Subgraph-aware Few-Shot Inductive Link Prediction via Meta-Learning

Shuangjia Zheng[1], Sijie Mai[1], Ya Sun[1], Haifeng Hu, Yuedong Yang

***Abstract*—Link prediction for knowledge graphs aims to predict missing connections between entities. Prevailing methods are limited to a transductive setting and hard to process unseen entities. The recently proposed subgraph-based models provide alternatives to predict links from the subgraph structure surrounding a candidate triplet. However, these methods require abundant known facts of training triplets and perform poorly on relationships that only have a few triplets. In this paper, we propose Meta-iKG, a novel subgraph-based meta-learner for few-shot inductive relation reasoning. Meta-iKG utilizes local subgraphs to transfer subgraph-specific information and to rapidly learn transferable patterns via meta-gradients. In this way, we find the model can quickly adapt to few-shot relationships using only a handful of known facts with inductive settings. Moreover, we introduce a large-shot relation updating procedure to ensure that our model can generalize well to both few-shot and large-shot relations. We evaluate Meta-iKG on inductive benchmarks sampled from the NELL and Freebase, and the results show that Meta-iKG outperforms the currently state-of-the-art methods in both few-shot scenarios and standard inductive settings.**

***Index Terms*—Inductive relation prediction, meta-learning, knowledge graph, subgraph scoring**

I. Introduction

Knowledge graphs (KGs) are repositories of large numbers of triplets in the form of relations between two entities, encoding knowledge and facts in the world. This kind of graph-structured knowledge has played a critical role across a wide range of tasks such as Semantic Search [1], Question Answering [2], and many more. However, due to the limitations of human knowledge and information extraction algorithms, they typically suffer from incompleteness, that is, absent links in the KGs. To automate the KG completion process, numerous latent representation methods have been proposed to condense each entity and relation into a low-dimensional and continuous vector space, which can then be utilized to infer missing links by operating the produced embeddings [3]–[7].

While these embedding-based models have shown promising performance, prevailing methods typically assume a fixed set of entities in the graph and neglect the evolving character of KGs. However, real-world KGs are often dynamic and ever-evolving [8], with new entities being added at any given moment, e.g., new users on online shopping platforms. Recently, taking inspiration from the success of graph neural networks (GNNs) in graph structure modeling [9], a few efforts have been made to subgraph-based inductive relation prediction, combining the beneficial qualities of both scalability and interpretability [10], [11]. The basic strategy behind this type of models is to score a target triplet based on its enclosing subgraph. It can facilitate the prediction of completely novel entities that are not surrounded by known nodes, since the domain-related initial embedding of these emerging entities is not required in the modeling.

Despite the impressiveness of model performance, this framework assumes that there are enough triplets to train robust and effective reasoning models for each relation in KGs, as GNNs usually need substantial instances to enable the model stability [12]. Nonetheless, previous works have observed that a large portion of KG relations are actually long-tailed [13], [14] and only occur a handful of times, which can be referred to as *few-shot* relations. Some pilot experiments have demonstrated that the few-shot scenario incurs the infeasibility of GNN models, resulting in catastrophic performance decline on those few-shot classes [15], [16]. The inability to handle the presence of very few samples is one of the major challenges for the current GNNs. To perform the few-shot link prediction task, several attempts [1], [17] have incorporated graph structured data into meta learning. They achieve fairly good performance in few-shot relation prediction yet still have two main limitations: (i) they only encode one-hop neighbors of entities and neglect the high-order neighborhood information around a target triplet; (ii) they are limited to transductive settings and cannot process unseen entities. In fact, few-shot relations often appear when new entities are encountered. Therefore, few-shot inductive relation reasoning is a non-trivial task of considerable importance that has remained under-explored.

Present Work. To address the aforementioned challenges, we propose Meta-iKG, a novel subgraph-based meta learner for few-shot inductive relation reasoning. Meta-iKG utilizes local subgraphs to transfer subgraph-specific information and to rapidly learn transferable patterns via meta-gradients. Specifically, we first translate link prediction as a subgraph modeling problem. Then, we regard triplet queries with the same relation r in KGs as a single task. Following the previous meta-learning paradigm [18]–[20], we use tasks of high-frequency relations to construct a meta-learner, which includes common features across different tasks. The meta-learner can be fast adapted to the tasks of few-shot relations, by providing a good initial point to train their relation-specific subgraph scoring functions. Moreover, different from standard meta-

Yuedong Yang and Haifeng Hu are with Sun Yat-sen University, China. E-mail: {yangyd25,huhaif}@mail.sysu.edu.cn

Shuangjia Zheng, Sijie Mai, and Ya Sun are with Sun Yat-sen University. E-mail: {zhengshj9, maisj, suny278}@mail2.sysu.edu.cn

[1]These authors contribute equally to this work.

This study has been supported by the National Key R&D Program of China (2020YFB0204803), Guangzhou S&T Research Plan (202007030010).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3177212, IEEE Transactions on Knowledge and Data Engineering

learning, we introduce a large-shot relation update procedure to eliminate the bias introduced by the few-shot relational meta-updating, which enables our Meta-iKG to generalize well to both large-shot and few-shot relations. We evaluate Meta-iKG on several novel inductive link prediction benchmarks sampled from NELL and Freebase, and experimental results show that Meta-iKG outperforms the state-of-the-art methods in both few-shot and standard inductive settings.

In brief, the main contributions are listed below:

- Introducing an inductive few-shot relation prediction problem which is different from previous works and more suitable for practical scenarios.
- Proposing a novel few-shot subgraph embedding model, Meta-iKG, that fits the few-shot nature of knowledge graph and can naturally generalize to the unseen entities.
- Experiments on eight inductive datasets demonstrate that our model achieves state-of-the-art AUC-PR and Hits@10 across most of them in both few-shot and standard inductive settings.

II. Related Work

Inductive relation prediction. A close research line is the rule-based approach that utilizes the observed co-occurrence of frequent patterns in the knowledge graph to identify logical rules [21]. Inspired by the statistical rule-induction approach, some differentiable rule learners including RuleN [22], NeuralLP [23], and DRUM [24] are introduced to simultaneously learn the logical rules and confidence scores in an end-to-end paradigm. However, they do not consider the neighborhood structure around the predicted triplets and thus are less expressive. Recent studies incorporate graph neural network into inductive relation reasoning to capture multi-hop information around the target triplet. GraIL [10], for example, proposes a subgraph-based relation reasoning framework to process unseen entities and CoMPILE [11] extends the idea by introducing a node-edge communicative message passing mechanism to model the directed subgraphs, which fits the directional nature of KGs. Our study can be interpreted as an extension of CoMPILE method to few-shot knowledge graph completion.

Few-shot relation prediction. Currently, the few-shot learning models mainly fall into four strategies [25]: (i) methods based on data augmentation [26]; (ii) methods based on learning of analogy tasks [27]; (iii) metric-based approaches [28], [29] and (iv) meta-optimizer based approaches [18], [30], [31]. Most few-shot relation prediction methods in the field of KG embedding are dominated by the last two types strategies. For example, GMatching [1] use metric methods to generalize over new relations from a handful of associative relations in a knowledge graph. Another line of research like MetaR [17] proposes to optimize the model parameters given the gradients on few-shot instances. They employ the model-agnostic meta-learning (MAML) [18] to train model by a small number of gradient updates, leading to the fast adaption on a new task. A few attempts have also been proposed that combine meta-learning with multi-hop reinforcement learning [32] or sequential network [33] to perform few-shot link prediction, while all of them only perform transductive relation prediction and cannot be directly transferred to the inductive setting. A key distinction between previous works and the one which we consider in this work is that we explicitly face the problem of few-shot scenario for inductive relation reasoning, which is more challenging as it needs to generalize to new types of relations. To the best of our knowledge, this work is the first research on few-shot learning for inductive relation reasoning.

III. Method

A. Formulation and Model Overview

A triplet in a KG is denoted as (s, r, t) where s, r, and t refers to the head entity, relation, and tail entity, respectively. Inductive relation prediction aims to evaluate the plausibility of a target triplet (s_T, r_T, t_T), where the embeddings of s_T and t_T are unavailable during reasoning. In this paper, we aim to enable the model to generalize well on the relations that only have few training triplets. Firstly we split the relations into few-shot and large-shot relations. If the number of triplets including a relation r is fewer than a threshold K_T, we denote r as a *few-shot relation*, otherwise, it is a *large-shot (normal) relation*. Following the idea of meta-learning [1], [17], [18], [32], we train triplets with large-shot relations to find well-initialized parameters and adapt the model on triplets with few-shot relations from the found initial parameters. An overview of our Meta-iKG can be seen in Fig. 1. In particular, Meta-iKG can be divided into two modules: (i) relation-specific learning and (ii) meta-learning. The purpose of relation-specific learning is to learn a GNN model with parameters $\boldsymbol{\theta}_r$ using a set of subgraphs surrounding a specific relationship r to identify the transferable patterns. Meta-learning is based on the relation-specific module for learning a meta model with parameters $\boldsymbol{\theta}$ and enables fast adaptation for new few-shot tasks. We introduce these two parts in following sections.

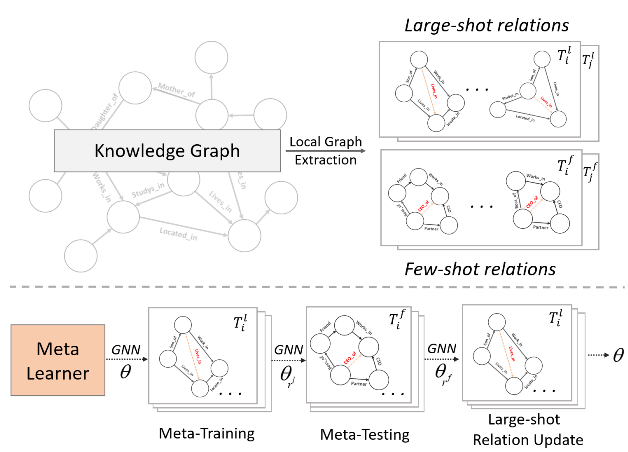

Fig. 1: Overview of Meta-iKG. A) Extracting local enclosing subgraphs around target entities. B) Meta-learner adapts to few-shot and large-shot relations via three-step optimization.

B. Subgraph-aware Relation-specific Learning

For each relation $r_T \in R$, we learn a relation-specific subgraph scoring function using the triplets with the relation

r_T to make predictions. Our subgraph-based scoring function consists of three subtasks: (i) extracting the enclosing subgraph surrounding two target entities, (ii) labeling the entities in the subgraph, and (iii) scoring the labeled subgraph using GNNs.

Subgraph extraction: We first extract the h-hop directed enclosing subgraph G between the target head and tail entities, where $h+1$ is the maximum distance from target head to target tail. This is based on the assumption that the paths connecting the two target entities include the information that can infer the target relation [10]. We refer the reader to Appendix for the details of subgraph extraction.

Inductive node labeling function: An inductive node labeling function is applied to mark nodes' different roles in the subgraph without leveraging external domain features. As such, the model has to learn the structural semantics in the subgraph, which is the key reason for the inductive nature of our model. Following [2], we initialize the node embedding $\boldsymbol{N}$ by the distances to the target entities to embed the relative position of each node in the subgraph. In detail, for a node i, its label is defined as $\boldsymbol{N}_i = \text{one-hot}(d_{si}) \oplus \text{one-hot}(d_{it}) \in \mathbb{R}^{2(h+1)}$ where d_{si} denotes the shortest distance from the target head entity to node i, and d_{it} denotes the shortest distance from node i to target tail entity. Following GraIL [10], the relative positions of s_T and t_T are labeled (0,1) and (1,0) so as to be identifiable by the model.

Directed Subgraph Scoring: In principle, our framework can be combined with a wide variety of GNN-based approaches, and here we focus on communicative message passing neural network following the idea of CoMPILE [11]. Formally, given a target triplet (s_T, r_T, t_T) with the enclosing subgraph G, the subgraph scoring function can be defined as:

$$S = \text{GNN}(G, \boldsymbol{N}_{s_T}, \boldsymbol{R}_{r_T}, \boldsymbol{N}_{t_T}; \boldsymbol{\theta}) \tag{1}$$

where S denotes the predicted plausibility of the target triplet, $\boldsymbol{N}_{s_T}$, $\boldsymbol{R}_{r_T}$, and $\boldsymbol{N}_{t_T}$ denotes the embedding of target head, target relation, and target tail, respectively. The relation embedding $\boldsymbol{R} \in \mathbb{R}^{N_r \times d}$ (N_r is the number of relations) is parameterized as a learnable matrix shared across subgraphs, and each row in $\boldsymbol{R}$ represents the embedding for a specific relation. Please refer to the Appendix for the details of GNN.

We use the following equation to compute the task loss:

$$L^{D}_{r_T}(\boldsymbol{\theta}) = \log(\exp(-\hat{S} \cdot S) + 1) \tag{2}$$

where $L^{D}_{r_T}(\boldsymbol{\theta})$ is the predictive loss for relation r_T with data $D = \{(s_T, r_T, t_T)\}$. $L^{D}_{r_T}(\boldsymbol{\theta})$ is formatted as the soft-margin loss. $\hat{S}$ denotes the true score of the triplet.

C. Meta-Learning

The goal of meta-learning is to learn well-initialized parameters, such that small changes in the parameters will produce significant improvements on the loss function of any task [18]. In this section, we describe our meta-learning strategy in detail which enables our model to generalize well on both few-shot and large-shot relations. Formally, we consider a meta GNN, i.e. CoMPILE with parameters $\boldsymbol{\theta}$. Firstly we divide the relations into few-shot and large-shot relations according to the threshold: $K_T = n_T/n_R \times \gamma$ where n_T denotes the number of training triplets, n_R denotes the number of relations, and γ is a scalar. In each iteration, we sample a batch of relations r^l from large-shot relation set $R^l = \{r | n_r > K_T, r \in R\}$ and a batch of relations r^f from few-shot relation set $R^f = \{r | n_r \leq K_T, r \in R\}$. Then we sample triplets belonging to r^l and r^f to form the support set D_S and query set D_Q, respectively.

The model is firstly updated by the support set D_S in large-shot tasks T^l, after which the parameters of the model become $\boldsymbol{\theta}_{r^l}$. Following MAML [18], the updated parameters $\boldsymbol{\theta}_{r^l}$ are computed using one or more gradient descent updates on the few-shot tasks T^f. Formally, we will go over many large-shot tasks (i.e., meta-training tasks) in a batch:

$$\boldsymbol{\theta}_{r^l} = \boldsymbol{\theta} - \boldsymbol{\alpha} \cdot \sum_{T^l_i} \nabla_{\boldsymbol{\theta}} L^{D_S}_{r^l_i}(\boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{\alpha}$ is the learning rate parameter, D_S is a support set randomly sampled from the triplets belonging to large-shot relations r^l for task T^l. After the relation-specific parameters $\boldsymbol{\theta}_{r^l}$ is learned, we evaluate $\boldsymbol{\theta}_{r^l}$ on the query set D_Q belonging to few-shot relations r^f for task T^f. The meta-gradient computed from this evaluation is used to update the meta policy network CoMPILE with parameters $\boldsymbol{\theta}$. Specifically, we update $\boldsymbol{\theta}$ using the few-shot tasks T^f (i.e., meta-testing tasks) as follows:

$$\boldsymbol{\theta}_{r^f} = \boldsymbol{\theta} - \beta \sum_{T^f_i} \nabla_{\boldsymbol{\theta}} L^{D_Q}_{r^f_i}(\boldsymbol{\theta}_{r^l}), \tag{4}$$

where β is the meta-learning rate. Eq. 4 computes the meta-gradients of $\boldsymbol{\theta}$ across two steps using the query set belonging to few-shot relations to finally update the parameters. By this means, Meta-iKG can learn to fast adapt to the few-shot relations with the aim of the well-initialized parameters $\boldsymbol{\theta}_{r^l}$ updated by the support set of large-shot relations.

The above two steps are the regular operations of meta-learning. While promising on the few-shot relations, these traditional operations of meta-learning will introduce bias to the updated parameters, because the final updated parameters $\boldsymbol{\theta}_{r^f}$ are only updated by the query set belonging to few-shot relations (see Eq. 4). Therefore, the model may not perform well on large-shot relations, as the support set belonging to the large-shot relations is only used to provide a good initialization of the parameters for computing the meta-gradients. To this end, we introduce a novel updating operation on $\boldsymbol{\theta}_{r^f}$ using the support set D_S with a smaller learning rate $\beta^{'}$, named 'large-shot relation update procedure':

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_{r^f} - \beta^{'} \sum_{T^l_i} \nabla_{\boldsymbol{\theta}_{r^f}} L^{D_S}_{r^l_i}(\boldsymbol{\theta}_{r^f}), \tag{5}$$

Combining Eq. 4 and Eq. 5, the final updated parameters depend on both the few-shot and large-shot relations. This simple operation enables Meta-iKG to generalize well on the whole inductive dataset. Notably, we apply Meta-SGD [19] to update the meta-learner, in which case the learning rate parameter $\boldsymbol{\alpha}$ is meta-learnable.

IV. Experiments

A. Dataset

We use the inductive datasets extracted from FB15K-237 [34] and NELL-995 [1]. Theoretically, in the inductive setting, both rule-based and subgraph-based algorithms rely on the paths between the target entities to infer relation. If there is no enclosing subgraph between the target entities, the model has no information to infer the relation but relies on the bias in the dataset to make prediction, which makes no sense computationally. Therefore, we follow CoMPILE [11] to extract the inductive datasets which have filtered out the triplets that have no enclosing subgraph between the target entities under hop h and ensure more accurate evaluation of the models. Note that CoMPILE extracts three versions of inductive datasets for each dataset, while we extract a new version of inductive datasets for FB15k-237 and Nell-995 (i.e., the v4), respectively. The inductive datasets are the filtered versions of the corresponding inductive datasets extracted in GraIL [10]. Since we use more challenging datasets, the performance of the models is generally weaker than the performance reported in GraIL [10]. Please refer to the Appendix for the statistics of the filtered datasets.

B. Experimental Details

Evaluation Protocol: To be consistent with prior methods, we use AUC-PR and Hits@10 to evaluate the models. To compute AUC-PR, for each test triplet, we sample one negative triplet and evaluate which triplet has larger score. This procedure is repeated for ten times to obtain the average AUC-PR for each run. For Hits@10, we rank each test triplet among the sampled 49 negative head/tail triplets and evaluate whether test triplet score makes it into top 10. The negative triplet is obtained by replacing the head or tail of the test triplet with other entities. To train the model, we assign 1 and -1 as the score of true and negative triplet respectively. We run the model for five times and average the testing results to obtain the final performance. We observe that the Mata-iKG has stable performance across a wide range of experimental settings, and the variances are quite low. Therefore, we do not present the standard errors in the tables.

Hyper-parameter Setting: We run the models for 20 epochs in the standard inductive datasets, and each epoch contains 100 meta-updates. We use Adam [35] as optimizer with learning rate being 0.001. The hop number h is 3 which is consistent with GraIL and CoMPILE. The number of GNN layers l is set to 3 or 4 which depends on the datasets. The few-shot factor γ is set to 0.1 at the training stage. For the sake of complexity, the subgraph will be pruned if it contains too many nodes, and we ensure that the pruned subgraph can also have a complete path between target entities.

C. Baselines

We compare Meta-iKG with the state-of-the-art subgraph-based inductive models GraIL [10] and CoMIPLE [11] as well as the rule-based algorithm RuleN [22]. For meta-learning strategy, we use both MAML [18] and Meta-SGD [19] as two different meta-learning strategies to create model variants.

D. Comparison on Few-Shot Testing Set

In this section, we first evaluate Meta-iKG on few-shot relations, where the relations whose number of training triplets K is fewer than 5, 10, and K_T are selected to evaluate the stability and robustness on different settings of few-shot inductive relation prediction (K_T is the threshold to split the large-shot and few-shot relations in training stage). For comparison, we also present the results of subgraph-based baselines CoMPILE [11] and GraIL [10]. According to Table I and II, we find that (a) both versions of Meta-iKG significantly outperform CoMPILE and GraIL on the majority of settings and datasets, and the improvement is much larger compared to that in the standard inductive datasets (see Table III and Table IV for the results on standard inductive datasets); (b) although the MAML version of Meta-iKG performs weaker than CoMPILE and GraIL on some standard datasets, it manages to outperform the baselines significantly on the few-shot scenarios. These results demonstrate the effectiveness and stability of Meta-iKG in few-shot inductive relation prediction.

E. Comparisons on Standard Inductive Datasets.

To clarify the importance of improving the predictive power of few-shot relations, we compare our proposed Meta-iKG with other baselines on the standard inductive datasets. From the Table III and Table IV, we can conclude that: (a) Meta-iKG, especially the Meta-SGD version, achieves the best performance on the majority of the inductive datasets in terms of both the AUC-PR and Hits@10 evaluation metrics by a significant margin; (b) the MAML version of our Meta-iKG performs worse than the Meta-SGD version, demonstrating the importance of the meta-trainable learning rate in the fast adaption of meta-learning. Compared to the MAML version that just provides a good initialization of the model parameter, the Meta-SGD version also learns the update direction and learning rate of the model parameter, leading a stronger capacity. These results demonstrate the effectiveness of our meta-learning strategy in inductive relation prediction, which can generalize well on both large-shot and few-shot relations via the updating procedure in Eq. 3, Eq. 4 and Eq. 5. Meta-iKG enables a better prediction capacity on few-shot relations without sacrificing the performance on overall datasets.

F. Analysis on the Model Complexity

Since the meta-learning algorithm does not introduce additional module to process the subgraph but merely modifies the optimization algorithm of the model, the time complexity remains unchanged compared to the original model during prediction. However, due to the flexible size of the subgraph, the common proxy to evaluate the time complexity of the model, i.e., FLOPs (which computes the number of operations go through in the forward pass given a sample), is changeable and depends on the subgraph. As for the space complexity, for Meta-iKG (MAML), since it introduces no additional parameters, the number of parameters is the same as that of CoMPILE (which is 41,185 when the dimensionality of relation embedding d is 32 and the number of layers l is 3 for FB15k-237-v1 dataset). For Meta-iKG (Meta-SGD), it introduces learnable learning rate parameter $\boldsymbol{\alpha}$ which has the same size as the parameter of the meta-learner (i.e., CoMPILE). Notably, the $\boldsymbol{\alpha}$ is not added to the model itself,

	FB15k-237-v1			FB15k-237-v2			Nell-995-v1	Nell-995-v2		
Model	$K \leq 5$	$K \leq 10$	$K \leq K_T$	$K \leq 5$	$K \leq 10$	$K \leq K_T$	$K \leq K_T$	$K \leq 5$	$K \leq 10$	$K \leq K_T$
GraIL	50.00	37.50	43.79	83.33	80.00	80.77	0.00	64.00	66.67	65.71
CoMPILE	43.75	42.86	46.43	80.08	80.00	76.92	0.00	78.18	72.41	67.14
Meta-iKG (MAML)	**75.00**	56.25	**57.14**	**86.67**	88.33	**88.46**	**50.00**	**80.00**	74.36	72.50
Meta-iKG (Meta-SGD)	**75.00**	**60.71**	53.57	**86.67**	**90.00**	**88.46**	**50.00**	78.00	**78.21**	**76.25**

TABLE I: **Comparison on Few-shot Relations (Hits@10).** For Nell-995-v1, since there is no relation whose number of training triplets is fewer than 10, we only present the result of $K \leq K_T$ setting.

	FB15k-237-v3			FB15k-237-v4	Nell-995-v3			Nell-995-v4		
Model	$K \leq 5$	$K \leq 10$	$K \leq K_T$	$K \leq K_T$	$K \leq 5$	$K \leq 10$	$K \leq K_T$	$K \leq 5$	$K \leq 10$	$K \leq K_T$
GraIL	54.67	56.29	67.75	67.86	53.17	63.81	64.88	68.78	68.73	68.88
CoMPILE	58.33	57.14	70.00	68.57	**65.87**	70.95	70.25	72.50	72.20	72.55
Meta-iKG (MAML)	**66.67**	**64.29**	**80.00**	**76.42**	**65.87**	71.43	71.07	74.25	73.90	74.34
Meta-iKG (Meta-SGD)	58.33	57.14	73.33	73.57	**65.87**	**72.86**	**73.14**	**78.63**	**78.29**	**78.28**

TABLE II: **Comparison on Few-shot Relations (Hits@10).** For FB15k-237-v4, since there are few testing triplets that belong to the relations whose number of training triplets is fewer than 10, we only present the $K \leq K_T$ case.

	FB15k-237				NELL-995			
Model	v1	v2	v3	v4	v1	v2	v3	v4
RuleN	79.60	82.67	83.03	84.01	67.12	80.52	73.91	77.07
GraIL	80.45	83.66	84.35	83.08	69.35	85.04	**84.43**	80.19
CoMPILE	79.95	83.56	83.97	83.87	68.36	85.50	84.04	79.89
Meta-iKG (MAML)	80.31	82.95	82.52	**84.23**	72.12	84.11	82.47	79.25
Meta-iKG (Meta-SGD)	**81.10**	**84.26**	**84.57**	83.70	**72.50**	**85.97**	84.05	**81.24**

TABLE III: **Comparison between Models (AUC-PR).**

	FB15k-237				NELL-995			
Model	v1	v2	v3	v4	v1	v2	v3	v4
RuleN	65.35	71.68	67.84	70.53	53.70	69.77	64.29	57.92
GraIL	66.52	73.82	70.15	68.30	55.56	76.40	75.66	71.24
CoMPILE	66.52	72.37	69.77	70.27	62.35	76.51	75.58	68.19
Meta-iKG (MAML)	66.52	72.37	68.81	**74.32**	60.49	74.07	**77.99**	71.63
Meta-iKG (Meta-SGD)	**66.96**	**74.08**	**71.89**	72.28	**64.20**	**77.91**	77.41	**73.12**

TABLE IV: **Comparison between Models (Hits@10).**

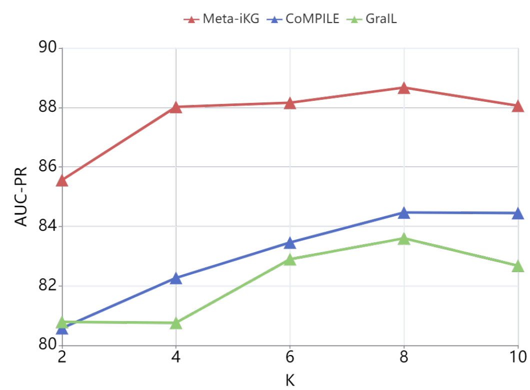

Fig. 2: Analysis on the Influence of K.

but is used to optimize the model. Therefore, the number of parameters of the model remains unchanged.

G. Analysis on the Influence of K

We study the effect of the number of training triplets K per relation on the performance. We select the relations whose number of training triplets is no less than 10 and no larger than K_T, and some training triplets are randomly removed until there are only K triplets per selected relation. K is selected from 2 to 10 for each relation in our experiments, and we also evaluate the predictive result of these relations without removing any triplets for comparison. For comprehensive comparison, we present the corresponding results of subgraph-based methods GraIL [10] and CoMPILE [11]. From the results in Fig. 2 we can infer that when $K \geq 6$, the performance of the model tends to converge (the performance has no significant difference with the non-removed version whose AUC-PR is 88.89), suggesting that Meta-iKG can reach satisfactory performance with a small number of training samples. Compared to GraIL and CoMPILE, our model shows consistent improvement with K set to different values. These results further demonstrate the effectiveness of our Meta-iKG in few-shot inductive relation prediction.

H. Ablation Studies on Meta-learning

We investigate the effectiveness of the introduced large-shot relation update procedure and the relation split operation. As presented in the 'W/O LRUP' case on Table V, when the large-shot relation update procedure is removed, a significant drop on performance is observed, suggesting that the model cannot generalize well on the standard inductive datasets. Note that the model still performs well on few-shot relation (the Hits@10 is 62.5% on FB15k-237-v1 dataset when $K \leq K_T$), which supports our claim that the model cannot perform well on large-shot relation using the regular meta-learning procedure. With the proposed LRUP, our model can generalize well on both few-shot and large-shot relations.

In the case of 'W/O RSO', we do not split the relations into large-shot and few-shot relations, but randomly sample a batch of relations for meta-training and meta-testing during each iteration. Without PRO, the results show that there is a decline of performance on both the datasets (over 2.5% drops on AUC-PR of FB15k-237-v1 dataset), which demonstrates the effectiveness of our strategy to adapt the model trained on large-shot relations to few-shot relations. Notably, the RSO is specifically designed to promote the learning of few-shot relations, and therefore it may have no great impact on the whole dataset. To verify this, we further evaluate the effectiveness of RSO on the few-shot relations. The results show that when the RSO is removed, the Hits@10 for FB15k-237-v1 and FB15k-237-v2 datasets are 47.14 and 82.68 respectively, both significantly lower than 53.57 and 88.46 when RSO is used (in the case of $K \leq K_T$).

I. Analysis on the Hyperparameter h

In this section, we analyze the effect of the hyperparameter hop h. The hop h is set from 2 to 4 in our experiments, where

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3177212, IEEE Transactions on Knowledge and Data Engineering

	FB15k-237-v1		FB15k-237-v2	
	AUC-PR	Hits@10	AUC-PR	Hits@10
W/O LRUP	63.97	43.48	63.37	39.08
W/O RSO	78.34	65.22	84.14	73.42
Meta-iKG	**81.10**	**66.96**	**84.26**	**74.08**

TABLE V: **Ablation Studies on Inductive Datasets.** The 'LRUP' refers to our designed large-shot relation update procedure, and 'RSO' represents relation split operation.

a larger h usually indicates more connections between the target entities in the subgraph. As shown in Table VI, when h is set to 2, a significant drop on performance is observed, mainly for the reason that the connections between target entities become rare such that the relation inference cannot be conducted. When h is set to 4, the model's performance is slightly lower compared to the performance when h is set to 3. We speculate that this is because a large number of noisy nodes and edges are added into the subgraph when h is 4, such that the discrimination of subgraph drops and overfitting occurs. We noticed that there are strong connections between target entities for the majority of triplets even when h is 3. Moreover, the subgraph becomes much larger and the complexity is much higher when h is 4. Therefore, for the sake of performance and complexity, we set h to 3.

	FB15k-237-v1		Nell-995-v1	
h	Hits@10	AUC-PR	Hits@10	AUC-PR
2	65.22	78.35	59.88	67.70
3	**66.96**	**81.10**	**64.20**	**72.50**
4	65.65	80.98	**64.20**	69.52

TABLE VI: **Analysis on the Hyperparameter h.** Since the majority of the subgraphs are empty when h is 1 which have no real meaning, we do not present the results of hop 1.

V. Conclusion

We present Meta-iKG, a novel method for few-shot inductive relational inference. Meta-iKG uses local subgraphs to convey subgraph-specific information and to learn transferable patterns faster via meta-gradients. We evaluate Meta-iKG on two novel several-shot inductive link prediction benchmarks, and the experimental results show that Meta-iKG outperforms state-of-the-art methods.

References

[1] C. Xiong, R. Power, and J. Callan, "Explicit semantic ranking for academic search via knowledge graph embedding," in *WWW*, 2017, pp. 1271–1279.

[2] M. Zhang and Y. Chen, "Link prediction based on graph neural networks," in *NeurIPS*, 2018, pp. 5165–5175.

[3] A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *NeurIPS*, 2013, pp. 2787–2795.

[4] T. Ebisu and R. Ichise, "Generalized translation-based embedding of knowledge graph," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 5, pp. 941–951, 2020.

[5] B. Yang, S. W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *ICLR*, 2015.

[6] S. Zheng, J. Rao, Y. Song, J. Zhang, X. Xiao, E. F. Fang, Y. Yang, and Z. Niu, "Pharmkg: a dedicated knowledge graph benchmark for bomedical data mining," *Briefings in bioinformatics*, vol. 22, no. 4, p. bbaa344, 2021.

[7] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2d knowledge graph embeddings," in *AAAI*, vol. 32, no. 1, 2018.

[8] R. Trivedi, H. Dai, Y. Wang, and L. Song, "Know-evolve: Deep temporal reasoning for dynamic knowledge graphs," in *ICML*, 2017, pp. 3462–3471.

[9] I. Roy, A. De, and S. Chakrabarti, "Adversarial permutation guided node representations for link prediction," *arXiv preprint arXiv:2012.08974*, 2021.

[10] K. K. Teru, E. Denis, and W. L. Hamilton, "Inductive relation prediction by subgraph reasoning," in *ICML*, 2020, pp. 9448–9457.

[11] S. Mai, S. Zheng, Y. Yang, and H. Hu, "Communicative message passing for inductive relation reasoning," *AAAI*, 2021.

[12] N. Keriven, A. Bietti, and S. Vaiter, "Convergence and stability of graph convolutional networks on large random graphs," *arXiv preprint arXiv:2006.01868*, 2020.

[13] X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, and M. Sun, "Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation," *arXiv preprint arXiv:1810.10147*, 2018.

[14] W. Xiong, M. Yu, S. Chang, X. Guo, and W. Y. Wang, "One-shot relational learning for knowledge graphs," *arXiv preprint arXiv:1808.09040*, 2018.

[15] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," *arXiv preprint arXiv:1711.04043*, 2017.

[16] F. Zhou, C. Cao, K. Zhang, G. Trajcevski, T. Zhong, and J. Geng, "Meta-gnn: On few-shot node classification in graph meta-learning," in *CIKM*, 2019, pp. 2357–2360.

[17] M. Chen, W. Zhang, W. Zhang, Q. Chen, and H. Chen, "Meta relational learning for few-shot link prediction in knowledge graphs," *arXiv preprint arXiv:1909.01515*, 2019.

[18] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *ICML*, 2017, pp. 1126–1135.

[19] Z. Li, F. Zhou, F. Chen, and H. Li, "Meta-sgd: Learning to learn quickly for few-shot learning," *arXiv preprint arXiv:1707.09835*, 2017.

[20] L. Liu, T. Zhou, G. LONG, J. Jiang, and C. Zhang, "Many-class few-shot learning on multi-granularity class hierarchy," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2020.

[21] L. Galárraga, C. Teflioudi, K. Hose, and F. M. Suchanek, "Fast rule mining in ontological knowledge bases with amie," *The VLDB Journal*, vol. 24, no. 6, pp. 707–730, 2015.

[22] C. Meilicke, M. Fink, Y. Wang, D. Ruffinelli, R. Gemulla, and H. Stuckenschmidt, "Fine-grained evaluation of rule-and embedding-based systems for knowledge graph completion," in *International Semantic Web Conference*. Springer, 2018, pp. 3–20.

[23] F. Yang, Z. Yang, and W. W. Cohen, "Differentiable learning of logical rules for knowledge base reasoning," in *NeurIPS*, 2017, pp. 2319–2328.

[24] A. Sadeghian, M. Armandpour, P. Ding, and D. Z. Wang, "Drum: End-to-end differentiable rule mining on knowledge graphs," in *NeurIPS*, 2019, pp. 15 347–15 357.

[25] Y. Hu, A. Chapman, G. Wen, and D. W. Hall, "What can knowledge bring to machine learning?—a survey of low-shot learning for structured data," *ACM Transactions on Intelligent Systems and Technology*, vol. 13, no. 3, pp. 1–45, 2022.

[26] B. Haney and A. Lavin, "Fine-grain few-shot vision via domain knowledge as hyperspherical priors," *arXiv preprint arXiv:2005.11450*, 2020.

[27] S. Benaim and L. Wolf, "One-shot unsupervised cross domain translation," *NeurIPS*, vol. 31, 2018.

[28] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2. Lille, 2015.

[29] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *arXiv preprint arXiv:1606.04080*, 2016.

[30] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv preprint arXiv:1803.02999*, 2018.

[31] Y. Lee and S. Choi, "Gradient-based meta-learning with learned layerwise metric and subspace," in *ICML*. PMLR, 2018, pp. 2927–2936.

[32] X. Lv, Y. Gu, X. Han, L. Hou, J. Li, and Z. Liu, "Adapting meta knowledge graph information for multi-hop reasoning over few-shot relations," *arXiv preprint arXiv:1908.11513*, 2019.

[33] M. Mirtaheri, M. Rostami, X. Ren, F. Morstatter, and A. Galstyan, "One-shot learning for temporal knowledge graphs," *arXiv preprint arXiv:2010.12144*, 2020.

[34] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gamon, "Representing text for joint embedding of text and knowledge bases," in *EMNLP*, 2015, pp. 1499–1509.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.