ISDEA+: A SCALABLE DOUBLE-EQUIVARIANT GNN FOR ZERO-SHOT KNOWLEDGE GRAPH COMPLETION

Yucheng Zhang & Bruno Ribeiro
Department of Computer Science
Purdue University
West Lafayette, IN 47906, USA
{zhan4332,ribeirob}@purdue.edu

ABSTRACT

Pre-trained zero-shot knowledge graph completion is the task of predicting missing triplets among new entities with new relations that are not seen in training. ISDEA (Gao et al., 2023) was recently introduced to tackle this challenge by ensuring models are equivariant to both entity permutations and relation permutations —- a property denoted as double equivariance by Gao et al. (2023). However, ISDEA's message passing layer is complex, resulting in substantial computational overheads and limiting scalability. In our work, we propose ISDEA+, an efficient and scalable variant of ISDEA that maintains the double equivariance property through a new, streamlined message-passing scheme. ISDEA+ significantly enhances the training and inference speed, achieving over a 50-fold acceleration, reduces GPU memory usage by 30% while also attaining a 9.5% improvement in the Hits@10 metric on real-world knowledge graph datasets.

1 INTRODUCTION

Pre-trained zero-shot knowledge graph completion is crucial for handling open-world scenarios where knowledge graphs are constantly evolving and new entities and relations continuously emerge. Earlier studies in this domain primarily focused on the inductive task involving only new entities during inference (Zhu et al., 2021). However, recent works have considered pre-trained zero-shot models over both new entities and relations (Gao et al., 2023; Lee et al., 2023; Galkin et al., 2023).

A knowledge graph is represented as a tuple of (V, E, R) where V denotes a set of entities, R signifies a set of relations and E consists of a set of edges $(V \times R \times V)$. There is a training graph $G_{\text{train}} = (V_{\text{train}}, E_{\text{train}}, R_{\text{train}})$ and an inference graph $G_{\text{inf}} = (V_{\text{inf}}, E_{\text{inf}}, R_{\text{inf}})$. In the zero-shot settings of our interest, the training graph and the inference graph have totally different entity and relation sets – namely, $V_{\text{train}} \cap V_{\text{inf}} = \emptyset$ and $R_{\text{train}} \cap R_{\text{inf}} = \emptyset$. We consider two types of tasks in this scenario: entity prediction task $(i, k, ?)$, which involves predicting the tail entity given a head entity i and a query relation k; and relation prediction task $(i, ?, j)$, which aims to determine if there's a link between a given head entity i and a tail entity j, as well as identifying the specific relation type.

Contributions and related work. Among the existing zero-shot knowledge completion approaches, ISDEA (Gao et al., 2023) is the only one specifically designed to predict both missing tail entities $(i, k, ?)$ and missing relations $(i, ?, j)$. Galkin et al. (2023) uses a node marking approach that, to the best of our knowledge, is incompatible with missing relation predictions. ISDEA is equivariant to both entity permutations and relation permutations – namely, double-equivariant (Gao et al., 2023). However, owing to its intricate message passing layer, ISDEA is computationally expensive and lacks scalability when handling large graphs. In this paper, we introduce ISDEA+, a novel double-equivariant graph neural network. ISDEA+ maintains the double-equivariant property of ISDEA, while significantly enhancing efficiency and scalability compared to ISDEA. Finally we demonstrate that ISDEA+ exhibits improved performance on zero-shot knowledge graph completion tasks.

2 ISDEA+ FRAMEWORK

Given a query triplet (i, k, j), ISDEA+ computes entity embedding first. $X_{i,k}$ denotes the final entity embedding for entity i under relation k. In ISDEA+ with T layers, messages traverse along two predefined meta-paths (Wang et al., 2019; Yun et al., 2019) of length T: $P_k = v_1 \xrightarrow{k} v_2 \xrightarrow{\forall k' \in R} \dots \xrightarrow{\forall k' \in R} v_{T+1}$ and $P_{\neg k} = v_1 \xrightarrow{k' \neq k} v_2 \xrightarrow{\forall k' \in R} \dots \xrightarrow{\forall k' \in R} v_{T+1}$. This design is justified as a query triplet (i, k, j) essentially seeks to determine whether entity j is a 1-hop neighbor of entity i under relation k. Hence, during message passing, the 1-hop neighbor linked by relation k holds significant relevance to the head entity.

At the t-th layer, two vectors $h_{i,k}^{(t)}$ and $h_{i,\neg k}^{(t)}$ are learnt with the following message-passing scheme.

$$h_{i,k}^{(1)} = \text{GNN}_1^{(0)}\left(h_{i,k}^{(0)}, \left\{\!\!\left\{ h_{j,k}^{(0)} \middle| j \in \mathcal{N}_k(i) \right\}\!\!\right\}\right), \qquad h_{i,\neg k}^{(1)} = \text{GNN}_2^{(0)}\left(h_{i,\neg k}^{(0)}, \left\{\!\!\left\{ h_{j,\neg k}^{(0)} \middle| j \in \bigcup_{k' \neq k} \mathcal{N}_{k'}(i) \right\}\!\!\right\}\right), \quad \text{if } t=0,$$

$$h_{i,k}^{(t+1)} = \text{GNN}_1^{(t)}\left(h_{i,k}^{(t)}, \left\{\!\!\left\{ h_{j,k}^{(t)} \middle| j \in \bigcup_{k' \in R} \mathcal{N}_{k'}(i) \right\}\!\!\right\}\right), \quad h_{i,\neg k}^{(t+1)} = \text{GNN}_2^{(t)}\left(h_{i,\neg k}^{(t)}, \left\{\!\!\left\{ h_{j,\neg k}^{(t)} \middle| j \in \bigcup_{k' \in R} \mathcal{N}_{k'}(i) \right\}\!\!\right\}\right), \quad \text{if } t>0,$$

where $N_k(i)$ denotes the 1-hop neighbours of entity i with relation k. $h_{i,k}^{(0)}$ and $h_{j,\neg k}^{(0)}$ are initialized as all-one vectors, ensuring ISDEA+'s double-equivariance to both entity and relation permutations. Appendix A.1 describes the difference between ISDEA and ISDEA+'s message passing layers.

The embedding vectors from each layer are concatenated, and the learned embedding from the two meta-paths are summed together to form the final entity embedding $X_{i,k}$ of entity i under relation k: $X_{i,k} = \text{MLP}_1\left(h_{i,k}^{(0)} || h_{i,k}^{(1)} || \dots || h_{i,k}^{(T)}\right) + \text{MLP}_2\left(h_{i,\neg k}^{(0)} || h_{i,\neg k}^{(1)} || \dots || h_{i,\neg k}^{(T)}\right)$. ISDEA+ computes triplet embedding $\Gamma(i, k, j)$ by concatenating entity embedding and the directed shortest distance $d(i, j)$ and $d(j, i)$ between two entities: $\Gamma(i, k, j) = X_{i,k} || X_{j,k} || d(i, j) || d(j, i)$.

3 Experiments

This section aims to address two key questions: **Q1:** Is ISDEA+ capable of scaling effectively to large graphs? **Q2:** How is the performance of ISDEA+ compared with ISDEA on zero-shot knowledge graph completion task?

For addressing **Q1**, we conducted experiments on PediaTypes (Gao et al., 2023), recording the time taken for computation and the GPU memory usage. Table 1 presents the results. Compared to other models, ISDEA+ is the fastest during training and requires the least GPU memory.

Table 1: Runtime (training and inference) and GPU memory usage on PediaTypes. ISDEA+ achieves 50 times speed up compared with ISDEA.

Models	Training Time	Inference Time	GPU Memory
InGram (Lee et al., 2023)	0.09s	**5.60s**	3.0GB
DEq-InGram (Gao et al., 2023)	0.09s	10.39s	3.0GB
ISDEA (Gao et al., 2023)	2.56s	359.45s	9.8GB
ISDEA+ (Ours)	**0.05s**	6.53s	**1.9GB**

To address **Q2**, we conduct experiments on PediaTypes. Table 2 showcases the Hits@10 results for relation prediction against 50 negative triplets, comparing ISDEA+ with other inductive models. The "A-B" dataset format denotes training on dataset A and zero-shot testing on dataset B, involving entirely unseen entities and relations. ISDEA+ demonstrates average 9.5% Hits@10 improvement over previous zero-shot models that can perform relation prediction tasks. Appendix A.4 shows results on entity prediction tasks.

Table 2: Relation Prediction Hits@10 on PediaTypes. ISDEA+ is the best on 7 out of 8 datasets.

Models	EN-FR	FR-EN	EN-DE	DE-EN	DB-WD	WD-DB	DB-YG	YG-DB
InGram	78.74$\pm$07.48	62.11$\pm$13.60	48.72$\pm$08.94	65.60$\pm$14.42	77.75$\pm$06.60	63.32$\pm$02.78	67.98$\pm$25.45	64.98$\pm$26.69
DEq-InGram	87.94$\pm$05.68	80.47$\pm$09.90	68.89$\pm$05.45	80.79$\pm$10.51	91.47$\pm$01.53	77.03$\pm$04.09	77.72$\pm$21.92	89.30$\pm$05.53
ISDEA	84.94$\pm$05.00	84.75$\pm$02.51	95.26$\pm$00.63	94.23$\pm$00.71	82.22$\pm$02.44	88.87$\pm$02.94	**91.42**$\pm$01.79	85.34$\pm$01.49
ISDEA+	**99.12**$\pm$00.24	**98.84**$\pm$00.06	**99.20**$\pm$00.13	**98.99**$\pm$00.12	**98.56**$\pm$00.12	**98.03**$\pm$00.17	88.78$\pm$03.23	**96.45**$\pm$00.24

4 Conclusions

In this work we proposed ISDEA+, a more scalable double equivariant ISDEA (Gao et al., 2023) that can also perform both entity and relation prediction (improving the capabilities of existing double equivariant zero-shot models such as Lee et al. (2023); Galkin et al. (2023)). ISDEA+ shows improved performance in zero-shot knowledge graph completion tasks for both relation and entity prediction tasks. In our future work, we aim to extend our experiments to encompass a broader range of knowledge graph datasets.

REFERENCES

Fabrizio Frasca, Emanuele Rossi, Davide Eynard, Ben Chamberlain, Michael Bronstein, and Federico Monti. Sign: Scalable inception graph neural networks. *arXiv preprint arXiv:2004.11198*, 2020.

Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards foundation models for knowledge graph reasoning. *arXiv preprint arXiv:2310.04562*, 2023.

Jianfei Gao, Yangze Zhou, Jincheng Zhou, and Bruno Ribeiro. Double equivariance for inductive link prediction for both new nodes and new relation types. *arXiv preprint arXiv:2302.01313*, 2023.

Jaejun Lee, Chanyoung Chung, and Joyce Jiyoung Whang. Ingram: Inductive knowledge graph embedding via relation graphs. *arXiv preprint arXiv:2305.19987*, 2023.

Haggai Maron, Or Litany, Gal Chechik, and Ethan Fetaya. On learning sets of symmetric elements. In *International conference on machine learning*, pp. 6734–6744. PMLR, 2020.

Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The world wide web conference*, pp. 2022–2032, 2019.

Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. Graph transformer networks. *Advances in neural information processing systems*, 32, 2019.

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 APPENDIX

A.1 COMPARISON WITH ISDEA

Both ISDEA and ISDEA+ incorporate the DSS layers introduced by Maron et al. (2020). Specifically, when updating $h_{i,k}^{(t)}$ both models handle relation k and other relations $k' \neq k$ separately, employing two distinct graph neural networks. However, their divergence lies in the message passing layers. In ISDEA, entity embeddings are updated using the following formula:

$$h_{i,k}^{(t+1)}=\mathrm{GNN}_1^{(t)}\left(h_{i,k}^{(t)},\left\{\!\!\left\{\, h_{j,k}^{(t)} \,\middle|\, j\in\mathcal{N}_k(i) \right\}\!\!\right\}\right)+\mathrm{GNN}_2^{(t)}\left(\sum_{k'\neq k}^{R} h_{i,k'}^{(t)},\left\{\!\!\left\{ \sum_{k'\neq k}^{R} h_{j,k'}^{(t)} \,\middle|\, j\in \bigcup_{k'\neq k}\mathcal{N}_{k'}(i) \right\}\!\!\right\}\right)$$

The primary drawback in the scalability of ISDEA is the interdependence between the embedding $h_{i,k}^{(t)}$ for entity i under relation k and the embeddings of other relations $k' \neq k$ (specifically $h_{i,k'}^{(t)}$ and $h_{j,k'}^{(t)}$). Consequently, during a single forward pass, ISDEA computes $h_{i,k}^{(t)}$ for all $k \in R$ even though only one of these embeddings is directly used for the query triplet.

ISDEA+ effectively addresses this limitation by disentangling $h_{i,k}^{(t)}$ from $h_{i,k'}^{(t)}$. However, ISDEA+ retains the capability to distinguish between relation k and other relations with $h_{i,\neg k}^{(t)}$.

A.2 TIME COMPLEXITY ANALYSIS

The number of entities is represented by $|V|$; edges by $|E|$ and relations by $|R|$. Suppose entity embeddings have d dimensions and T message passing layers are employed.

Mini-batch training is the de facto mechanism used in modern deep learning algorithm. Suppose the batchsize is B. Then in one training step, ISDEA has time complexity of $O(T|R|(|E|d + |V|d^2))$. In one mini-batch there are at most B distinct query relations. The time complexity of ISDEA+ is at most $O(TB(|E|d + |V|d^2))$ for every step which is $|R|/B$ times faster than ISDEA.

In our implementation, we make sure that each mini batch contains precisely one relation (details in A.3). Then the time complexity of ISDEA+ is $O(T(|E|d + |V|d^2))$ and it is $|R|$ times faster than ISDEA.

A.3 Implementation Details

We use SIGN (Frasca et al., 2020) as the message passing layer in ISDEA+:

$$X_{i,k}^{(t+1)} = X_{i,k}^{(t)} W_1^{(t)} + \left(\sum_{j \in N(i)} \frac{1}{\sqrt{d_i d_j}} X_{j,k}^{(t)} \right) W_2^{(t)},$$

where d_i is the degree of entity i; $W_1^{(t)}$ and $W_2^{(t)}$ are parameters to be learnt at layer t. Following SIGN, we've omitted the utilization of non-linear activation functions to expedite both the training and inference processes.

ISDEA+ employs 3 message passing layers with 32 neurons in each layer. The model is trained for 50 epochs using the Adam optimizer with a batch size of 256 and a learning rate of 0.0001.

We employ the same negative sampling strategy following ISDEA. For every positive triplet (i, k, j), we corrupt the tail entity to create negative triplet (h, r, t') for n_t times and corrupt the relation to create negative triplet (h, r', t) for n_r times. During training, n_t and n_r are set to 2 and binary cross-entropy loss is used as loss function. However, during testing, n_t and n_r are set to 50. The evaluation metrics employed include Mean Reciprocal Rank (MRR) and Hits@10.

To speed up training, negative sampling is conducted prior to the mini-batching process, ensuring each mini-batch comprises only one relation. Consequently, different mini-batches might contain varying quantities of positive and negative triplets, yet all associated triplets within a batch share the same relation. As a result, during each training step, we execute the forward pass just once.

A.4 More Experiment Results

In this part we show the Hits@10 result of entity prediction on PediaTypes.

Table 3: Entity Prediction Hits@10 on PediaTypes. ISDEA+ is the best on 4 out of 8 datasets.

Models	EN-FR	FR-EN	EN-DE	DE-EN	DB-WD	WD-DB	DB-YG	YG-DB
InGram	78.74±07.48	62.11±13.60	48.72±08.94	65.60±14.42	77.75±06.60	63.32±02.78	67.98±25.45	64.98±26.69
DEq-InGram	94.47±00.60	**88.90**±02.06	93.85±00.36	94.02±00.74	71.94±07.37	**91.47**±00.62	71.53±04.78	**80.53**±07.96
ISDEA	76.28±00.05	77.51±01.46	82.24±00.94	81.80±00.68	66.69±01.01	75.19±03.12	**72.87**±01.03	76.41±01.52
ISDEA+	**95.39**±00.30	81.57±03.17	**97.66**±00.19	**95.03**±00.44	**86.60**±00.59	90.93±00.24	69.62±01.10	73.16±00.82