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

Shuangjia Zheng, Sijie Mai, Ya Sun, 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 recent proposed subgraph-based models provided 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 learn transferable patterns faster 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 update procedure to traditional meta-learning to ensure that our model can generalize well both on few-shot and large-shot relations. We evaluate Meta-iKG on inductive benchmarks sampled from NELL and Freebase, and the results show that Meta-iKG outperforms the current state-of-the-art methods both in 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 amounts 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 variety 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 that condense each entity and relation into a low-dimensional continuous vector space, which can then be utilized to infer missing links by operating the produced embeddings [3]–[9].

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 [10], 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, a few efforts have been made to subgraph-based inductive relation prediction, combining the beneficial qualities of both scalability and interpretability [11], [12]. The basic strategy behind this type of model 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 the graph neural network usually needs substantial instances to enable the model stability [13]. However, previous works have shown that a large part of KG relations is actually long-tail [14], [15] and only occur only 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 [16], [17]. The inability to handle the presence of very few samples is one of the major challenges for the current GNNs. In fact, few-shot relations often appear when new entities are encountered. Therefore, few-shot inductive relation reasoning is a practical issue of considerable importance that has not been fully resolved.

Present Work. 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 learn transferable patterns faster 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 function. Moreover, different from standard meta-learning, we introduce 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 both on 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 both in few-shot and standard inductive settings.
In brief, the main contributions are listed below:

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

II. RELATED WORK

**Inductive relation prediction.** A close research line is the rule-based approach [21] and recent proposed differentiable rule learners such as RuleN [22], NeuralLP [23], and DRUM [24] that simultaneously learn the logical rules and confidence scores in an end-to-end paradigm. Recent studies incorporate graph neural network into inductive relation reasoning to capture multi-hop information around the target triple. GraIL [11] proposed a subgraph-based relation reasoning framework to process unseen entities and CoMPILE [12] extends the idea by introducing a node-edge communicative message passing mechanism to model the directed subgraphs, which fits the directional nature of knowledge graph and can naturally deal with asymmetric and anti-symmetric relations. Our study can be interpreted as an extension of CoMPILE method to few-shot knowledge graph completion.

**Few-shot relation prediction.** The few-shot learning models mainly fall into two categories: (i) metric based approaches [25], [26]; (ii) meta-optimizer based approaches [18], [27], [28]. The former one learns an effective metric and corresponding matching function among a set of training instances. For example, GMatching [1] and MetaR [29] use metric methods to generalize over new relations from a handful of associative relations in a knowledge graph. They achieves fairly good performance in one-shot relation prediction yet still have two main limitations: (i) they only encode one-hop neighbors of entities and ignore the high-order neighborhood information around a target triplet. (ii) they are limited to transductive settings and cannot process unseen entities. Note that MetaR [29] aims to quickly optimize the model parameters given the gradients on few-shot data instances. One widely-used strategy is the model-agnostic meta-learning (MAML) [18] which trains model by a small number of gradient updates and leads to fast adapting on a new task. A few attempts have been proposed that combine meta learning with multi-hop reinforcement learning [30] or sequential network [31] to perform few-shot link prediction. 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 not available during reasoning. In this paper, to enable the model to generalize well on the relations that only have few training triplets, we split the relations into few-shot and large-shot relations. If the number of triplets including a relation \(r\) is fewer than a specific threshold \(K_r\), we denote \(r\) as a few-shot relation, otherwise, it is a large-shot (normal) relation. Following the basic idea of meta-learning [1], [18], [29], [30], we train triplets with large-shot relations to find well-initialized parameters and adapt the models 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, the proposed learning framework 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 \(\theta\) for a set of subgraphs surrounding a specific relationship \(r\) (task) to identify the transferable patterns. Meta-learning is based on the relation-specific module for learning a meta model with parameters \(\theta\) 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 query relation \(r^q \in R\), we learn a relation-specific subgraph scoring function using the triplets with the relation \(r^q\) to make predictions. Our subgraph-based scoring function can be divided into three subtasks: (i) extracting the enclosing subgraph surrounding the two target entities, (ii) labeling the entities in the subgraphs with their relative position, and (iii) scoring the labeled subgraphs 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 could infer the target relation [11]. Given a triplet \((s, r, t)\), we define \(s\) as the 1-hop incoming neighbors of \(t\), and \(t\) as the 1-hop outgoing neighbor of \(s\), and \(h\)-hop incoming/outgoing neighbors vice
verse. Firstly we extract the \( h \)-hop incoming neighbors for target head and \( h \)-hop outgoing neighbors for target tail. Then we find the common entities of the extracted neighbors of the two target entities. Finally, we build the subgraph by adding the edges (i.e., the triplets) whose heads and tails are in the common entities or target entities.

**Inductive node labeling function.** Before feeding the \( h \)-hop subgraph to GNN, we first apply an inductive node labeling function to it, which uses different labels to mark nodes’ different roles in the subgraph without leveraging any external domain features and global information. As such, the labeling function to it, which uses different labels to mark hop subgraph to GNN, we first apply an inductive node common entities or target entities.

In detail, for a node \( i \) within the subgraph, its label is defined as \( N_i = \text{one-hot}(d_{si}) \oplus \text{one-hot}(d_{ti}) \in \mathbb{R}^{2(h+1)} \) where \( d_{si} \) denotes the shortest distance from the target head entity to node \( i \), and \( d_{ti} \) denotes the shortest distance from the node \( i \) to target tail entity. Note that similar to GraIL [11], the two target nodes, \( s_T \) and \( t_T \) are uniquely 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 [12]. Formally, given a target triplet \((s_T, r_T, t_T)\) with the enclosing subgraph \( G \), the subgraph scoring function can be defined as:

\[
S = \text{GNN}(G, N_{s_T}, R_{r_T}, N_{t_T}; \theta)
\]

where \( S \) denotes the predicted plausibility of the target triplet, \( N_{s_T}, R_{r_T}, \), and \( N_{t_T} \) denotes the embedding of target head, target relation, and target tail, respectively. Please refer to CoMPiLe [12] or our Appendix for the detailed procedure. The relation embedding \( R \) is parameterized as a learnable matrix and is shared across subgraphs.

**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 generalizes well both on few-shot and large-shot relations. Formally, we consider a meta GNN, i.e. CoMPiLe with parameters \( \theta \). 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 \( \gamma \) is a scalar. In each iteration, we sample a batch of relations \( r^f \) from large-shot relation set \( R^f = \{r|n_r > K_T, r \in R\} \) and a batch of relations \( r^l \) from few-shot relation set \( R^l = \{r|n_r \leq K_T, r \in R\} \). Then we sample triplets belonging to \( r^f \) and \( r^l \) 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 \( \theta_{r,f} \). Following MAML [18], the updated parameters \( \theta_{r,f} \) are computed using one or more gradient descent updates on the few-shot tasks \( T^l \). Formally, we will go over many large-shot tasks (i.e., meta-training tasks) in a batch:

\[
\theta_{r,f} = \theta - \alpha \cdot \sum_{T^l} \nabla_{\theta} D_S^{r,f}(\theta),
\]

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

\[
\theta_{r,f} = \theta - \beta \cdot \sum_{T^l} \nabla_{\theta} L_{r,f}^{D_Q}(\theta_{r,f}),
\]

where \( \beta \) is the meta-learning rate. By this means, Meta-iKG can learn to fast adapt to the few-shot relations with the aim of the well-initialized parameters \( \theta_{r,f} \) updated by the support set in 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 \( \theta_{r,f} \) depend on the few-shot relations and therefore the model may not perform well on large-shot relations. To this end, we innovative to introduce another updating operation on \( \theta_{r,f} \) using the support set \( D_S \) belonging to large-shot tasks \( T^l \) with a smaller learning rate \( \beta' \), which is called ‘large-shot relation update procedure’. The equation is shown as below:

\[
\theta \leftarrow \theta_{r,f} - \beta' \cdot \sum_{T^l} \nabla_{\theta} L_{r,f}^{D_S}(\theta_{r,f}),
\]

where \( \theta \) is the final updated parameter. This simple operation enables Meta-iKG to generalize well both on large-shot and few-shot relations. Note that we apply Meta-SGD [19] as our meta-learner, in which case the learning rate \( \alpha \) is meta-learnable. We detail the Meta-iKG in Algorithm 1.

**IV. EXPERIMENTS**

**A. Dataset**

FB15K-237 [32] and NELL-995 [1] are two commonly used databases for link prediction. We follow CoMPiLe [12]

**Algorithm 1: Meta-iKG**

| Input: GNN parameter \( \theta \), learning rate \( \beta \) and \( \beta' \), meta-learning rate \( \alpha \), task distribution \( p(T^f) \) and \( p(T^l) \) | Output: GNN parameter \( \theta \) |
|---|---|
| step 1: Initialize \( \theta \) and \( \alpha \) |    |
| step 2: while not done do | Sampling large-shots \( T^l \) from \( p(T^l) \) | Obtaining \( \theta_{r,f} \) using \( T^l, \theta \) and \( \alpha \) via Eq. 2 |
| | Sampling few-shot tasks \( T^f \) from \( p(T^f) \) | Obtaining \( \theta_{r,f} \) using \( T^f, \theta_{r,f} \) and \( \beta \) via Eq. 3 |
| | Updating \( \theta \) using \( T^f, \theta_{r,f} \) and \( \beta \) via Eq. 4 |
| | Updating \( \alpha \) by: \( \alpha \leftarrow \alpha - \beta \sum_{T^f} \nabla_{\alpha} L_{r,f}^{D_Q}(\alpha) \) | end |
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. The statistics of the filtered inductive FB15k-237 and Nell-995 datasets are shown in Table I and Table II respectively. 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 $v_4$), respectively. The inductive datasets are the filtered versions of the corresponding inductive datasets extracted in GraIL [11]. Since we use more challenge inductive datasets, so the performance of the models is generally weaker than the performance reported on GraIL [11].

The subgraphs of train triplets and validation triplets are extracted from the train graph, and the subgraphs of the test triplets are extracted from the test graph. The test graph contains entities that do not presented in the train graph, and the train graph contains all the relations in the test graph. The size of the train graph refers to the number of triplets contained in the train graph, and the size of test graph vice verse. Note that the train triplets are all in the train graph, but the test triplets are not in the test graph. Also, the validation triplets are not in the train graph. These settings are the same as those in GraIL and CoMPILE.

To perform the Hits@10 experiment, we need to sample 49 negative head triplets and 49 negative tail triplets for each test triplet (this setting is consistent with GraIL and CoMPILE; the GraIL paper said they sampled 50 negative triplets, but actually it is 49 in the code), where the negative head/tail triplets are generated by replacing the head/tail of the test triplet with other entities. Therefore, we do not use the test triplets that cannot find enough negative head or tail triplets to evaluate the model. We evaluate our model both on original inductive setting and on few-shot inductive setting where the relations whose number of training triplets (denoted as $K$) is fewer than 5, 10, or $K_T$ are selected for testing.

### B. Experimental Details

#### Evaluation Protocol
To be consistent with the 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. 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 triplets are obtained by replacing the head or tail of the test triplets with other entities. We train the model for four times and average the testing results to obtain the final performance.

#### Hyper-parameter Setting
We implement our model on Pytorch. We run the model for 20 epochs in the original inductive datasets, and each epoch contains 100 meta-updates. We use Adam [33] as optimizer with learning rate being 0.001. The hop number $h$ is set to 3 which is consistent with GraIL and CoMPILE. The number of iterations $l$ is set to 3 or 4 which depends on the datasets. The few-shot factor $\gamma$ is set to 0.1. 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 the target head and target tail. The codes and datasets will be publicly available once accepted.

### C. Baselines

We compare Meta-iKG with the state-of-the-art subgraph-based inductive models GraIL [11] and CoMPILE [12] 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. Comparisons on Inductive Datasets.

We compare our proposed Meta-iKG with other baselines on the inductive datasets. From the Table III and Table IV, we can conclude that: (a) our Meta-iKG, especially the Meta-SGD version, achieves the best performance on the majority of the

---

**Table I: Inductive FB15k-237 Datasets**

| Version | Train Relations | Train Graph | Train Triplets | Validation Triplets | Test Relations | Test Graph | Test Triplets |
|---------|----------------|-------------|----------------|---------------------|---------------|------------|---------------|
| v1      | 183            | 4,245       | 4,040          | 475                 | 146           | 1,993      | 108           |
| v2      | 213            | 9,739       | 9,462          | 1,142               | 176           | 4,145      | 380           |
| v3      | 218            | 17,986      | 17,703         | 2,179               | 187           | 7,406      | 779           |
| v4      | 222            | 27,203      | 26,917         | 1,658               | 204           | 11,714     | 1,369         |

**Table II: Inductive Nell-995 Datasets**

| Version | Train Relations | Train Graph | Train Triplets | Validation Triplets | Test Relations | Test Graph | Test Triplets |
|---------|----------------|-------------|----------------|---------------------|---------------|------------|---------------|
| v1      | 14             | 4,687       | 3,610          | 379                 | 14            | 833        | 81            |
| v2      | 88             | 8,219       | 7,118          | 921                 | 79            | 4,586      | 430           |
| v3      | 142            | 16,393      | 14,453         | 1,848               | 122           | 8,048      | 686           |
| v4      | 77             | 7,546       | 6,710          | 419                 | 61            | 7,073      | 638           |

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

| Model | FB15k-237 | NELL-995 |
|-------|-----------|----------|
| RuleN | v1        | v2       | v3       | v4        | v1       | v2       | v3       | v4       |
| GraIL | 79.60     | 80.44    | 80.44    | 80.44    | 76.72    | 75.39    | 75.39    | 75.39    |
| CoMPILE | 79.95    | 80.36    | 80.36    | 80.36    | 76.36    | 75.00    | 75.00    | 75.00    |
| Meta-iKG (Meta-SGD) | 81.10 | 84.26 | 84.57 | 83.70 | 72.50 | 85.97 | 84.05 | 81.24 |

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

| Model | FB15k-237 | NELL-995 |
|-------|-----------|----------|
| RuleN | v1        | v2       | v3       | v4       | v1       | v2       | v3       | v4       |
| GraIL | 65.35     | 70.08    | 67.84    | 30.53    | 53.19    | 59.17    | 63.37    | 57.59    |
| CoMPILE | 65.25    | 72.82    | 70.15    | 68.30    | 55.56    | 76.40    | 75.66    | 71.24    |
| Meta-iKG (Meta-SGD) | 66.96 | 74.08 | 71.49 | 72.28 | 64.20 | 77.91 | 77.41 | 73.12 |
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. These results demonstrate the effectiveness of our unique meta-learning strategy in inductive relation prediction, which can generalize well both on large-shot and few-shot relations via the updating procedure in Eq. 2, Eq. 3 and Eq. 4. Our model enables a better prediction capacity on few-shot relations without sacrificing the performance on the overall datasets.

E. Comparison on Few-Shot Testing Sets.

In this section, we evaluate our Meta-iKG on the 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 during training to split the large-shot and few-shot relations). For comparison, we also present the results of CoMPILE, which is the baseline message passing model for our Meta-iKG. According to Table V and VI, we find that (a) both versions of Meta-iKG significantly outperform the baseline CoMPILE on the majority of settings, and the improvement is much larger compared to that in the original inductive datasets; (b) although the MAML version of Meta-iKG performs weaker than CoMPILE on some original datasets, it manages to outperform CoMPILE significantly on the few-shot scenarios. These results demonstrate the effectiveness and stability of our meta-learning strategies in few-shot inductive relation prediction by learning transferable patterns of subgraphs faster via meta gradients.

F. Analysis on the Model Complexity

Since the meta-learning algorithm does not introduce additional module to process the enclosing 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 size of the subgraph. As for the space complexity, for the MAML version of our Meta-iKG, since it introduces no additional parameters, the number of parameters is the same as that of the original CoMPILE (which is 41,185 when the dimensionality of relation embedding $d$ is set to 32 and the number of layers $l$ is set to 3). For the Meta-SGD version of our Meta-iKG, since it introduces learnable learning rate parameter $\alpha$ which has the same size as the parameter of the GNN (i.e., CoMPILE), it has twice as many parameters as that of CoMPILE. Notably, the learning rate parameter $\alpha$ is not added to the model itself, but is used to optimize the model.

G. Analysis on the Influence of $K$

We study the effect of the number of training triplets $K$ per relation on the performance of FB15k-237-v2 dataset. We select the relations whose number of training triplets is no less than and no larger than $K_T$ (10 $\leq K \leq K_T$), and some training triplets are randomly removed until there are only $K$ triples 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. 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. These results further demonstrate the effectiveness of our meta-learning method in few-shot inductive relation prediction.

| Model            | FB15k-237-v1 $K \leq 5$ | FB15k-237-v2 $K \leq 5$ | Nell-995-v1 $K \leq 5$ | Nell-995-v2 $K \leq 5$ |
|------------------|-------------------------|-------------------------|------------------------|------------------------|
| CoMPILE          | 43.75                   | 42.86                   | 80.08                  | 76.92                  |
| Meta-iKG (MAML)  | 75.00                   | 56.25                   | 57.14                  | 66.67                  |
| Meta-iKG (Meta-SGD) | 75.00               | 60.71                   | 53.57                  | 66.67                  |

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.

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 VII when the large-shot relation update procedure is removed, significant drop on performance is observed, which suggests 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 support our claim that the model cannot perform well on large-shot relation using the regular meta-learning procedure. In the case of ‘W/O PRO’, we do not split the relations into large-shot and few-shot relations,
and 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 that adapting the model trained on large-shot relations to few-shot relations.

| Model                  | $K \leq 5$ | $K \leq 10$ | $K \leq K_T$ | $K \leq K_T$ | $K \leq 5$ | $K \leq 10$ | $K \leq K_T$ |
|------------------------|------------|-------------|--------------|--------------|------------|-------------|--------------|
| COMPARE                | 73.57      | 73.14       | 70.01        | 68.57        | 70.01      | 71.43       | 71.07        |
| Meta-iKG (MAML)        | 66.67      | 62.49       | 80.00        | 76.42        | 65.87      | 71.43       | 71.07        |
| Meta-iKG (Meta-SGD)    | 58.33      | 57.14       | 73.33        | 73.57        | 65.87      | 72.86       | 73.14        |

Table VII: Ablation Studies on Inductive Datasets. The ‘LRUP’ refers to our designed large-shot relation update procedure, and ‘RPO’ represents relation split operation.

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 Proceedings of the 26th international conference on world wide web, 2017, pp. 1277–1279.
[2] M. Zhang and Y. Chen, “Link prediction based on graph neural networks,” in Advances in Neural Information Processing Systems, 2018, pp. 5165–5175.
[3] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi-relational data,” in Advances in neural information processing systems, 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] M. Nickel, Ł. Rosasco, and T. Poggio, “Holographic embeddings of knowledge graphs,” arXiv preprint arXiv:1501.04935, 2015.
[6] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, “Embedding entities and relations for learning and inference in knowledge bases,” arXiv preprint arXiv:1412.6575, 2014.
[7] T. Trouillon, C. R. Dance, É. Gaussier, J. Welbl, S. Riedel, and G. Bouchard, “Knowledge graph completion via complex tensor factorization,” The Journal of Machine Learning Research, vol. 18, no. 1, pp. 4735–4772, 2017.
[8] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, “Convolutioal 2d knowledge graph embeddings,” arXiv preprint arXiv:1707.01476, 2017.
[9] H. Xiao, Y. Chen, and X. Shi, “Knowledge graph embedding based on multi-view clustering framework,” IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 2, pp. 585–596, 2021.
[10] R. Trivedi, H. Dui, Y. Wang, and L. Song, “Know-evolve: Deep temporal reasoning for dynamic knowledge graphs,” arXiv preprint arXiv:1705.05742, 2017.
[11] K. K. Teru, E. Denis, and W. L. Hamilton, “Inductive relation prediction by subgraph reasoning,” in ICML, 2020.
[12] S. Mai, S. Zheng, Y. Yang, and H. Hu, “Communicative message passing for inductive relation reasoning,” arXiv preprint arXiv:2012.08911, 2020.
[13] N. Keriven, A. Bietti, and S. Vaiter, “Convergence and stability of graph convolutional networks on large random graphs,” arXiv preprint arXiv:2006.10186, 2020.
[14] 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.
[15] 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.
[16] V. Garcia and J. Bruna, “Few-shot learning with graph neural networks,” arXiv preprint arXiv:1711.04043, 2017.
[17] 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 Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 2537–2540.
[18] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” in International Conference on Machine Learning. PMLR, 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] G. Koch, R. Zemel, and R. Salakhutdinov, “Siamese neural networks for one-shot image recognition,” in ICML deep learning workshop, vol. 2. Lille, 2015.
[22] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, “Matching networks for one shot learning,” arXiv preprint arXiv:1606.04080, 2016.
[23] A. Nichol, J. Achiam, and J. Schulman, “On first-order meta-learning algorithms,” arXiv preprint arXiv:1803.02999, 2018.
[24] Y. Lee and S. Choi, “Gradient-based meta-learning with learned layerwise metric and subspace,” in International Conference on Machine Learning. PMLR, 2018, pp. 2927–2936.
[25] 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.
[26] 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.
[27] M. Mirfakhr, M. Rostami, X. Ren, F. Morstatter, and A. Galstyan, “One-shot learning for temporal knowledge graphs,” arXiv preprint arXiv:2010.12144, 2020.
[28] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gafton, “Representing text for joint embedding of text and knowledge bases,” in Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 1499–1509.
[29] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in ICLR, 2015.