Fusing Relation Graph and Mutual Information for Inductive Link Prediction in Knowledge Graphs

Hongbo Liu , Jicang Lu, Yaohui Hao, Taojie Zhu

Information Engineering University, Zhengzhou 450000, China
lhb921@163.com

Abstract. Inductive link prediction in knowledge graphs refers to the task of inferring links involving entities unseen during training, where the relations may be seen or unseen. Most existing approaches are limited to predicting only seen relations and struggle to generalize to unseen relations, which restricts their utility in dynamic settings. To address this challenge, we introduce RGIILP, a novel framework designed for inductive link prediction. Specifically, we construct a relation graph from the source knowledge graph and design a neural network model that enables interactive feature propagation between entities and relations. Furthermore, we introduce a mechanism that maximizes the mutual information between global and local representations to capture the graph's global structural information. Experiments on several benchmark datasets demonstrate that RGIILP outperforms existing state-of-the-art methods for inductive link prediction task.

Keywords: Inductive Link Prediction, Knowledge Graphs, Relation Graph, Mutual Information.

1 Introduction

Knowledge graphs play an increasingly vital role in modern research. However, knowledge graphs constructed from real-world data invariably suffer from incompleteness. Link prediction serves as a key technique to address this issue. Conventional link prediction methods typically first learn embeddings of entities and relations derived from a source knowledge graph, and then apply these representations to score triples in a target graph. However, such approaches are incapable of performing link prediction in graphs containing newly emerging, unseen entities and relations. To tackle this challenge, a range of inductive link prediction methods have been developed. Methods such as GraIL[1], CoMPILE[2] and TACT[3] learn topological structures from knowledge graphs to infer links involving previously unobserved entities, yet they are still unable to handle unseen relations. INGRAM[4] addresses link prediction with unseen relations by introducing an interactive learning structure between the knowledge graph and a corresponding relation graph. The model employs a Minimum Spanning Tree (MST) to ensure basic graph connectivity and adopts strategies like dynamic partitioning and random re-initialization to ensure effective learning of all entities and relations. However, the MST is inherently a form of subgraph sampling, meaning the relation graph

is constructed from local structure rather than the complete graph. This may restrict the model's capability to effectively encode complex relational patterns across the full graph.

To overcome the aforementioned shortcomings, we introduce a novel model that integrates relation graph learning with mutual information for inductive link prediction in knowledge graphs. The model introduces a mutual information learning mechanism to capture global structural features of the knowledge graph. Specifically, we first construct a relation graph, and then develop a neural network that enables iterative interaction between the knowledge graph and the relation graph. Finally, the model jointly optimizes the objectives of mutual information learning and entity-relation interaction within a unified training framework. The main contributions of the paper is summarized as follows:

(1) We propose to incorporate mutual information into the inductive link prediction framework, which enhances the capture of global structural information by maximizing the agreement between node-level and graph-level representations.

(2) We construct a relation graph derived from the knowledge graph and develop an entity-relation interaction learning architecture that enables iterative updating of entity and relation representations across both graphs.

(3) The model is evaluated on benchmark datasets including FB15k-237, WikiData and NELL-995. Experimental results confirm that the proposed model achieves superior performance compared to the existing baseline methods.

2 Related Work

2.1 Inductive link prediction in KG

Rule-based approaches, such as RuleN[5], TensorLog [6], Neural-LP[7], DRUM[8], operate by mining high-quality logical rules from the knowledge graph, then leveraging these rules to predict relations between unknown entities. However, these rule mining algorithms primarily focus on discovering Horn rules, which have limited expressive power. Furthermore, generating high-quality rules for massive knowledge graphs entails considerable computational resources.

Subgraph-based methods, including GraiL, ComPILE, operate under the hypothesis that the local neighborhood surrounding a target relation contains meaningful structural patterns, which can be utilized to predict potential semantic associations between unknown entities. These approaches typically extract a local subgraph closely related to the target relation and employ graph neural networks to learn such patterns.

Additionally, two other influential approaches have been developed to tackle the inductive reasoning problem from distinct perspectives. NBFNet [9] integrates the Bellman-Ford algorithm with graph neural networks, simulating an iterative search over all potential paths in the graph for representation learning and inference. RED-GNN [10] constructs a relation-specific digraph (r-digraph) and utilizes dynamic programming with recursive encoding to achieve efficient and interpretable inference.

However, a fundamental limitation of the aforementioned methods is their inability to handle unseen relations. Consequently, these approaches are poorly suited for real-

world knowledge graphs that evolve continuously, such as NELL and Wikidata, where new relations frequently emerge through open information extraction or collaborative curation.

To address the prediction of unseen relations, a variety of innovative approaches have emerged. RMPI [11] incorporates ontological information from the knowledge graph as auxiliary input to facilitate generalization. INGRAM proposes a novel interactive learning architecture between the knowledge graph and a derived relation graph. Despite their contributions, both methods exhibit notable shortcomings. RMPI's performance is highly dependent on the availability and quality of external ontological knowledge, which is often incomplete or expensive to obtain. Although INGRAM constructs a relation graph to model relational interactions, it builds this graph based on a Minimum Spanning Tree (MST) sampled from the original knowledge graph. This construction is inherently local and may fail to capture the rich, global relational dependencies present in the full graph, thereby limiting its capacity to learn comprehensive relational semantics for robust generalization.

2.2 Mutual Information

Mutual Information (MI) maximization, which assesses the inter-variable dependency, has been established as a powerful principle for learning correlated and informative representations. In representation learning, maximizing the MI between input data and its learned representations encourages the model to preserve salient features. For graph-structured data, Deep Graph Infomax (DGI)[12] adapts this principle by maximizing the MI between local node representations and a global graph-level summary. This objective facilitates the model's capacity to identify the correlation between local substructures and the global graph topology, fostering representations that are both locally discriminative and globally coherent. DGI and its variants have shown promising results in unsupervised graph representation learning.

3 Model

3.1 OverView of the model's structural framework

The paper introduces an inductive link prediction model that integrates relation graphs and mutual information. An overview of the model's architecture is provided in Figure 1. The framework consists of two main components: an entity-relation interaction learning module and a mutual information module. The entity-relation interaction module employs a network structure based on interactive learning between entities and relations, comprising an entity layer and a relation layer. The entity layer updates entity embeddings by aggregating information from both the knowledge graph structure and the associated relation graph. Subsequently, the relation layer refines relation representations based on these updated entity embeddings. The mutual information module enhances global representational consistency by maximizing the mutual information between localized node-level embeddings and a holistic graph-level summary. Then the two modules are trained jointly within a unified learning framework, where the entity-

relation interaction loss and the mutual information objective are optimized simultaneously. This enables the model to perform inductive link prediction by effectively balancing localized relational reasoning with the preservation of global graph semantics.

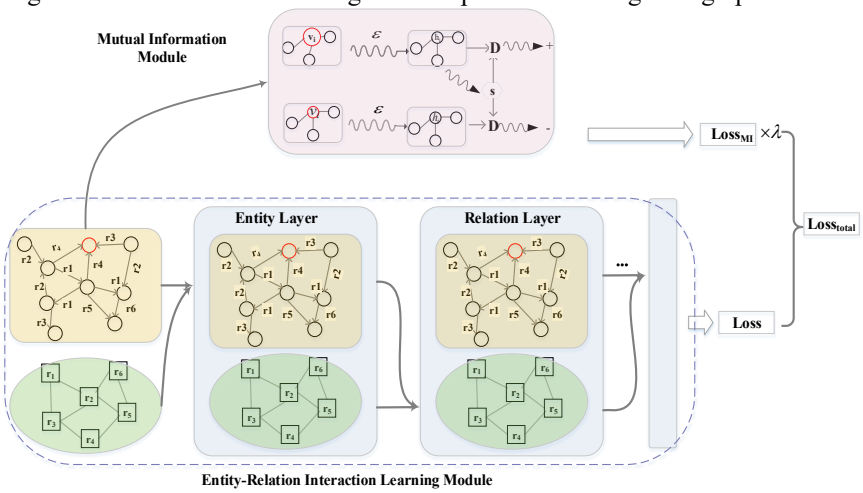

Fig. 1. Overall framework of RGIILP

3.2 Entity-Relation Interaction Learning Module

The entity-relation interaction learning module aims to jointly model both the knowledge graph and its corresponding relation graph. This module primarily consists of two components: relation graph construction and entity-relation interaction learning.

Relation Graph Construction

For a given knowledge graph G=(E,R), where E and R correspond to the sets of entities and relations, respectively. The corresponding relation graph L(G) is defined as follows:

a. The vertex set V(L(G))=R(G), meaning that each vertex in L(G) corresponds to a directed edge in G.

b. The edge set E(L(G)) is defined based on the connectivity of edges in G. Specifically, for any two edges r_1=(x,y) and r_2=(w,z) in G, if y=w (it means that the tail entity of r_1 is the head entity of r_2), then a directed edge (r_1,r_2) is created in L(G). This can be formally expressed as:

$$(r_1, r_2) \in E(L(G)) \Leftrightarrow r_1 = (x, y), r_2 = (y, z) \quad in \quad G(x, y, z \in E) \tag{1}$$

Entity-Relation Interaction Learning

Entities and relations are fundamental components of a knowledge graph. However, methods that model only entities fail to capture the rich relational semantics between

them, while those that focus exclusively on relations cannot adequately incorporate entity-specific features and contextual information. To effectively model the interactions and mutual dependencies between entities and relations, we employ a neural network architecture for entity-relation interactive learning. The architecture comprises two distinct functional layers: an entity layer and a relation layer. The entity layer learns entity representations from the original knowledge graph by aggregating information from neighboring entities and their associated relations, effectively fusing local structural information with broader relational context. The relation layer learns relation representations from the relation graph by aggregating information from the entities connected by each relation, thereby capturing their semantic and structural properties.

During message propagation, entities and relations engage in continuous information exchange to achieve collaborative feature updates. Specifically, the entity layer transmits updated entity representations to the relation layer, which in turn feeds back updated relation representations to the entity layer. Through multiple iterations, both entity and relation representations undergo continuous refinement, ultimately achieving collaborative optimization of these representations.

(1)Relation Layer

The relation layer propagates messages in the relation graph, whereby node representations are updated by aggregating the representation vectors of the node and its neighbouring nodes. In order to differentiate the importance of different relations, an attention-based weighting factor is introduced during the aggregation process. The relation representations obtained from the pre-trained model during the graph construction serve as the initial values for nodes in the relation graph. The attention weight for a node *i* and its neighbor is computed based on their similarity, formulated as:

$$\alpha_{ij}^{l} = \frac{\exp(\sigma(P_e^l[z_i^l \,||\, z_j^l]))}{\sum_{z_j \in N_i} \exp(\sigma(P_e^l[z_i^l \,||\, z_j^l]))} \tag{2}$$

Where α_{ij} denotes the attention weight assigned to the directed edge(*j*,*i*); z_i^l corresponds to the latent embedding of node *i* at the l-th layer; z_j^l represents the relation representation associated with node *i*, P_e^l refers to the weight matrix corresponding to relation "e" at the *l*-th layer, $\sigma()$ designates the nonlinear activation function, N_i indicates the set of neighboring nodes adjacent to node *i*, $||$ signifies the vector concatenation operation.

The representation of node *i* is iteratively updated through weighted aggregation of neighborhood information. Specifically, the representation of *i* in relation graph at the *l*-th layer is defined as follows:

$$z_i^{l+1} = \sigma(\sum_{z_j \in N_i} \alpha_{ij}^l W^l z_j^l) \tag{3}$$

Where W denotes the weight matrix, z_j^l represents the representation of the relation connected to node i in the l-th layer, $\sigma()$ denotes the ReLU (Rectified Linear Unit) activation function.

(2)Entity layer

The entity layer propagates messages based on the knowledge graph. The representation of entity v_i is updated by aggregating the representations of its neighboring nodes, its own representation, and the relation representations associated with these neighbors. The specific calculation is defined as follows:

$$v_i^{l+1} = \sigma(\beta_{ii}^l P_v^l [v_i^l \| z_i^l] + \sum_{v_j \in N_i'} \sum_{r_k \in R_{ji}} \beta_{ijk}^l P_v^l [v_j^l \| z_k^l]) \tag{4}$$

Where P_v^l is the parameter matrix associated with entity v in the l-th layer; v_j^l represents the embedding of a neighboring node adjacent to entity i; z_k^l corresponds to the representation of the relation associated with entity i; β_{ijk}^l indicates the attention coefficient between entity v_i and v_j through relation r_k; β_{ii}^l refers to the self-attention coefficient of entity v_i.

(3)Entity-Relation Interaction

To characterize the associations between relations and entities, this module adopts the scoring function from the DistMult as the triple scoring function:

$$f(h,r,t) = \sum_{i=1}^{d} h_i \cdot r_i \cdot t_i \tag{5}$$

where d denotes the vector's dimensionality, and h_i, t_i, and r_i represent the i-th components of the **h**, **r** and **t**, respectively.

Then we adopt a margin-based loss function designed to guarantee that the positive triple is scored higher than its negative counterpart by a margin of at least γ.

$$Loss = \sum_{(h_i,r_k,t_i) \in G}^{|N|} \max(0, f(h_i,r_k,t_i) - f(\overline{h_i},r_k,\overline{t_i}) + \gamma) \tag{6}$$

where γ is a margin hyperparameter, (h,r,t)denotes a positive triple, and $(\overline{h},r,\overline{t})$ represents a negative triple obtained via random negative sampling.

3.3 Mutual Information Module

Mutual information (MI) serves as a metric for the divergence between two random variables X and Y . For discrete random variables, the formal definition is:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{7}$$

For continuous random variables,it is defined as:

$$I(X;Y) = \int_Y \int_X p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \int_Y \int_X p(y \mid x) p(x) \log \frac{p(y \mid x)}{p(y)} \tag{8}$$

Consequently, the mutual information between two random variables X and Y is equivalent to the Kullback-Leibler (KL) divergence between their joint distribution and the product of their marginal distributions.

$$I(X;Y) = D_{KL}\left(p(x,y) \,\|\, p(x)p(y)\right) \tag{9}$$

Where $p(x,y)$ denotes the joint distribution of variables x and y, $p(x)$ and $p(y)$ represent their marginal distributions, respectively.

From Equation (9), maximizing mutual information essentially aims to minimize the distance between the joint distribution and the product of independent distributions, thereby capturing high-order correlations between the variables. However, directly optimizing the KL divergence in high-dimensional continuous spaces faces challenges such as numerical instability and high variance. To address this, a common practice is to maximize a lower bound of mutual information. As noted in literature [13], three prevalent estimators for the mutual information lower bound are the Donsker-Varadhan (DV) representation, Jensen-Shannon divergence (JSD), and Information Noise-Contrastive Estimation (InfoNCE). The InfoNCE estimator, known for its lower variance within contrastive learning frameworks and its compatibility with stochastic gradient descent, is adopted in this paper.

The InfoNCE loss is built upon the noise-contrastive estimation framework. According to literature [14], the relationship between the InfoNCE loss and mutual information is given by:

$$I(X;Y) \geq LogN - Loss_{InfoNCE} \tag{10}$$

where N is the total number of samples. Therefore, minimizing the InfoNCE loss is equivalent to maximizing a computationally tractable lower bound of the mutual information $I(X;Y)$.

The primary objective of the InfoNCE loss is to ensure that the similarity score between a query sample and its positive sample is significantly higher than the sum of its similarity scores with all negative samples. Specifically, given a query x, a positive sample is y+, and k negative samples y_1^- , y_2^- …, y_k^-, the loss is defined as:

$$Loss_{InfoNCE} = -\log \frac{\exp(f(x, y^{+}) / \tau)}{\sum_{j=1}^{k} \exp(f(x, y_j^{-}) / \tau)} \tag{11}$$

where $f(x,y)$ is a similarity score function, typically the inner product or cosine similarity. The hyperparameter τ serves as a scaling factor that determines whether the output distribution is sharp or smooth.

In graph representation learning, InfoNCE loss is often computed using approaches such as global-local, graph-node, or multi-view. This paper employs the global-local approach to construct positive and negative samples. Specifically, for a given subgraph or neighborhood context, a global vector is generated by aggregating all node information using a readout function. Simultaneously, the representation of each entity or relation within the subgraph serves as a local vector. Contrastive learning tasks are subsequently formulated for both nodes and edges. Let g_e denote the global vector representation for entities, and e_i denote the local embedding for the i-th entity from the same subgraph. Similarly, let g_r denote the global vector representation for relations, and r_i denote the local vector for the i-th relation from the same subgraph. Positive and negative entity sample pairs are constructed as $(g_e,e_i)$and (g_e,e_j)(where e_j is an entity representation from other subgraphs within the same batch), respectively. Positive and negative relation sample pairs are constructed as $(g_r,r_i)$and (g_r,r_j)(where r_j is a relation representation from other subgraphs within the same batch), respectively. Based on these, the InfoNCE losses for entities and relations are computed separately, and their sum serves as the total contrastive loss.

3.4 Joint Training

The overall objective function is formulated as a weighted aggregation of the entity-relation interaction loss and the mutual information maximization loss. It is defined as:

$$L_{total} = Loss + \lambda \bullet Loss_{ML} \tag{12}$$

where λ balances the weight of the mutual information (MI) regularization term within the overall objective function. Through this joint training strategy, the model integrates local context with global structural information simultaneously.

4 Experiments

This section presents a comprehensive overview of the experimental configuration, including the datasets, baselines, evaluation metrics, and implementation details. We then compare the proposed model with the baselines on inductive link prediction tasks and evaluate its generalization ability to unseen relations. Finally, we present the ablation studies and an analysis of parameter sensitivity.

4.1 Experimental Configuration

DataSets. The experiment utilizes datasets proposed in INGRAM, which were obtained by modifying the standard benchmarks - Wikidata, FB15k-237, and NELL-995. The numerical suffix following each dataset name indicates the proportion of triples containing unknown relations. For example, FB-75 signifies that 75% of the triples in this dataset involve unknown relations, while the remaining 25% involve known relations.

Baselines and Parameters. We evaluate RGIILP against several state-of-the-art inductive link prediction approaches including Neural-Lp[7], DRUM[8], GraiL[1], CompILE[2], NBFNet[9], RED-GNN[10], INGRAM[4] and HyRel[15]. The experiments were conducted in a PyTorch environment under the Adam optimization scheme. The model underwent 6,000 iterations of training, starting with a learning rate of 5×10^{-4}. In the entity-relation interaction learning module, both the entity layer and the relation layer are composed of two layers. During training, ten negative sample was generated for each positive triple, and the mutual information loss accounts for 10% of the total loss.

4.2 Main Results

Experimental results for the inductive link prediction task, comparing RGIILP with baseline methods, are summarized in Tables 1-3. Due to the high computational complexity of node labeling in GraiL and CompILE, these models were not evaluated on the WK-100/75/50/25 and FB-100/75/50/25 datasets. Evaluation of the results leads to the following observations:

(1) Compared to existing baseline models (Neural-LP, DRUM, NBFNet, RED-GNN, INGRAM, and HyRel), the proposed RGIILP demonstrates a clear advantage in MRR and Hits@1. Specifically, RGIILP achieves the best MRR on 9 out of the 12 data subsets and leads in Hits@1 on 7 subsets. In terms of Hits@10, RGIILP obtains the highest score on 5 subsets; on the remaining 7 subsets, its performance is slightly below that of INGRAM and HyRel. These results confirm that RGIILP outperforms the baselines in overall ranking quality.

(2) Rule-based models (Neural-LP and DRUM) consistently yield the lowest performance across all 12 subsets, particularly in terms of MRR as well as Hits@1 and Hits@10. Moreover, as the percentage of unseen relations in the test set increases (from 25% to 100%), the values of MRR gradually decline, as do the scores for Hits@1 and Hits@10. This suggests that rule-based methods exhibit certain limitations when handling previously unseen relation types. Such models typically first mine rules from the graph structure and then employ them for relational prediction. For relations absent from the training set, no relevant rules can be learned; consequently, a higher proportion of unseen relations in the testing set leads to lower MRR, Hits@1 and Hits@10 scores.

(3) On the NELL-995 dataset, relation-graph-based models (RED-GNN, INGRAM, HyRel) consistently achieve superior performance over subgraph-based models (GraIL, CompILE), which in turn outperform rule-based models (Neural-LP, DRUM) across MRR, Hits@1, and Hits@10. This observed hierarchy can be attributed to the varying

capabilities of these approaches in capturing and utilizing relational semantics. Rule-based methods primarily rely on memorizing and composing frequent relational patterns observed during training, and consequently struggle to activate relevant inference rules when faced with unseen relation patterns. This leads to their limited generalization ability to unknown relations. In contrast, subgraph-based models not only learn structural patterns but also aggregate neighboring node and relation information through graph neural networks. This design endows them with a degree of analogical reasoning, enabling better handling of unseen relations compared to purely rule-based methods. Relation-graph-based models go a step further by explicitly learning relational representations alongside graph structures, which allows them to capture semantic associations and interaction patterns between relations. This capability supports effective transfer and composition of relational semantics, thereby facilitating more robust reasoning about unseen relations and resulting in the strongest overall performance.

(4)Compared to the second-best model INGRAM, RGIILP achieves average improvements of 7.23% in MRR and 7.63% in Hits@10 on NELL, 1.6% in MRR and 0.28% in Hits@10 on Wikidata, and 3.83% in MRR and 3.13% in Hits@10 on FB15k-237. These results demonstrate that the mutual-information and entity-relation co-training mechanisms effectively enhance inductive link prediction. Compared to the best-performing model HyRel, RGIILP shows gains on NELL and Wikidata, but slight declines on FB15k-237. This phenomenon can be ascribed to the inherent properties of the FB15k-237 dataset, which exhibits larger scale and higher relational density; in such settings, the mutual-information mechanism may introduce more noise or excessive smoothing, thereby degrading the model's effectiveness.

(5) On the three fully unseen datasets where both entities and relations are unknown (WK-100, NL-100, and FB-100), RGIILP surpasses the existing approaches in both MRR and Hits@1. This demonstrates the strong effectiveness of the mutual information enhanced entity-relation co-training approach for inductive link prediction involving completely unseen entities and relations.

Table 1. Inductive Link Prediction Performance(%) of RGIILP on WikiData, with best performances in bold

WIKIData		NeuralLP	DRUM	NBFNet	RED-GNN	INGRAM	HyRel	RGIILP
WK-100	MRR	0.9	1	1.4	9.6	10.7	9.1	**11.7**
	Hit@1	0.5	0. 4	0. 5	7	7.2	5.9	**7.5**
	Hit@10	1.6	1.9	2.6	13.6	16.9	16.5	**18**
WK-75	MRR	2	2	7.2	17.2	24.7	25.5	**27**
	Hit@1	0. 4	0. 7	2.8	11	17.9	18.7	**26**
	Hit@10	5.4	4.3	17.2	29	36.2	**38.9**	37.6
WK-50	MRR	2.5	1.7	6.2	5.8	6.8	6.8	**8.5**
	Hit@1	0.7	0.2	**3.6**	3.3	3.4	3.6	3.2
	Hit@10	5.4	4.6	10.5	9.3	13.5	13.8	**14**
WK-25	MRR	6.8	6.4	15.4	17	18.6	19.1	**20**
	Hit@1	4.6	3.5	9.2	11.1	12.4	12.5	**14**
	Hit@10	10.4	11.6	30.1	26.3	30.9	**31.6**	29

Table 2. Inductive Link Prediction Performance (%) of RGIILP on NELL995, with best performances in bold

NELL995		GraIL	COMPILE	NeuralLP	DRUM	NBFNet	RED-GNN	INGRAM	HyRel	RGIILP
NL-100	MRR	13.5	12.3	8.4	7.6	9.6	21.2	30.9	39.4	**42.5**
	Hit@1	11.4	7.1	3.5	4.4	3.2	11.4	21.2	**29.9**	29.4
	Hit@10	17.3	20.9	18.1	13.8	19.9	38.5	50.6	**57.4**	52.2
NL-75	MRR	9.6	17.8	11.7	15.2	13.7	20.3	26.1	30.5	**32.7**
	Hit@1	3.6	9.3	4.8	7.2	7.7	12.9	16.7	20.5	**21.9**
	Hit@10	20.5	36.1	27.3	31.3	25.5	35.3	46.4	50.2	**56.2**
NL-50	MRR	16.2	19.4	10.1	10.7	22.5	17.9	28.1	32.1	**33.7**
	Hit@1	10.4	12.5	6.4	7	16.1	11.5	19.3	22.2	**22.5**
	Hit@10	28.8	33	19	19.3	34.6	28	45.3	52	**56**
NL-25	MRR	21.6	18.9	14.8	16.1	28.3	21.4	33.4	34.8	**38.5**
	Hit@1	16	11.5	10.1	11.9	22.4	16.6	24.1	**26.3**	25.8
	Hit@10	36.6	32.4	27.1	26.4	41.7	26.6	50.1	54.1	**58.5**

Table 3. Inductive Link Prediction Performance (%) of RGIILP on FB15k237,with best performances in bold

FB15k237		NeuralLP	DRUM	NBFNet	RED-GNN	INGRAM	HyRel	RGIILP
FB-100	MRR	2.6	3.4	7.2	12.1	22.3	28.2	**30.1**
	Hit@1	0.7	1.1	2.6	5.3	14.6	18.8	**23.3**
	Hit@10	5.7	7.7	15.4	26.3	37.1	**46.3**	39
FB-75	MRR	5.6	6.5	8.9	10.7	18.9	**27.7**	23.4
	Hit@1	3	3.4	4.8	5.7	11.9	**19.6**	18.3
	Hit@10	9.9	12.1	16.6	20.1	32.5	**43.3**	43.1
FB-50	MRR	8.8	10.1	13	12.9	11.7	**17.8**	13
	Hit@1	4.3	6.1	7.1	7.2	6.7	**10.1**	8.3
	Hit@10	18.4	19.1	**25.9**	25.1	21.8	**33.3**	22
FB-25	MRR	16.4	17.5	**22.4**	14.5	13.3	21	15
	Hit@1	9.8	10.9	**13.7**	7.7	6.7	11.4	7
	Hit@10	30.9	32	41	28.4	27.1	**42**	26.9

5 Conclusion

The paper introduces RGIILP, an inductive link prediction framework specifically developed to infer missing links where relation types are previously unseen. To implement RGIILP, we first construct a relation graph corresponding to the knowledge graph. Subsequently, Joint embedding learning is then performed by integrating the knowledge graph with its corresponding relation graph, thereby fully capturing the information of both entities and relations. Simultaneously, mutual information between global and local representations is maximized to capture knowledge graph's global structure. Extensive experiments on multiple benchmark datasets demonstrate that RGIILP achieves superior performance compared to state-of-the-art inductive link prediction methods, especially in generalizing to previously unseen relations.

References

1. Teru, K., Denis, E., Hamilton, W.: Inductive relation prediction by subgraph reasoning. In: Proceedings of the 37th International Conference on Machine Learning. PMLR, vol. 119, pp. 9448–9457 (2020)
2. Mai, S., Zheng, S., Yang, Y., et al.: Communicative message passing for inductive relation reasoning. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press, Palo Alto, CA, USA (2021)
3. Chen, J., He, H., Wu, F., et al.: Topology-aware correlations between relations for inductive link prediction in knowledge graphs. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press, Palo Alto, CA, USA (2021)
4. Lee, J., Chung, C., Whang, J.J.: InGram: inductive knowledge graph embedding via relation graphs. In: Proceedings of the 40th International Conference on Machine Learning. PMLR, vol. 202, pp. 18796–18809 (2023)
5. Meilicke, C., Fink, M., Wang, Y., et al.: Fine-grained evaluation of rule- and embedding-based systems for knowledge graph completion. In: Proceedings of the 18th International Semantic Web Conference. LNCS, vol. 11136, pp. 3–20. Springer, Cham (2018)
6. Cohen, W.W.: Tensorlog: a differentiable deductive database. arXiv preprint arXiv:1605.06523 (2016)
7. Wang, P.W., Stepanova, D., Domokos, C., et al.: Differentiable learning of numerical rules in knowledge graphs. In: Proceedings of the 8th International Conference on Learning Representations (2020)
8. Sadeghian, A., Armandpour, M., Ding, P., et al.: Drum: End-to-end differentiable rule mining on knowledge graphs. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 15347–15357. Curran Associates, Inc., Red Hook, NY, USA (2019)
9. Zhu, Z., Zhang, Z., Xhonneux, L.P., et al.: Neural bellman-ford networks: a general graph neural network framework for link prediction. In: Proceedings of the 35th International Conference on Neural Information Processing Systems, pp. 29476–29490. Curran Associates, Inc., Red Hook, NY, USA (2021)
10. Zhang, Y., Yao, Q.: Knowledge graph reasoning with relational digraph. In: Proceedings of the ACM Web Conference 2022, pp. 912–924. ACM, New York, NY, USA (2022)
11. Geng, Y., Chen, J., Pan, J.Z., et al.: Relational message passing for fully inductive knowledge graph completion. In: Proceedings of the 39th IEEE International Conference on Data Engineering, pp. 1221–1233. IEEE (2023)
12. Veličković, P., Fedus, W., Hamilton, W.L., et al.: Deep graph infomax. In: Proceedings of the 7th International Conference on Learning Representations (2019)
13. Belghazi M I, Rajeswar S, Baratin A, et al. MINE: mutual information neural estimation[C]//Proceedings of the International Conference on Machine Learning, 2018:1-9.
14. Oord A, Li Y, et al. Representation learning with contrastive predictive coding[C]//Proceedings of the International Conference on Learning Representations (ICLR), 2019.
15. Yang J, Jiang X, et al. Generalize to fully unseen graphs: Learn transferable hyper-relation structures for inductive link prediction[C]//Proceedings of the ACM International Conference on Multimedia (MM), 2024：1274-1282.