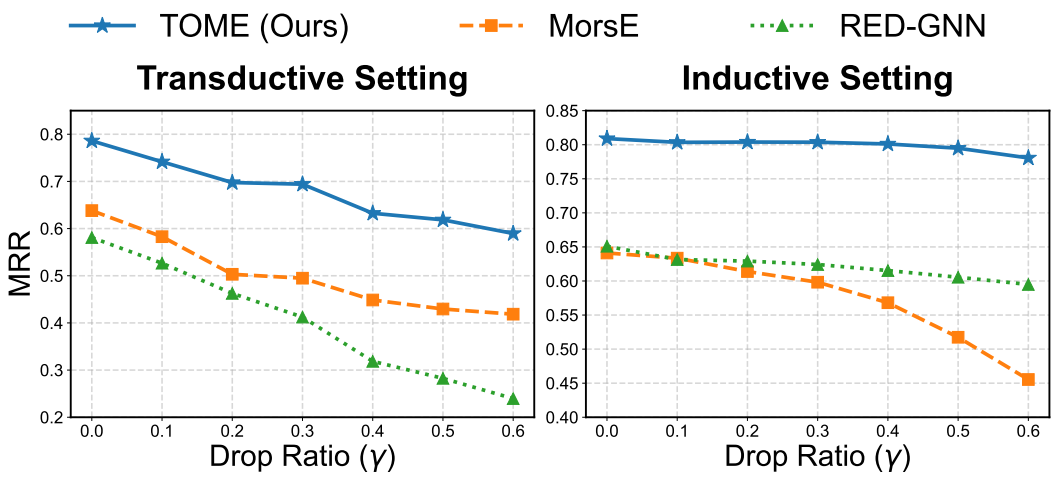

Figure 3: MRR under source graph drop ratios (γ). The left panel reports transductive evaluation, while the right panel reports inductive evaluation.

G Statistical Significance Tests

To assess the statistical significance of TOME's improvements, we conduct paired t-tests against RED-GNN Zhang & Yao (2022), RED-GNN+CP Su et al. (2025), and MorsE Chen et al. (2022b). The evaluation covers 144 settings across the inductive and transductive link prediction experiments.

TOME achieves higher performance than the compared methods in 138 out of the 144 settings. The aggregated results yield p-values below 10^{-3} for MRR and approximately 2×10^{-3} for Hits@1. These results provide statistical evidence that the observed improvements are consistent across datasets, evaluation settings, and evaluation metrics.

H Source-Graph Sparsity Analysis

In real-world applications, the training data available for a source knowledge graph may be incomplete. To evaluate the robustness of TOME under source-side data scarcity, we conduct experiments on WN18RR (V4) by randomly removing triples from the source graph $\mathcal{G}_{\text{train}}$. The drop ratio $\gamma \in \{0.0, 0.1, \ldots, 0.6\}$ denotes the proportion of removed triples.

We consider two evaluation settings:

1. **Transductive evaluation:** The model is trained and evaluated on the same sparsified source graph $\mathcal{G}_{\text{train}}$. This setting measures reasoning stability when the graph used for training and evaluation is incomplete.
2. **Inductive evaluation:** The model is trained on the sparsified source graph $\mathcal{G}_{\text{train}}$ and evaluated on the complete, unseen target graph $\mathcal{G}_{\text{test}}$. This setting measures whether the model can transfer structural reasoning learned from limited source-side evidence.

Figure 3 reports the MRR under different source-graph drop ratios.

In the transductive setting, TOME exhibits greater robustness to source-side sparsity than the non-meta baseline RED-GNN Zhang & Yao (2022). When $\gamma = 0.6$, RED-GNN's MRR decreases by 0.3414, from 0.5810 to 0.2396. In comparison, the MRR reductions of MorsE Chen et al. (2022b) and TOME are 0.2196 and 0.1964, respectively. TOME achieves an MRR of 0.5896 at $\gamma = 0.6$, which is higher than the 0.5810 MRR obtained by RED-GNN on the complete graph. This result indicates that episodic meta-training can improve the stability of structural reasoning under severe source-side sparsity.

In the inductive setting, TOME also maintains relatively stable performance when the source graph is incomplete. Under the highest drop ratio, MorsE's MRR decreases by 0.1857, from 0.6411 to 0.4554,

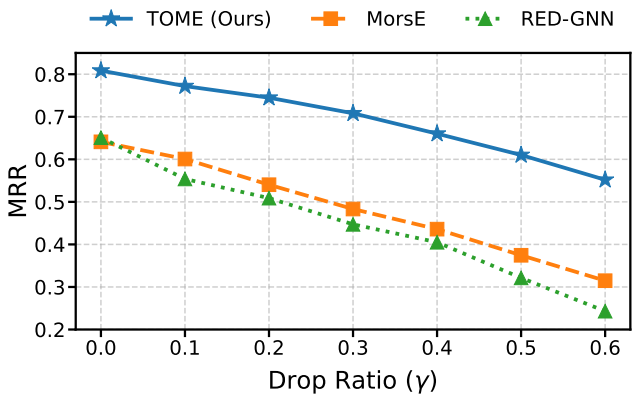

Figure 4: Inductive MRR under target graph drop ratios (γ). Meta-learning exhibits superior robustness over standard GNNs; query-aware reasoning ensures stable inference.

whereas TOME's MRR decreases by only 0.0284, from 0.8090 to 0.7806. The smaller performance degradation suggests that TOME can exploit transferable topological patterns without depending entirely on complete source-side neighborhoods.

Overall, the source-sparsity results show that TOME combines the transferability of episodic meta-training with the stability of query-aware topological reasoning. Even when 60% of the source triples are removed, TOME retains an inductive MRR of 0.7806, which is higher than the MRR of MorsE trained on the complete graph under $\gamma = 0.0$.

I Target Knowledge Graph Sparsity Analysis

Beyond training data scarcity, evaluating reasoning capability in incomplete deployment environments (i.e., the target knowledge graph $\mathcal{G}_{\text{test}}$) is equally critical. In this experiment, we randomly drop triplets from the WN18RR (V4) target graph. Models trained on the complete source graph are deployed for zero-shot link prediction across varying drop ratios $\gamma \in \{0.0, 0.1, \ldots, 0.6\}$. Figure 4 illustrates the inductive MRR trajectories, yielding two main observations:

First, meta-learning better adapts to sparse target environments. As γ increases from 0.0 to 0.6, the MRR of standard topological models (e.g., RED-GNN Zhang & Yao (2022)) suffers a severe decrease of 0.4080 (0.6510 $\rightarrow$ 0.2431). Meta-learning-based models (MorsE Chen et al. (2022b) and TOME) mitigate this decline, limiting MRR drops to 0.3264 and 0.2566, respectively. This confirms that transferable topological knowledge prevents overfitting to the specific density of the source graph. Instead of relying on complete local neighborhoods, these models leverage learned topological priors to infer missing semantics.

Second, TOME further improves overall robustness via query-aware in-situ reasoning. While MorsE is more robust than RED-GNN, its absolute performance remains constrained (MRR=0.3147 at $\gamma = 0.6$) because fragmented neighborhoods fail to initialize stable entity embeddings. TOME circumvents this by using the query relation as a navigation signal. When missing edges disrupt optimal paths, its in-situ mechanism dynamically re-evaluates the surviving topology to extract alternative evidence chains. Consequently, TOME maintains a highly reliable inductive MRR of 0.5519 even under a severe sparsity of $\gamma = 0.6$.

Summary. TOME effectively integrates meta-learning transferability with query-aware topological stability. This integration ensures reliability for real-world inductive link prediction in sparse deployment environments.

J Ablation Study

To examine the contributions of episodic meta-training and query-conditioned message passing, we conduct ablation studies on the WN18RR (V1) and FB15k-237 (V1) inductive splits. In addition to

Table 5: Ablation study of TOME on WN18RR (V1) and FB15k-237 (V1) inductive splits. Best results are **bolded**.

Architecture	WN18RR (V1)				FB15k-237 (V1)			
	MRR	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10
RED-GNN + Meta	0.807	75.6	85.2	89.5	0.544	45.5	64.1	68.5
RED-GNN + In-situ Msg	0.712	66.2	77.5	81.1	0.373	30.9	40.2	45.4
TOME w/o Meta	0.689	64.2	73.2	77.1	0.333	28.1	37.2	42.0
TOME w/o r_q	0.684	62.3	75.0	79.2	0.321	24.8	38.1	44.1
TOME w/o Msg-r_q	0.829	77.8	88.5	90.8	0.528	44.2	61.2	66.5
TOME w/o Attn-r_q	0.831	78.1	88.9	91.2	0.566	47.8	66.0	70.6
TOME (Ours)	**0.836**	**78.7**	**89.4**	**91.9**	**0.572**	**48.3**	**66.6**	**71.2**

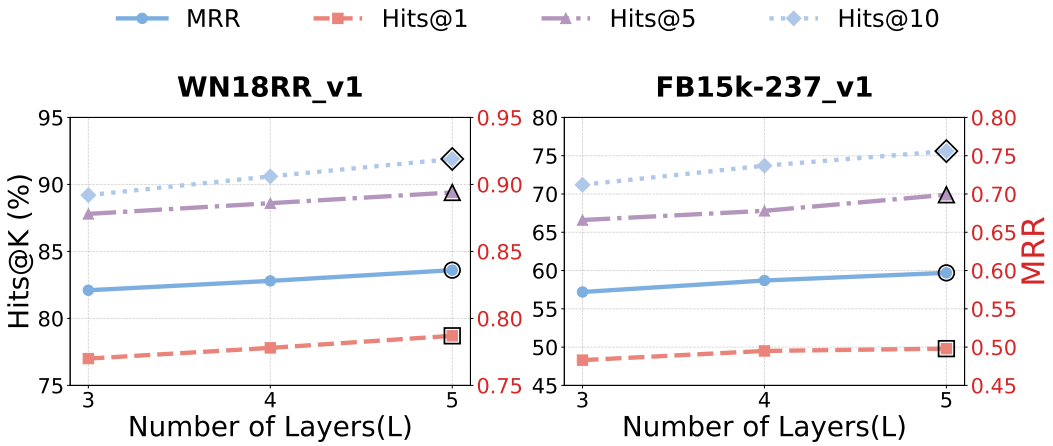

Figure 5: Impact of reasoning layer depth (L) on WN18RR (V1) and FB15k-237 (V1). Stable gains at deeper layers validate the noise-filtering capability of query-aware reasoning.

the TOME variants, we include two RED-GNN-based controls to compare episodic meta-training and in-situ message construction in a common backbone. The variants are defined as follows:

- **RED-GNN + Meta:** We apply episodic meta-training to the RED-GNN backbone.
- **RED-GNN + In-situ Msg:** We augment the RED-GNN backbone with query-conditioned in-situ message construction, without episodic meta-training.
- **TOME w/o Meta:** We train TOME with standard batch training on the source graph, omitting episodic meta-training.
- **TOME w/o r_q:** We remove the query relation r_q from both message formulation and attention, making the model query-agnostic.
- **TOME w/o Msg-r_q:** We remove r_q from message formulation while retaining it in attention.
- **TOME w/o Attn-r_q:** We retain r_q in message formulation but remove it from attention.

RED-GNN-Based Variants. The two RED-GNN-based controls provide complementary reference points. RED-GNN + Meta obtains MRR/Hits@1 scores of 0.807/75.6% on WN18RR and 0.544/45.5% on FB15k-237, while RED-GNN + In-situ Msg obtains 0.712/66.2% and 0.373/30.9%, respectively. The full TOME model outperforms both variants on both datasets. Compared with RED-GNN + Meta, it improves MRR by 0.029 and 0.028, and Hits@1 by 3.1 and 2.8 percentage points on WN18RR and FB15k-237, respectively. The gains over RED-GNN + In-situ Msg are larger, particularly on FB15k-237. These results show that the complete TOME configuration is more effective than either RED-GNN-based reference variant.

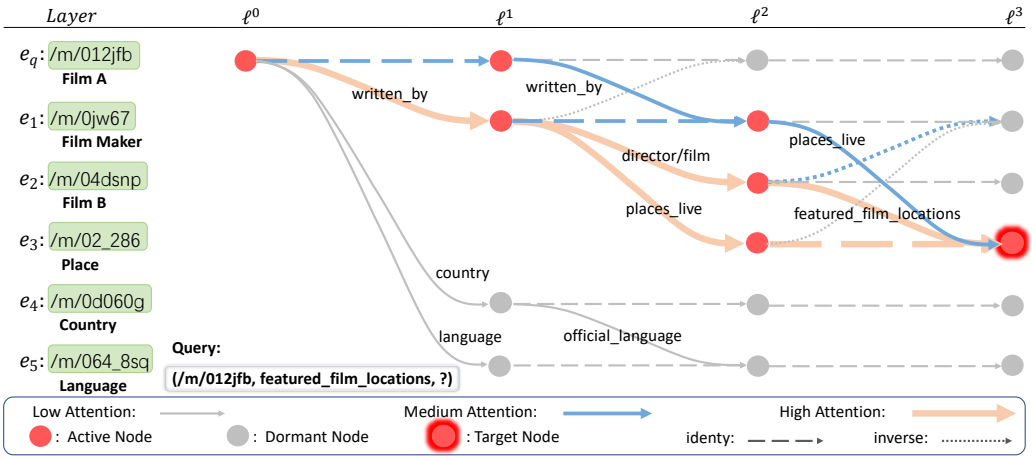

Figure 6: Visualization of the reasoning process for query (/m/012jfb, featured_film_locations, ?). Colors indicate attention (orange: high, blue: medium, grey: low); line styles denote relation semantics (solid: forward, dashed: `identity`, dotted: `inverse`).

Impact of Episodic Meta-Training. Comparing TOME with TOME w/o Meta isolates the effect of episodic meta-training within the TOME architecture. Removing meta-training decreases MRR from 0.836 to 0.689 on WN18RR and from 0.572 to 0.333 on FB15k-237. Hits@1 also drops from 78.7% to 64.2% and from 48.3% to 28.1%, respectively. The consistent declines across both datasets indicate that episodic meta-training is an important component for TOME's inductive transfer.

Effect of Query Conditioning. Removing r_q from both message formulation and attention leads to substantial performance drops. Relative to full TOME, TOME w/o r_q decreases MRR by 0.152 and Hits@1 by 16.4 percentage points on WN18RR, and by 0.251 and 23.5 percentage points on FB15k-237. This demonstrates the importance of conditioning the reasoning process on the query relation.

Where to Inject the Query Relation. The comparison between TOME w/o Msg-r_q and TOME w/o Attn-r_q distinguishes message-level from attention-level query conditioning. On FB15k-237, removing r_q from message formulation reduces MRR by 0.044 and Hits@1 by 4.1 percentage points, whereas removing it from attention reduces them by 0.006 and 0.5 percentage points. On WN18RR, the corresponding MRR decreases are 0.007 and 0.005, with Hits@1 decreases of 0.9 and 0.6 percentage points. Thus, message-level conditioning has a larger effect, particularly on FB15k-237, while attention-level conditioning provides an additional, smaller improvement. The full model, which injects r_q into both message formulation and attention, achieves the best results across the reported metrics.

Impact of Model Depth (L). Figure 5 shows the results for reasoning depths $L \in \{3, 4, 5\}$. Increasing L from 3 to 5 improves MRR from 0.572 to 0.597 on FB15k-237 and from 0.821 to 0.836 on WN18RR; Hits@10 on FB15k-237 increases from 71.2% to 75.6%. These results support using $L = 5$, which captures additional multi-hop evidence without a performance decline in the evaluated settings.

K Qualitative Case Study

We visualize the recursive encoding process to examine how TOME identifies and aggregates query-relevant topological evidence. We select a real-world query from the test graph $\mathcal{G}_{\text{test}}$: `(/m/012jfb, featured_film_locations, ?)`. The target answer is `/m/02_286`.

Figure 6 illustrates the multi-hop inference process over a localized relational digraph (r-digraph). The reasoning unfolds across four message-passing layers (ℓ^0 to ℓ^3), with information flow modulated by query-aware attention weights. For clarity, identity and inverse labels, as well as some low-attention edges, are omitted from the visualization to highlight the main topological paths contributing to the prediction.

TOME extracts two primary high-attention evidence paths, shown by the orange lines. The first path traverses the filmmaker node /m/0jw67 at ℓ^1 and reaches the target node /m/02_286 at ℓ^2 through the places_live relation. The second path passes through another film node /m/04dsnp at ℓ^2 and reaches the target at ℓ^3. After the target is reached at ℓ^2, TOME retains its state through a high-attention identity edge, shown by the dashed orange line. This allows the accumulated evidence to be preserved across subsequent reasoning layers.

Medium-attention edges, shown in blue, represent secondary structural paths. In contrast, low-attention edges, shown in grey, correspond to topological distractors, such as the country node /m/0d060g and the language node /m/064_8sq introduced at ℓ^1. Their weak relevance to the query results in low attention scores. As the neighborhood expands through relations such as official_language, the in-situ message-passing mechanism suppresses these less informative branches. This case study illustrates how TOME preserves query-relevant evidence while limiting the influence of irrelevant topology.