A Proof of Theorem 1 and 2

Theorem 1 *Single-source edge-wise GNN can learn rule-induced subgraph representation if* $\uplus = +, \oplus = +, \diamond = +$, $\otimes^1 = \times, \otimes^2 = \times$. *i.e., there exists nonzero* $\alpha_{i,j}$ *such that*

$$\mathbf{e}^k_{u,r_t,v} = \sum_{i=1}^{k} \underbrace{\sum_{(u,r_t,v)} \sum_{(v,y_0,x_0)} \dots \sum_{(x_{i-3},y_{i-2},u)}}_{i} \alpha_{i1}\mathbf{r}_{r_t} \times \alpha_{i2}\mathbf{r}_{y_0} \times \dots \times \alpha_{ii}\mathbf{r}_{y_{i-2}} \tag{15}$$

Proof In this case, the rule-induced subgraph representation is:

$$\mathcal{S}_{u,r_t,v} = \sum_{i=1}^{k} \underbrace{\sum_{(u,r_t,v)} \sum_{(v,y_0,x_0)} \dots \sum_{(x_{i-3},y_{i-2},u)}}_{i} \alpha_{i1}\mathbf{r}_{r_t} \times \alpha_{i2}\mathbf{r}_{y_0} \times \dots \times \alpha_{ii}\mathbf{r}_{y_{i-2}} \tag{16}$$

Then we will show that single-source edge-wise GNN can learn this rule-induced sugraph representation in **induction**.

$k = 1$. we have

$$\mathbf{e}^1_{u,r_t,v} = \mathbf{h}^1_u + \mathbf{r}_{r_t} = \mathbf{r}_{r_t} + \sum_{(x_0,y_0,u)\in\mathcal{T}} (\mathbf{h}^0_{x_0} + \mathbf{e}^0_{x_0,y_0,u}) \times \mathbf{r}_{y_0}$$

Note that $\mathbf{e}^0_{x_0,y_0,u} \neq 0$ if and only if $(x_0, y_0, u) = (u, r_t, v)$. However, this is impossible as $u \neq v$. Thus $\mathbf{e}^1_{u,r_t,v}$ satisfies the definition of rule-induced subgraph representation.

$k = 2$. we have:

$$\begin{aligned}
\mathbf{e}^2_{u,r_t,v} = \mathbf{h}^2_u + \mathbf{e}^1_{u,r_t,v} &= \sum_{(x_0,y_0,u)\in\mathcal{T}} (\mathbf{h}^1_{x_0} + \mathbf{e}^1_{x_0,y_0,u}) \times \mathbf{r}_{y_0} + \mathbf{r}_{r_t} \\
&= \sum_{(x_0,y_0,u)\in\mathcal{T}} 2\mathbf{h}^1_{x_0} \times \mathbf{r}_{y_0} + \mathbf{r}_{r_t} \\
&= \sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} 2(\mathbf{h}^0_{x_1} + \mathbf{e}^0_{x_1,y_1,x_0}) \times \mathbf{r}_{y_1} \times \mathbf{r}_{y_0} + \mathbf{r}_{r_t} \\
&= \sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} 2\mathbf{e}^0_{x_1,y_1,x_0} \times \mathbf{r}_{y_1} \times \mathbf{r}_{y_0} + \mathbf{r}_{r_t}
\end{aligned} \tag{17}$$

We can find that $e^0_{x_1,y_1,x_0} \neq 0$ if and only if $(x_1, y_1, x_0) = (u, r_t, v)$, i.e. there exists both (u, r_t, v) and (v, r_t, u). Obviously, $\mathbf{e}^2_{u,r_t,v}$ satisfies the definition of rule-induced subgraph representation.

Assume that this conclusion exists for $n \leq k - 1$. Now we check the k-th term.

$$\mathbf{e}^k_{u,r_t,v} = \mathbf{h}^k_u + \sum_{i=1}^{k-1} \underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \dots \sum_{(v,y_{i-2},x_0)}}_{i-1} \alpha_{i1}\mathbf{r}_{r_t} \otimes \alpha_{i2}\mathbf{r}_{y_{i-2}} \otimes \dots \otimes \alpha_{ii}\mathbf{r}_{y_0} \tag{18}$$

First, we consider $\mathbf{h}_u^k$.

$$
\begin{aligned}
\mathbf{h}_u^k &= \sum_{(x_0,y_0,u)} (\mathbf{h}_{x_0}^{k-1} + \mathbf{e}_{x_0,y_0,u}^{k-1}) \times \mathbf{r}_{y_0} \\
&= \sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} (\mathbf{h}_{x_1}^{k-2} + \mathbf{e}_{x_1,y_1,x_0}^{k-2}) \times \mathbf{r}_{y_1} \times \mathbf{r}_{y_0} + \sum_{(x_0,y_0,u)} \mathbf{e}_{x_0,y_0,u}^{k-1} \times \mathbf{r}_{y_0} \\
&= \underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \dots \sum_{(x_{k-1},y_{k-1},x_{k-2})}}_{k} \mathbf{e}_{x_{k-1},y_{k-1},x_{k-2}}^{0} \times \mathbf{r}_{y_{k-1}} \times \mathbf{r}_{y_{k-2}} \times \dots \times \mathbf{r}_{y_0} \\
&+ \underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \dots \sum_{(x_{k-2},y_{k-2},x_{k-3})}}_{k-1} \mathbf{e}_{x_{k-2},y_{k-2},x_{k-3}}^{1} \times \mathbf{r}_{y_{k-2}} \times \dots \times \mathbf{r}_{y_0} \\
&+ \dots \\
&+ \sum_{(x_0,y_0,u)} \mathbf{e}_{x_0,y_0,u}^{k-1} \times \mathbf{r}_{y_0}
\end{aligned}
\tag{19}
$$

Notice that $\underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \dots \sum_{(x_{k-1},y_{k-1},x_{k-2})}}_{k} \mathbf{e}_{x_{k-1},y_{k-1},x_{k-2}}^{0} \times \mathbf{r}_{y_{k-1}} \times \mathbf{r}_{y_{k-2}} \times \dots \times \mathbf{r}_{y_0} \neq 0$ if and only if $(x_{k-1}, y_{k-1}, x_{k-2}) = (u, r_t, v)$. In this situation, this term is exactly the k-th term in the expression of $\mathbf{e}_{u,r_t,v}^k$. Now we want to prove that:

$$
\begin{aligned}
&\underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \dots \sum_{(x_{k-2},y_{k-2},x_{k-3})}}_{k-1} \mathbf{e}_{x_{k-2},y_{k-2},x_{k-3}}^{1} \times \mathbf{r}_{y_{k-2}} \times \dots \times \mathbf{r}_{y_0} \\
&+ \dots \\
&+ \sum_{(x_0,y_0,u)} \mathbf{e}_{x_0,y_0,u}^{k-1} \times \mathbf{r}_{y_0}
\end{aligned}
\tag{20}
$$

can be fused in top $k-1$ term of Equation. 16. Let's check the j-th term of Equation. 20.

$$
\begin{aligned}
&\underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \cdots \sum_{(x_{j-1},y_{j-1},x_{j-2})}}_{j} \mathbf{e}^{k-j}_{x_{j-1},y_{j-1},x_{j-2}} \times \mathbf{r}_{y_{j-1}} \times \ldots \times \mathbf{r}_{y_0} \\
&= \underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \cdots \sum_{(x_{j-1},y_{j-1},x_{j-2})}}_{j} (\mathbf{e}^{k-j-1}_{x_{j-1},y_{j-1},x_{j-2}} + \mathbf{h}^{k-j}_{x_{j-1}}) \times \mathbf{r}_{y_{j-1}} \times \ldots \times \mathbf{r}_{y_0} \\
&= \underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \cdots \sum_{(x_{j-1},y_{j-1},x_{j-2})}}_{j} (\mathbf{e}^{0}_{x_{j-1},y_{j-1},x_{j-2}} + \mathbf{h}^{k-j}_{x_{j-1}} + \ldots + +\mathbf{h}^{1}_{x_{j-1}}) \times \mathbf{r}_{y_{j-1}} \times \ldots \times \mathbf{r}_{y_0} \\
&= \underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \cdots \sum_{(x_{j-1},y_{j-1},x_{j-2})}}_{j} \mathbf{e}^{0}_{x_{j-1},y_{j-1},x_{j-2}} \times \mathbf{r}_{y_{j-1}} \times \ldots \times \mathbf{r}_{y_0} \\
&+ \underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \cdots \sum_{(x_{j-1},y_{j-1},x_{j-2})} \sum_{(x_j,y_j,x_{j-1})}}_{j+1} \\
&(\mathbf{h}^{k-j-1}_{x_j} + \ldots + \mathbf{h}^{0}_{x_j} + \mathbf{e}^{k-j-1}_{x_j,y_j,x_{j-1}} + \ldots + \mathbf{e}^{0}_{x_j,y_j,x_{j-1}}) \times \mathbf{r}_{y_{j-1}} \times \ldots \times \mathbf{r}_{y_0}
\end{aligned} \tag{21}
$$

Note that $\mathbf{e}^0_{x_{j-1},y_{j-1},x_{j-2}} \neq 0$ if and only if $(x_{j-1}, y_{j-1}, x_{j-2}) = (u, r_t, v)$, thus the term can be fused into the j-th term of Equation. 16. $\mathbf{e}^0_{x_j,y_j,x_{j-1}}$ can be fused into the $(j+1)$-th term and so on. Therefore, we have:

$$
\begin{aligned}
&\underbrace{\sum_{(x_0,y_0,u)} \sum_{(x_1,y_1,x_0)} \cdots \sum_{(x_{j-1},y_{j-1},x_{j-2})}}_{j} \mathbf{e}^{k-j}_{x_{j-1},y_{j-1},x_{j-2}} \times \mathbf{r}_{y_{j-1}} \times \ldots \times \mathbf{r}_{y_0} \\
&= \sum_{i=1}^{k} \underbrace{\sum_{(u,r_t,v)} \sum_{(v,y_0,x_0)} \cdots \sum_{(x_{i-3},y_{i-2},u)}}_{i} \alpha_{i1}\mathbf{r}_{r_t} \times \alpha_{i2}\mathbf{r}_{y_0} \times \ldots \times \alpha_{ii}\mathbf{r}_{y_{i-2}}
\end{aligned} \tag{22}
$$

There, we prove that single-source edge-wise GNN can learn rule-induced subgraph representation in this case. □

Theorem 2 *Single-source edge-wise GNN can learn rule-induced subgraph representation if* $\uplus = \oplus, \oplus = \oplus, \diamond = \oplus$, $\otimes^1 = \otimes, \otimes^2 = \otimes$, *where* $\oplus$ *and* $\otimes$ *are binary operators that satisfy* $0 \oplus a = a, 0 \otimes a = 0$. *i.e., there exists nonzero* $\alpha_{i,j}$ *such that*

$$
\mathbf{e}^k_{u,r_t,v} = \bigoplus_{i=1}^{k} \underbrace{\bigoplus_{(u,r_t,v)} \bigoplus_{(v,y_0,x_0)} \cdots \bigoplus_{(x_{i-3},y_{i-2},u)}}_{i} \alpha_{i1}\mathbf{r}_{r_t} \otimes \alpha_{i2}\mathbf{r}_{y_0} \otimes \ldots \otimes \alpha_{ii}\mathbf{r}_{y_{i-2}} \tag{23}
$$

Proof Without loss of generality, we can replace $+$ with $\oplus$ and $\times$ with $\otimes$ to represent a binary operator, then we directly get this theorem. Note that we should ensure that $\oplus$ and $\otimes$ satisfy $0 \oplus a = a, 0 \otimes a = 0$, which we use in the process of proof. □

B Details of Datasets

We summarize the details of inductive relation prediction benchmark datasets in Table 5.

Table 5: Statistics of three inductive datasets, which contain four different versions individually. We use #E and #R and #TR to denote the number of entities, relations, and triples.

		WN18RR			**FB15k-237**			**NELL-995**		
		#R	#E	#TR	#R	#E	#TR	#R	#E	#TR
v1	train	9	2746	6678	183	2000	5226	14	10915	5540
	test	9	922	1991	146	1500	2404	14	225	1034
v2	train	10	6954	18968	203	3000	12085	88	2564	10109
	test	10	2923	4863	176	2000	5092	79	4937	5521
v3	train	11	12078	32150	218	4000	22394	142	4647	20117
	test	11	5084	7470	187	3000	9137	122	4921	9668
v4	train	9	3861	9842	222	5000	33916	77	2092	9289
	test	9	7208	15157	204	3500	14554	61	3294	8520

C Implementation Details

In general, our proposed method is implemented in DGL[34] and PyTorch[33] and trained on single GPU of NVIDIA GeForce RTX 3090. We apply Adam optimizer[37] with an initial learning rate of 0.0005. Observing that batch size has little effect on the performance of the model, We adjust batch size as large as possible for different datasets to accelerate training. We use the binary cross entropy loss.The maximum number of training epochs is set to 10. During training, we add reversed edges to fully capture relevant rules. The number of hop h is set to 3 which is consistent with existing subgraph-based methods. We conduct grid search to obtain optimal hyperparameters, where we search subgraph types in {enclosing, unclosing}, embedding dimensions in {16, 32}, number of GNN layers in {3, 4, 5, 6} and dropout in {0, 0.1, 0.2}. Configuration for the best performance of each dataset is given within the code.