



Expanding the Scope: Inductive Knowledge Graph Reasoning with Multi-starting Progressive Propagation

Zhoutian Shao¹ , Yuanning Cui¹ , and Wei Hu^{1,2} (✉)

¹ State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

ztshao.nju@gmail.com, yncui.nju@gmail.com, whu@nju.edu.cn

² National Institute of Healthcare Data Science, Nanjing University, Nanjing 210093, China

Abstract. Knowledge graphs (KGs) are widely acknowledged as incomplete, and new entities are constantly emerging in the real world. Inductive KG reasoning aims to predict missing facts for these new entities. Among existing models, graph neural networks (GNNs) based ones have shown promising performance for this task. However, they are still challenged by inefficient message propagation due to the distance and scalability issues. In this paper, we propose a new inductive KG reasoning model, **MStar**, by leveraging conditional message passing neural networks (C-MPNNs). Our key insight is to select multiple query-specific starting entities to expand the scope of progressive propagation. To propagate query-related messages to a farther area within limited steps, we subsequently design a highway layer to propagate information toward these selected starting entities. Moreover, we introduce a training strategy called LinkVerify to mitigate the impact of noisy training samples. Experimental results validate that MStar achieves superior performance compared with state-of-the-art models, especially for distant entities.

Keywords: Knowledge graphs · Inductive reasoning · Conditional message passing

1 Introduction

Knowledge graphs (KGs) have become a valuable asset for many downstream AI applications, including semantic search, question answering, and logic reasoning [4, 11, 33]. Real-world KGs, such as Freebase [1], NELL [21], and DBpedia [15], often suffer from the incompleteness issue that lacks massive certain triplets [5, 12]. The KG reasoning task aims to alleviate incompleteness by discovering missing triplets based on the knowledge learned from known facts. Early studies [38] assume that KGs are static, ignoring the potential unseen entities

and emerging triplets in the continuously updated real-world KGs. This motivates the task of inductive KG reasoning [32, 46], which allows for incorporating emerging entities and facts during inference.

Table 1. Hits@10 results of RED-GNN [47] in our empirical study. We divide all triplets in the FB15k-237 (v1) dataset [32] into four groups according to the shortest distance between head and tail entities. “ ∞ ” denotes that the head entity cannot reach the corresponding tail entity in the KG. The maximum shortest distance is 10 in the FB15k-237 (v1) when ignoring triplets belonging to ∞ .

Distance	Proportions	Layers = 3	Layers = 6	Layers = 9
[1, 4)	70.25%	.611	.594	.587
[4, 7)	22.44%	.000	.102	.154
[7, 10]	3.90%	.000	.000	.088
∞	3.41%	.000	.000	.000

Due to their excellent efficiency and performance, conditional message passing neural networks (C-MPNNs), such as NBFNet [50] and RED-GNN [47], have emerged as one of the premier models in the field of inductive KG reasoning. To transmit conditions, existing C-MPNNs only incorporate conditional information into the head entity and propagate along the relational paths progressively. However, this single-starting-entity strategy results in a limited conditional message passing scope, leading to the failure of message passing from the head entity to distant target entities. This inspires us to extend the scope of conditional message passing to support reasoning on target entities in a farther area.

We conduct an empirical study to analyze the drawbacks of the limited message passing scope. Specifically, we report the results of a C-MPNN, RED-GNN [47], on predicting target entities at different distances in Table 1. It can be observed that RED-GNN performs poorly for queries with distant target entities, even stacking more message-passing layers. This indicates that existing C-MPNNs cannot effectively propagate conditional messages toward distant target entities, hindering performance on these queries. Although stacking more GNN layers can alleviate this issue to some extent, it causes high computation and performance declines on the queries with target entities nearby.

In this paper, we propose a novel inductive KG reasoning model MStar based on **Multi-Starting** progressive propagation, which expands the scope of efficient conditional message passing. Our key insight is to utilize more conditional starting entities and create shortcuts between the head entity and them. Specifically, we design a starting entities selection (SES) module and a highway layer to select multiple starting entities and create shortcuts for conditional message passing, respectively. First, the SES module encodes entities using a pre-embedded GNN and then selects multiple query-dependent starting entities, which may include entities distant from the head entity. These entities broaden the scope of subsequent progressive propagation and allow MStar to propagate along query-related

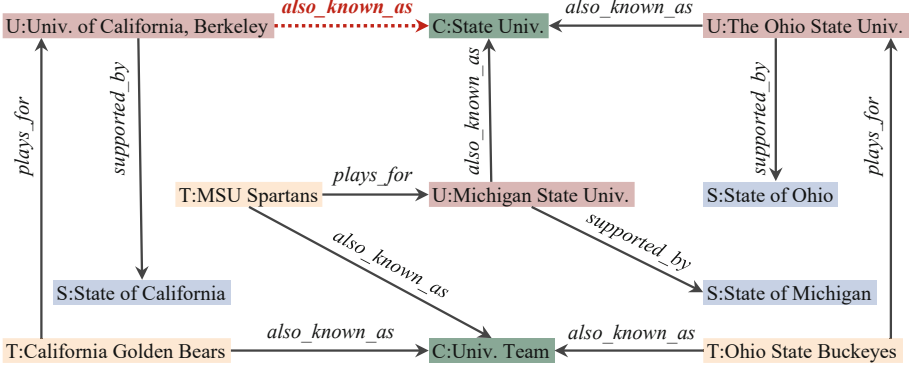


Fig. 1. A motivating example of distant target tail entities for predicting (*Univ. of California, Berkeley* → *also_known_as* → *State Univ.*). Prefix “U”, “S”, and “T” represent university, state, and basketball teams, respectively. Prefix “C” represents category-type entities. Different colors and prefixes symbolize distinct entity types.

relational paths to enhance reasoning concerning distant entities. Second, we create shortcuts from the head entity to the selected multiple starting entities in the highway layer. The design of the highway layer is inspired by skip connection from ResNet [8]. The conditional message can be passed to distant entities through the highway layer. For example, in Fig. 1, a 3-layer RED-GNN would fail to predict the target answer, because the length of the shortest path between head entity *Univ. of California, Berkeley* and the target entity *State Univ.* is larger than 3. In contrast, our MStar can select multiple starting entities, e.g., *Michigan State Univ.* and *The Ohio State Univ.*, and transmit conditional messages to them through the highway layer. Thus, MStar can achieve a better reasoning performance than other C-MPNNs on this query. After the highway layer, we follow it with a multi-condition GNN to perform message passing based on the embeddings of multiple starting entities. We also propose a training sample filtering strategy called LinkVerify to reduce the impact of the unvisited target entities. Overall, MStar visits more query-related distant entities in limited steps and provides more conditional information to these entities compared with existing models.

Our main contributions in this paper are summarized as follows:

- We propose a novel inductive KG reasoning framework based on C-MPNN, named **MStar**. It extends the scope of conditional message passing to improve the predictions of distant target entities.
- We design two modules, SES and highway layer. The SES module performs starting entities selection for visiting distant entities. The highway layer provides shortcuts for efficient conditional message passing, alleviating computation waste during additional propagation.
- We conduct extensive experiments on inductive datasets to demonstrate the effectiveness of our framework and each module. The results show that MStar

outperforms the existing state-of-the-art reasoning models and improves the performance on queries with distant target entities.

The rest of this paper is organized as follows. We first discuss related works in Sect. 2. Then, we describe the reasoning task and propagation mechanisms in Sect. 3. The details of MStar are presented in Sect. 4, and the experimental results are reported in Sect. 5. Finally, in Sect. 6, we discuss the superiority of MStar and possible extensions in future work.

2 Related Work

2.1 Knowledge Graph Reasoning

KG reasoning has been an active research area due to the incompleteness of KGs. Typical KG reasoning models process each triplet independently and extract the latent semantics of entities and relations. To model the semantics of the triplets, TransE [2], TransH [39], TransR [17], and RotatE [29] compute translational distance variously. RESCAL [22], DistMult [44], and ComplEx [35] follow another reasoning paradigm based on semantic matching. Instead of exploring the information implied in a single triplet, R-GCN [28] and CompGCN [36] capture global structure evidence based on graph neural networks (GNNs). These models, however, learn unary fixed embedding from training, which cannot be generalized to emerging entities in the inductive KGs. Instead, our model embodies relational information to encode emerging entities.

2.2 Inductive Knowledge Graph Reasoning

One research line of inductive KG reasoning is rule mining, independent of entity identities. RuleN [20] and AnyBURL [19] try to prune the process of rule searching. Neural LP [45] and DRUM [27] propose to learn logical rules in an end-to-end differentiable manner, learning weights for each relation type and path. However, the rules are usually short due to the expensive computation for mining and may not be generalized to distant entities.

Another research line is subgraph extraction. GraIL [32] extracts subgraphs around each candidate triplet and labels the entities with the distance to the head and tail entities. CoMPiLE [18], TACT [3], SNRI [43], LogCo [23], and ConGLR [16] follow a similar subgraph-labeling paradigm. However, the subgraphs that these models extract convey insufficient information due to sparsity. These models constitute our baselines for inductive KG reasoning.

2.3 Conditional Message Passing Neural Networks

Recently, a variant of GNNs called conditional message passing neural networks (C-MPNNs) [10] propagates messages along the relational paths and encodes pairwise entity embeddings. Given a query head u and a query relation q as

conditions, C-MPNNs compute embeddings of $(v | u, q)$ for all entity v . To incorporate conditions into embeddings, NBFNet [50] and A*Net [49] initialize the head entity with the embedding of query relation and propagate in the full KG for each GNN layer. However, conditional information passing is still restricted in the neighborhood of the head entity. Differently, RED-GNN [47], AdaProp [48], and RUN-GNN [41] propagate the message progressively starting from the head entity without special initialization. During progressive propagation, the involved entity set is augmented step by step with the neighbor entities of the current set instead of being a full entity set. Thus, progressive propagation cannot even visit distant entities in limited steps. MStar alleviates the above problem by selecting multiple starting entities adaptively for progressive propagation and transmitting conditional information through shortcuts.

EL-GNN [25] is another work related to C-MPNNs. This study proposes that C-MPNNs learn the rules of treating the head entity as constant when the head entity is initialized with conditional information. Thus, EL-GNN learns more rules by assigning unique embeddings for entities whose out-degree in the KG reaches a specific threshold. However, the degree and entity-specific embeddings are fixed, which violates the nature of inductive KG reasoning. Our MStar selects starting entities according to the query and generates conditional entity embeddings, which can be applied to unseen entities.

2.4 Skip Connection

Skip connection [8] is a popular technique in deep learning that skips one or more layers. Skipping layers contributes to addressing vanishing or exploding gradients [31] by providing a highway for the gradients. ResNet [8] constructs the highway by adding input \mathbf{x} and output $F(\mathbf{x})$. DenseNet [9] provides multiple highways by concatenating the input of each layer. These models transmit the input in shallow layers directly to the target deeper layer in an efficient way. Inspired by skip connection, MStar constructs a highway with several new edges to transmit messages faster and propagate to farther entities.

3 Preliminaries

Knowledge Graph. A KG $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$ is composed of finite sets of entities \mathcal{E} , relations \mathcal{R} , and triplets \mathcal{F} . Each triplet $f \in \mathcal{F}$ describes a fact from head entity to tail entity with a specific relation, i.e., $f = (u, q, v) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, where u , q , and v denote the head entity, relation, and tail entity, respectively.

(Inductive) Knowledge Graph Reasoning. To complete the missing triplet in real-world KGs, KG reasoning is proposed to predict the target tail entity or head entity with a given query $(u, q, ?)$ or $(?, q, v)$. Given a source KG $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$, inductive KG reasoning aims to predict the triplets involved in the target KG $\mathcal{G}' = (\mathcal{E}', \mathcal{R}', \mathcal{F}')$, where $\mathcal{R}' \subseteq \mathcal{R}$, $\mathcal{E}' \not\subseteq \mathcal{E}$, and $\mathcal{F}' \not\subseteq \mathcal{F}$.

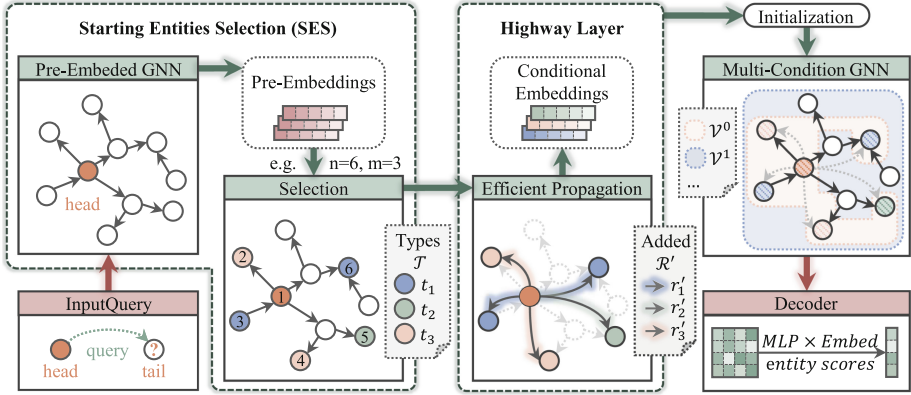


Fig. 2. Framework overview of MStar

Starting Entities in Progressive Propagation. GNNs transmit messages based on the message propagation framework [7, 40]. This framework prepares an entity set to transmit messages for each propagation step. Full propagation transmits messages among all entities at all times. Progressive propagation continuously incorporates the neighbor entities of the entity set in the previous step. Based on progressive propagation, we use starting entities to indicate the entities involved in the first layer of the GNN. Given the starting entities \mathcal{S} , the entities involved in the ℓ^{th} layer of the GNN can be formulated as

$$\mathcal{V}^\ell = \begin{cases} \mathcal{S} & \ell = 0 \\ \mathcal{V}^{\ell-1} \cup \{x \mid \exists (e, r, x) \in \mathcal{N}(e) \wedge e \in \mathcal{V}^{\ell-1}\} & \ell > 0 \end{cases},$$

where $\mathcal{N}(e)$ denotes the neighbor edges of the entity e . In particular, NBFNet puts all the entities into \mathcal{S} , i.e., $\mathcal{S} = \mathcal{E}$. RED-GNN only puts the head entity into \mathcal{S} , i.e., $\mathcal{S} = \{u\}$ with given query $(u, q, ?)$. Too few starting entities limit the scope of conditional message passing. On the contrary, too many start entities disperse the attention of GNNs on local information which is critical for reasoning. Our model MStar strikes a balance by including the head entity and some selected query-dependent starting entities that are helpful for reasoning.

4 Methodology

4.1 Model Architecture Overview

The overview of MStar is presented in Fig. 2. Specifically, we first employ the pre-embedded GNN to pre-encode all entities. Then, SES selects n query-dependent starting entities according to the pre-embeddings. The highway layer classifies starting entities into m types, considering the correlation between the head entity and other starting entities. To improve message-passing efficiency, the highway layer maps each entity type into a new relation and constructs shortcut edges

between the head entity and other starting entities. Based on the message passing on the shortcut edges, we use the highway layer to obtain conditional entity embeddings as the initialization for multi-condition GNN. Finally, the multi-condition GNN propagates relational information progressively conditioned on these starting entities and generates pairwise embeddings of each entity. According to the final entity embeddings, the decoder operates as a multilayer perceptron (MLP) and generates scores for each candidate entity.

4.2 Starting Entities Selection

As shown in Fig. 1, progressive propagation starts from the only entity (head entity) and cannot reach the distant entities. However, the excessive utilization of starting entities introduces noisy relational paths into the reasoning. Despite the expansion of the propagation, some starting entities still miss the target entities and visit other distant entities unrelated to the query. Thus, we propose to select multiple query-dependent starting entities adaptively to cover a farther area but not introduce irrelevant noise in reasoning.

Pre-Embedded GNN. To find the starting entities related to the query, we first introduce a pre-embedded GNN to learn the simple semantics of the entities. The pre-embedded GNN transmits messages among all entities in the KG following the full propagation paradigm. To explore query-related knowledge, the pre-embedded GNN encodes the relation conditioned on query relation q . Specifically, the computation for message passing is given by

$$\mathbf{h}_{\text{pre}|u,q}^\ell(e) = \frac{1}{|\mathcal{N}(e)|} \sum_{(e,r,x) \in \mathcal{N}(e)} \left(\mathbf{h}_{\text{pre}|u,q}^{\ell-1}(x) + \hat{\mathbf{r}}_q \right), \quad (1)$$

$$\hat{\mathbf{r}}_q = \mathbf{W}_r \mathbf{q} + \mathbf{b}_r, \quad (2)$$

where $\mathbf{h}_{\text{pre}|u,q}^\ell(e)$ denotes the embedding of the entity e in propagation step ℓ , \mathbf{q} is a learnable embeddings for relation q , $\mathbf{W}_r \in \mathbb{R}^{d \times d}$ is an r -specific learnable weight matrix, and $\mathbf{b}_r \in \mathbb{R}^d$ is an r -specific learnable bias. d is the dimension of both entity and relation embeddings. $\hat{\mathbf{r}}_q$ denotes the embedding of relation r conditioned on q . The pre-embedded GNN initializes $\mathbf{h}_{\text{pre}|u,q}^0$ as zero vectors and produces the entity embeddings $\mathbf{h}_{\text{pre}|u,q}^{L_1}$ after L_1 layers of message passing.

Selection. Provided with the embeddings of entities conditioned on u and q , we design a score function to select query-dependent starting entities. The score function measures the importance of entities relative to the head entity and query relation. Given an entity e , the importance score $\alpha_{e|u,q}$ is defined as

$$\alpha_{e|u,q} = \mathbf{W}_1 \left(\text{ReLU} \left(\mathbf{W}_2 \left(\mathbf{h}_{\text{pre}|u,q}^{L_1}(e) \oplus \mathbf{h}_{\text{pre}|u,q}^{L_1}(u) \oplus \mathbf{q} \right) \right) \right), \quad (3)$$

where $\mathbf{W}_1 \in \mathbb{R}^{1 \times d}$ and $\mathbf{W}_2 \in \mathbb{R}^{d \times 3d}$ are learnable weight matrices. \oplus denotes the concatenation of two vectors. We keep the top- n entities as starting entity set $\mathcal{S}_{u,q}$. $\mathcal{S}_{u,q}$ can propagate along the relational paths conditioned on the query.

4.3 Highway Layer

Given multiple starting entities, progressive propagation can traverse more entities, particularly those located at distant positions. The distant entities, however, receive nothing about the conditional information, due to the limited scope of conditional message passing. Inspired by the skip connection [8], which allows skip-layer feature propagation, we introduce a highway layer to tackle this issue.

Aiming to propagate conditional information to the starting entities, we consider constructing shortcut edges between the query head entity and the other starting ones. Due to the different semantics of the starting entities, we classify entities into m types based on the embeddings. Each type indicates that this group of entities has a specific semantic relationship with the head entity. Then, we map each entity type to a new semantic relation type and construct new edges. Given conditions u, q and entity e , the entity type is defined as follows:

$$\beta_{e|u,q} = \arg \max_t \mathbf{W}_t \mathbf{h}_{\text{pre}|u,q}^{L_1}(e), \quad t \in [1, m], \quad (4)$$

where t is a type of starting entities, and $\mathbf{W}_t \in \mathbb{R}^{1 \times d}$ is a t -specific learnable weight matrix.

Given starting entity types, the highway layer constructs shortcut edges as

$$\mathcal{H}_{u,q} = \left\{ (u, r'_{\beta_{e|u,q}}, e) \mid e \in \mathcal{S}_{u,q} - \{u\} \right\}, \quad (5)$$

where $r'_{\beta_{e|u,q}}$ denotes the new relation that we introduce, corresponding to the starting entity type. These edges act as a skip connection to support skipping propagation from the head to the starting entities.

Finally, the highway layer performs message passing on $\mathcal{H}_{u,q}$ to obtain the embeddings of the selected starting entities:

$$\mathbf{g}_{u,q}(e) = \sum_{(e,r,x) \in \mathcal{N}_{\text{highway}}(e)} \mathbf{g}_{u,q}(x) \odot \hat{\mathbf{r}}_q, \quad (6)$$

where $\mathbf{g}_{u,q}(e)$ denotes the embedding of entity e , $\mathcal{N}_{\text{highway}}(e)$ denotes the neighbor edges of the entity e in set $\mathcal{H}_{u,q}$, and \odot denotes the point-wise product between two vectors. To satisfy target entity distinguishability [10], we set a learnable embedding for the head entity u .

4.4 Multi-condition GNN

In MStar, we introduce a multi-condition GNN to produce the final entity embeddings. The multi-condition GNN is a C-MPNN conditioned on the head entity and query relation. Specifically, the multi-condition GNN initializes entity embeddings $\mathbf{h}_{u,q}^0$ as $\mathbf{g}_{u,q}$ and propagates from the starting entities progressively. Given the query triplet $(u, q, ?)$, we incorporate the query information into propagation in two ways.

First, we model the embedding of relation r in an edge as $\hat{\mathbf{r}}_q$ conditioned on the query relation q same as Eq. (2). Second, considering that the semantics of

edges are query-dependent, we use the attention mechanism [37] and assign a weight for every edge (e, r, x) in step ℓ :

$$\gamma_{(e,r,x)|u,q}^\ell = \sigma\left(\mathbf{W}_{\text{attn}}^\ell \text{ReLU}(\mathbf{W}_{\text{attn}_u}^\ell \mathbf{h}_{u,q}^{\ell-1}(e) + \mathbf{W}_{\text{attn}_r}^\ell \hat{\mathbf{r}}_q + \mathbf{W}_{\text{attn}_q}^\ell \mathbf{q})\right), \quad (7)$$

where $\mathbf{W}_{\text{attn}}^\ell \in \mathbb{R}^{1 \times d_\gamma}$, $\mathbf{W}_{\text{attn}_u}^\ell$, $\mathbf{W}_{\text{attn}_r}^\ell$ and $\mathbf{W}_{\text{attn}_q}^\ell \in \mathbb{R}^{d_\gamma \times d}$ are learnable weight matrices, d_γ is the dimension of attention, $\mathbf{h}_{u,q}^\ell(e)$ denotes the embedding of the entity e in multi-condition GNN at step ℓ , and σ denotes a sigmoid function.

Based on the two ways above, the entity embeddings are given by

$$\mathbf{h}_{u,q}^\ell(e) = \text{ReLU}\left(\mathbf{W}_o^\ell \sum_{(e,r,x) \in \mathcal{N}(e) \wedge \{e,x\} \subset \mathcal{V}_{u,q}^\ell} \gamma_{(e,r,x)|u,q}^\ell (\mathbf{h}_{u,q}^{\ell-1}(x) \odot \hat{\mathbf{r}}_q)\right), \quad (8)$$

where $\mathbf{W}_o^\ell \in \mathbb{R}^{d \times d}$ is a learnable weight matrix, $\mathcal{V}_{u,q}^\ell$ is the entity set in progressive propagation step ℓ , and $\mathcal{V}_{u,q}^0 = \mathcal{S}_{u,q}$.

4.5 Training Strategy: LinkVerify

To reason the likelihood of a triplet (u, q, e) , the decoder produces a score function $s(\cdot)$. Given the final output $\mathbf{h}_{u,q}^{L_2}$ after L_2 layers of multi-condition GNN, the score function is given by

$$s(u, q, e) = \mathbf{W}_3 \left(\text{ReLU} \left(\mathbf{W}_4 (\mathbf{h}_{u,q}^{L_2}(u) \oplus \mathbf{h}_{u,q}^{L_2}(e)) \right) \right), \quad (9)$$

where $\mathbf{W}_3 \in \mathbb{R}^{1 \times d}$ and $\mathbf{W}_4 \in \mathbb{R}^{d \times 2d}$ are learnable weight matrices. However, multi-condition GNN propagates progressively and probably misses several distant target tail entities during the training. In this situation, the prediction knows nothing about the target tail entity and brings a noisy score for training.

To alleviate the problem above, we propose a mechanism **LinkVerify** to filter noisy training samples. The noisy sample represents the triplet whose target tail entity is not involved in $\mathcal{V}_{u,q}^{L_2}$. Taking the inductive KG reasoning task as a multi-label classification problem, we use the multi-class log-loss [14, 47] to optimize the model. Associated with LinkVerify, the final loss is given by

$$\mathcal{L} = \sum_{(u,q,v) \in \mathcal{F}} \left(-s(u, q, v) + \log \left(\sum_{\forall e \in \mathcal{E}} \exp(s(u, q, e)) \right) \right) \times \mathbb{1}(v \in \mathcal{V}_{u,q}^{L_2}). \quad (10)$$

5 Experiments

In this section, we perform extensive experiments to answer the questions below:

- **Q1:** Does MStar perform well on inductive KG reasoning?
- **Q2:** How does each designed module influence the performance?
- **Q3:** Whether MStar can improve reasoning ability about distant entities or not?

Table 2. Statistics of the inductive datasets. \mathcal{G} and \mathcal{G}' denote the KGs in the training and test sets, respectively.

Datasets		FB15k-237			NELL-995			WN18RR		
Versions	KGs	$ \mathcal{R} $	$ \mathcal{V} $	$ \mathcal{F} $	$ \mathcal{R} $	$ \mathcal{V} $	$ \mathcal{F} $	$ \mathcal{R} $	$ \mathcal{V} $	$ \mathcal{F} $
v1	\mathcal{G}	183	2,000	5,226	14	10,915	5,540	9	2,746	6,678
	\mathcal{G}'	146	1,500	2,404	14	225	1,034	9	922	1,991
v2	\mathcal{G}	203	3,000	12,085	88	2,564	10,109	10	6,954	18,968
	\mathcal{G}'	176	2,000	5,092	79	4,937	5,521	10	2,923	4,863
v3	\mathcal{G}	218	4,000	22,394	142	4,647	20,117	11	12,078	32,150
	\mathcal{G}'	187	3,000	9,137	122	4,921	9,668	11	5,084	7,470
v4	\mathcal{G}	222	5,000	33,916	77	2,092	9,289	9	3,861	9,842
	\mathcal{G}'	204	3,500	14,554	61	3,294	8,520	9	7,208	15,157

5.1 Experiments Settings

Datasets. We perform inductive KG reasoning experiments on the benchmark datasets proposed in GraIL [32], which are derived from WN18RR [6], FB15k-237 [34], and NELL-995 [42]. Each benchmark dataset is divided into four versions (v1, v2, v3, v4), and the size typically increases following the version number. Each version consists of training and test graphs without overlapping entities. The training graphs contain triplets for training and validation, following a split ratio of 10:1. The statistics of the datasets are presented in Table 2.

Baselines. We compare MStar with 10 inductive baselines organized into three groups, including (i) three rule-based models: RuleN [20], Neural LP [45], and DRUM [27]; (ii) two subgraph-based models: GraIL [32] and CoMPILE [18]; (iii) five C-MPNN-based models: NBFNet [50], A*Net [49], RED-GNN [47], AdaProp [48], and RUN-GNN [41].

Evaluation and Tie Policy. Following [47–49], we evaluate all the models using the filtered mean reciprocal rank (MRR) and Hits@10 metrics. The best models are chosen according to MRR on the validation dataset. Subgraph-based models typically rank each test triplet among 50 randomly sampled negative triplets, whereas C-MPNNs evaluate each triplet against all possible candidates. In this paper, we follow the latter and take the results of rule-based and subgraph-based models from [48]. Missing results are reproduced by their official code.

There are different tie policies [30] to compute MRR when several candidate entities receive equal scores. In progressive propagation, all unvisited entities are assigned identical scores. Following [41, 47], we measure the average rank among the entities in the tie, as suggested in [26]. To keep the tie policy consistent, we re-evaluate AdaProp using the official code.

Implementation Details. We implement our model using the PyTorch framework [24] and employ the Adam optimizer [13] for training. Due to the relatively

Table 3. Inductive KG reasoning results (measured with MRR). The best scores are in **bold** and the second-best scores are underlined. “-” denotes the result unavailable, and values with suffix “*” are reproduced using the released code.

Models	FB15k-237				NELL-995				WN18RR			
	v1	v2	v3	v4	v1	v2	v3	v4	v1	v2	v3	v4
RuleN	.363	.433	.439	.429	.615	.385	.381	.333	.668	.645	.368	.624
Neural LP	.325	.389	.400	.396	.610	.361	.367	.261	.649	.635	.361	.628
DRUM	.333	.395	.402	.410	.628	.365	.375	.273	.666	.646	.380	.627
GraIL	.279	.276	.251	.227	.481	.297	.322	.262	.627	.625	.323	.553
CoMPILE	.287	.276	.262	.213	.330	.248	.319	.229	.577	.578	.308	.548
NBFNet	.270	.321	.335	.288	.584	.410	.425	.287	.686	.662	.410	.601
A*Net	-	-	-	-	-	-	-	-	-	-	-	-
RED-GNN	.341	.411	.411	.421	.591*	.373*	.391*	.195*	.693	.687	.422	.642
AdaProp	.279*	.467*	.470*	.440*	.725*	.416*	.413*	.338*	.706*	.703*	.433*	.651*
RUN-GNN	<u>.397</u>	<u>.473</u>	.468	<u>.463</u>	.617*	.413*	<u>.479*</u>	.282*	.699	.697	.445	.654
MStar	.458	.526	.506	.487	.787	.540	.496	.384	.733	<u>.702</u>	<u>.442</u>	.645

Table 4. Inductive KG reasoning results (measured with Hits@10)

Models	FB15k-237				NELL-995				WN18RR			
	v1	v2	v3	v4	v1	v2	v3	v4	v1	v2	v3	v4
RuleN	.446	.599	.600	.605	.760	.514	.531	.484	.730	.694	.407	.681
Neural LP	.468	.586	.571	.593	.871	.564	.576	.539	.772	.749	.476	.706
DRUM	.474	.595	.571	.593	<u>.873</u>	.540	.577	.531	.777	.747	.477	.702
GraIL	.429	.424	.424	.389	.565	.496	.518	.506	.760	.776	.409	.687
CoMPILE	.439	.457	.449	.358	.575	.446	.515	.421	.747	.743	.406	.670
NBFNet	.530	.644	.623	.642	.795	<u>.635</u>	.606	<u>.591</u>	.827	.799	.568	.702
A*Net	<u>.535</u>	.638	.610	.630	-	-	-	-	.810	.803	.544	.743
RED-GNN	.483	.629	.603	.621	.866*	.601*	.594*	.556*	.799	.780	.524	.721
AdaProp	.461*	<u>.665*</u>	<u>.636*</u>	.632*	.776*	.618*	.580*	.589*	.796*	.792*	.532*	.730*
RUN-GNN	.496	.639	.631	.665	.833*	.575*	<u>.659*</u>	.436*	.807	.798	<u>.550</u>	<u>.735</u>
MStar	.583	.702	.675	.665	.900	.735	.666	.617	<u>.817</u>	.803	.547	.726

small size of the inductive dataset and its susceptibility to overfitting, we apply early stopping to mitigate this issue. We tune the hyper-parameters using grid search and select the number of starting entities n in $\{1, 2, 4, 8, 16, 32, 64\}$, the number of starting entity types m in $\{2, 3, 5, 7, 9\}$. The best hyperparameters are selected according to the MRR metric on the validation sets. All experiments are conducted on a single NVIDIA RTX A6000 GPU with 48GB memory.

5.2 Main Results (Q1)

Tables 3 and 4 depict the performance of different models on inductive KG reasoning. MStar demonstrates the best performance across all metrics on FB15k-237 and NELL-995, and compares favorably with the top models on WN18RR. We observe that (i) subgraph-based models typically perform poorly. This is because subgraphs are often sparse or empty and provide less information, particularly for distant entities. (ii) Rule-based models are generally more competitive but are still weaker compared to C-MPNN-based models. However, DRUM outperforms existing models except MStar in Hits@10 on NELL-995 (v1). NELL-995 (v1) is a special dataset and the distance between the head and tail entities for all triplets in the test graph is no longer than 3, which is very short. Thus, we conjecture that the length of the learned rules limits the reasoning capabilities of rule-based models. Differently, MStar holds an edge over these two groups of models on all datasets. This suggests that multiple starting entities in MStar alleviate the distance limit issues as much as possible when reasoning.

Compared with the best C-MPNN-based results, MStar achieves an average relative gain of 9.9% in MRR, 5.2% in Hits@10 on FB15k-237, and 13.9% in MRR, 6.1% in Hits@10 on NELL-995. Existing C-MPNN-based models typically use all entities in the KG or only the head entity as starting entities, without providing conditional information to distant entities, which can introduce excessive noise or lack sufficient information. Instead, our MStar selects multiple query-dependent starting entities adaptively and propagates conditions farther through the highway for accurate reasoning. Moreover, LinkVerify in MStar additionally reduces noisy samples in training. We also observe that the improvement of the model on WN18RR is not as pronounced as on the other datasets. To provide insights into this phenomenon, we conduct further analysis in Sect. 5.4.

5.3 Ablation Study

Variants of MStar (Q2). In this section, we design several variants of MStar to study the contributions of three components: (i) selection, (ii) highway, and (iii) LinkVerify in training. The results are summarized in Tables 5 and 6, which indicate that all components contribute significantly to MStar on the three datasets.

First, the variant of w/o selection propagates only from the head entity which is the same as RED-GNN. According to the results, removing selection significantly decreases performance, highlighting the effectiveness of using multiple starting entities to explore reasoning patterns across a broader neighborhood.

Second, it can be observed that the performance of variant w/o highway is worse than MStar. This observation suggests that transmitting query-dependent information to the starting entities is a promising approach to expedite propagation for conditions and enhance reasoning accuracy.

Third, the variant of w/o LinkVerify is inferior to MStar all the time, as triplets with unvisited target entities in training KG introduce noise. Removing LinkVerify results in poorer performance, especially on smaller datasets. For

Table 5. Ablation study of the proposed framework (measure with MRR)

Models	FB15k-237				NELL-995				WN18RR			
	v1	v2	v3	v4	v1	v2	v3	v4	v1	v2	v3	v4
MStar	.458	.526	.506	.487	.787	.540	.496	.384	.733	.702	.442	.645
w/o Selection	.432	.491	.483	.457	.719	.479	.457	.280	.721	.674	.432	.643
w/o Highway	.411	.488	.460	.474	.774	.473	.494	.297	.726	.700	.438	.629
w/o LinkVerify	.426	.517	.498	.481	.661	.502	.482	.375	.729	.698	.420	.641

Table 6. Ablation study of the proposed framework (measured with Hits@10)

Models	FB15k-237				NELL-995				WN18RR			
	v1	v2	v3	v4	v1	v2	v3	v4	v1	v2	v3	v4
MStar	.583	.702	.675	.665	.900	.735	.666	.617	.817	.803	.547	.726
w/o Selection	.534	.686	.644	.629	.775	.693	.619	.425	.811	.778	.528	.717
w/o Highway	.532	.657	.609	.644	.855	.682	.648	.532	.814	.788	.543	.698
w/o LinkVerify	.568	.699	.657	.658	.785	.695	.645	.608	.811	.797	.508	.724

Table 7. Per-distance evaluation on FB15k-237 (v1) (measured with Hits@10). “ ∞ ” indicates that the head entity fails to reach the tail entity.

Distance	Proportions	RED-GNN	AdaProp	RUN-GNN	NBFNet	MStar
1	32.68%	.813	.933	.851	.545	.948
2	12.20%	.640	.520	.740	.760	.780
3	25.37%	.433	.269	.414	.490	.471
4	7.32%	.000	.000	.267	.333	.300
5	11.22%	.000	.000	.217	.261	.174
6	3.90%	.000	.000	.000	.438	.188
7	1.46%	.000	.000	.000	.333	.000
8	1.46%	.000	.000	.000	.333	.167
9	0.00%	.000	.000	.000	.000	.000
10	0.98%	.000	.000	.000	.250	.000
∞	3.41%	.000	.000	.000	.357	.214

instance, w/o LinkVerify decreases 7.0% for FB15k-237 (v1) and 1.3% for FB15k-237 (v4) relatively. This is because the noisy triplets negatively influence training when data is lacking. Thus, LinkVerify demonstrates to be more effective when applied to KGs with fewer triplets.

Per-distance Performance (Q3). To check the reasoning ability on distant tail entities, we compare MStar with several expressive models on FB15k-237 (v1). To make the comparison more precise, we split FB15k-237 (v1) into 11

Table 8. Proportions of long-distance triplets in the KGs. The shortest distance between head and tail entities in a long-distance triplet is longer than 3.

Datasets	FB15k-237		NELL-995		WN18RR	
Versions	\mathcal{G}	\mathcal{G}'	\mathcal{G}	\mathcal{G}'	\mathcal{G}	\mathcal{G}'
v1	15.78%	29.76%	39.64%	0.00%	34.31%	17.55%
v2	8.69%	15.48%	10.62%	2.52%	20.86%	16.33%
v3	3.41%	4.51%	11.16%	3.96%	22.32%	26.94%
v4	2.39%	2.74%	9.30%	6.98%	22.39%	20.50%

subsets according to the shortest distance between the head and tail entity for each triplet. The comparisons are conducted on each subset based on official code and parameters. RED-GNN, AdaProp and MStar use 3 layers of GNN. RUN-GNN and NBFNet use 5 and 6 layers of GNN, respectively. The results are shown in Table 7.

Compared to the models with a single starting entity (RED-GNN, AdaProp, and RUN-GNN), MStar performs better significantly on distant entities. For instance, RED-GNN fails to predict entities beyond 3 hops. Moreover, MStar can even reason about unreachable target entities. This is because MStar can select query-related starting entities that are disconnected from the head entity but in the neighborhood of the unreachable entities. These observations demonstrate that multiple starting entities can expand the reasoning area effectively, and the highway layer provides additional evidence for reasoning about distant entities.

Differently, the reasoning performance of NBFNet on close entities is significantly decreased despite the ability to reason about distant entities. For instance, NBFNet is inferior to the other models on Hits@10 for 1-distance triplets with a great gap of at least 0.268. This is because NBFNet propagates from query-independent starting entities and reasons along many noisy relational paths, which disrupts the inference about close entities. Instead, MStar improves the reasoning performance for distant entities and keeps the reasoning abilities for close entities simultaneously. This is achieved due to MStar propagating conditions along query-related relational paths and removing noisy links by LinkVerify.

5.4 Further Analysis

Perspective of Datasets. As shown in Tables 5 and 6, the improvement of MStar on WN18RR is not as great as the one on other datasets. As can be seen from Table 2, WN18RR (v1, v2, v3, v4) and NELL-995 (v1) have fewer relations. Due to the entity-independent nature of inductive KG reasoning, entity embeddings usually rely on the representation of relations. With fewer relations, entities carry more monotonous information. Therefore, it becomes challenging to select query-dependent entities and propagate messages to the target ones. To study the situation further, we count the proportion of triplets whose shortest distance between the head and tail entities exceeds 3. We regard these triplets

Table 9. Comparison of different starting entities selection methods

Models	FB15k-237 (v1)		NELL-995 (v1)		WN18RR (v1)	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
MStar	.462	.598	.801	.921	.736	.816
w/random	.427	.587	.787	.901	.698	.803
w/degree	.403	.553	.362	.595	.709	.810

as long-distance triplets. The result is shown in Table 8. We can see that NELL-995 (v1) owns zero long-distance triplets in the test graphs. Thus, NELL-995 (v1) can resolve the above issues by propagating conditional information to any target entity in 3 hops, even without multiple starting entities.

Perspective of Starting Entities Selection. MStar leverages an importance score function to select starting entities. The score function is conditioned on the query head and relation, aiming to explore query-dependent entities. Here, we consider two other score function variants, i.e., variant w/ random and variant w/ degree. Variant w/ random scores the entities with random values. Similar to EL-GNN [25], variant w/ degree assigns higher scores to entities with higher degrees. All variants keep top- n entities as starting ones.

Table 9 shows the comparison results. We can observe that random scores lead to a degraded performance. This is because random starting entities propagate along many noisy relational paths. Noisy paths hinder MStar’s ability to capture query-related rules and to reach distant target tail entities. Variant w/ degree is also inferior to our MStar, even worse than random scores. For instance, the performance of variant w/ degree on FB15k-237 (v1) decreases by 54.8% and 54.0% relative to MStar and variant w/ random, respectively. This is mainly due to the fact that the global feature degree fixes the starting entities and cannot support query-dependent propagation.

6 Conclusion and Future Work

In this paper, we explore the issue of inefficient message propagation for KG reasoning and propose a new inductive KG reasoning model called MStar. Specifically, we propose using multiple starting entities to expand the propagation scope. Moreover, we construct a highway between the head entity and the other starting entities to accelerate conditional message passing. Additionally, we introduce a training strategy LinkVerify to filter inappropriate samples. Experimental results demonstrate the effectiveness of MStar. In particular, ablation results validate the superiority of MStar for reasoning about distant entities. In future work, we plan to explore alternative modules for selecting and classifying starting entities. We also intend to investigate methods to effectively utilize noisy triplets during training instead of dropping them.

Acknowledgments. We thank the anonymous reviewers for their valuable comments. This work was supported by the National Natural Science Foundation of China under Grant 62272219 and the Collaborative Innovation Center of Novel Software Technology & Industrialization.

Supplemental Material Statement. The source code and hyperparameters are available at our GitHub repository: <https://github.com/nju-websoft/MStar>.

References

1. Bollacker, K.D., Evans, C., Paritosh, P.K., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Proc. of SIGMOD, pp. 1247–1250 (2008)
2. Bordes, A., Usunier, N., García-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Proc. of NeurIPS, vol. 26, pp. 2787–2795 (2013)
3. Chen, J., He, H., Wu, F., Wang, J.: Topology-aware correlations between relations for inductive link prediction in knowledge graphs. In: Proc. of AAAI, vol. 35, pp. 6271–6278 (2021)
4. Chen, X., Jia, S., Xiang, Y.: A review: Knowledge reasoning over knowledge graph. *Expert Syst. Appl.* **141**, 112948 (2020)
5. Chen, Z., Wang, X., Wang, C., Li, Z.: PosKHG: a position-aware knowledge hyper-graph model for link prediction. *Data Sci. Eng.* **8**, 135–145 (2023)
6. Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.: Convolutional 2D knowledge graph embeddings. In: Proc. of AAAI, vol. 32, pp. 1811–1818 (2018)
7. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural message passing for quantum chemistry. In: Proc. of ICML, vol. 70, pp. 1263–1272 (2017)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. of CVPR, pp. 770–778 (2016)
9. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proc. of CVPR, pp. 2261–2269 (2017)
10. Huang, X., Romero, M., Ceylan, İ.İ., Barceló, P.: A theory of link prediction via relational Weisfeiler-Leman on knowledge graphs. In: Proc. of NeurIPS, vol. 36, pp. 19714–19748 (2023)
11. Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S.: A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Networks Learn. Syst.* **33**, 494–514 (2022)
12. Jia, T., Yang, Y., Lu, X., Zhu, Q., Yang, K., Zhou, X.: Link prediction based on tensor decomposition for the knowledge graph of COVID-19 antiviral drug. *Data Intell.* **4**, 134–148 (2022)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Proc. of ICLR, pp. 1–13 (2015)
14. Lacroix, T., Usunier, N., Obozinski, G.: Canonical tensor decomposition for knowledge base completion. In: Proc. of ICML, vol. 80, pp. 2869–2878 (2018)
15. Lehmann, J., et al.: DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web* **6**, 167–195 (2015)
16. Lin, Q., et al.: Incorporating context graph with logical reasoning for inductive relation prediction. In: Proc. of SIGIR, pp. 893–903 (2022)

17. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proc. of AAAI, vol. 29, pp. 2181–2187 (2015)
18. Mai, S., Zheng, S., Yang, Y., Hu, H.: Communicative message passing for inductive relation reasoning. In: Proc. of AAAI, vol. 35, pp. 4294–4302 (2021)
19. Meilicke, C., Chekol, M.W., Ruffinelli, D., Stuckenschmidt, H.: Anytime bottom-up rule learning for knowledge graph completion. In: Proc. of IJCAI, pp. 3137–3143 (2019)
20. Meilicke, C., Fink, M., Wang, Y., Ruffinelli, D., Gemulla, R., Stuckenschmidt, H.: Fine-grained evaluation of rule- and embedding-based systems for knowledge graph completion. In: Proc. of ISWC, vol. 11136, pp. 3–20 (2018)
21. Mitchell, T.M., et al.: Never-ending learning. *Commun. ACM* **61**, 103–115 (2018)
22. Nickel, M., Tresp, V., Kriegel, H.: A three-way model for collective learning on multi-relational data. In: Proc. of ICML, pp. 809–816 (2011)
23. Pan, Y., Liu, J., Zhang, L., Zhao, T., Lin, Q., Hu, X., Wang, Q.: Inductive relation prediction with logical reasoning using contrastive representations. In: Proc. of EMNLP, pp. 4261–4274 (2022)
24. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. In: Proc. of NeurIPS, vol. 32, pp. 8024–8035 (2019)
25. Qiu, H., Zhang, Y., Li, Y., Yao, Q.: Logical expressiveness of graph neural network for knowledge graph reasoning. In: Proc. of ICLR, pp. 1–21 (2020)
26. Rossi, A., Barbosa, D., Firmani, D., Matinata, A., Merialdo, P.: Knowledge graph embedding for link prediction: a comparative analysis. *ACM Trans. Knowl. Discov. Data* **15**, 14:1–14:49 (2021)
27. Sadeghian, A., Armandpour, M., Ding, P., Wang, D.Z.: DRUM: End-to-end differentiable rule mining on knowledge graphs. In: Proc. of NeurIPS, vol. 32, pp. 15321–15331 (2019)
28. Schlichtkrull, M.S., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: Proc. of ESWC, vol. 10843, pp. 593–607 (2018)
29. Sun, Z., Deng, Z., Nie, J., Tang, J.: RotatE: knowledge graph embedding by relational rotation in complex space. In: Proc. of ICLR, pp. 1–18 (2019)
30. Sun, Z., Vashishth, S., Sanyal, S., Talukdar, P.P., Yang, Y.: A re-evaluation of knowledge graph completion methods. In: Proc. of ACL, pp. 5516–5522 (2020)
31. Szegedy, C., et al.: Going deeper with convolutions. In: Proc. of CVPR, pp. 1–9 (2015)
32. Teru, K.K., Denis, E.G., Hamilton, W.L.: Inductive relation prediction by subgraph reasoning. In: Proc. of ICML, vol. 119, pp. 9448–9457 (2020)
33. Tian, L., Zhou, X., Wu, Y.P., Zhou, W.T., Zhang, J.H., Zhang, T.S.: Knowledge graph and knowledge reasoning: a systematic review. *J. Electron. Sci. Technol.* **20**, 100159 (2022)
34. Toutanova, K., Chen, D., Pantel, P., Poon, H., Choudhury, P., Gamon, M.: Representing text for joint embedding of text and knowledge bases. In: Proc. of EMNLP, pp. 1499–1509 (2015)
35. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: Proc. of ICML, vol. 48, pp. 2071–2080 (2016)
36. Vashishth, S., Sanyal, S., Nitin, V., Talukdar, P.P.: Composition-based multi-relational graph convolutional networks. In: Proc. of ICLR, pp. 1–16 (2020)
37. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: Proc. of ICLR, pp. 1–12 (2018)
38. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **29**, 2724–2743 (2017)

39. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proc. of AAAI, vol. 28, pp. 1112–1119 (2014)
40. Wu, F., Jing, X., Wei, P., Lan, C., Ji, Y., Jiang, G., Huang, Q.: Semi-supervised multi-view graph convolutional networks with application to webpage classification. *Inf. Sci.* **591**, 142–154 (2022)
41. Wu, S., Wan, H., Chen, W., Wu, Y., Shen, J., Lin, Y.: Towards enhancing relational rules for knowledge graph link prediction. In: Proc. of EMNLP Findings, pp. 10082–10097 (2023)
42. Xiong, W., Hoang, T., Wang, W.Y.: DeepPath: a reinforcement learning method for knowledge graph reasoning. In: Proc. of EMNLP, pp. 564–573 (2017)
43. Xu, X., Zhang, P., He, Y., Chao, C., Yan, C.: Subgraph neighboring relations infomax for inductive link prediction on knowledge graphs. In: Proc. of IJCAI, pp. 2341–2347 (2022)
44. Yang, B., Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: Proc. of ICLR, pp. 1–12 (2015)
45. Yang, F., Yang, Z., Cohen, W.W.: Differentiable learning of logical rules for knowledge base reasoning. In: Proc. of NeurIPS, vol. 30, pp. 2319–2328 (2017)
46. Zhang, W., Yao, Z., Chen, M., Huang, Z., Chen, H.: NeuralKG-ind: a Python library for inductive knowledge graph representation learning. In: Proc. of SIGIR, pp. 3140–3144 (2023)
47. Zhang, Y., Yao, Q.: Knowledge graph reasoning with relational digraph. In: Proc. of WWW, pp. 912–924 (2022)
48. Zhang, Y., Zhou, Z., Yao, Q., Chu, X., Han, B.: AdaProp: learning adaptive propagation for graph neural network based knowledge graph reasoning. In: Proc. of KDD, pp. 3446–3457 (2023)
49. Zhu, Z., et al.: A*Net: a scalable path-based reasoning approach for knowledge graphs. In: Proc. of NeurIPS, vol. 36, pp. 59323–59336 (2023)
50. Zhu, Z., Zhang, Z., Xhonneux, L.A.C., Tang, J.: Neural Bellman-Ford networks: a general graph neural network framework for link prediction. In: Proc. of NeurIPS, vol. 34, pp. 29476–29490 (2021)