A Effectiveness of the threshold η

In practice, $|T_{r,head}|$ or $|T_{r,tail}|$ can be a large set containing long tail noisy types, which may weaken the performance of the prior model. To remove noisy types, we introduce a threshold η. Type $t \in T_{r,*}$ will be kept if

$$w_{r,*}(t) \geq W_{r,*}^{min} + \eta \times (W_{r,*}^{max} - W_{r,*}^{min}) \tag{1}$$

where $W_{r,*}^{max}$ and $W_{r,*}^{min}$ are the maximun and the minimal weight of the type set $T_{r,*}$ respectively. If $w_{r,*}(t)$ is not sufficiently high(i.e., Eq.1 is not satisfied), type t will be removed from $T_{r,*}$ without further consideration. $* = \{\text{head}, \text{tail}\}$. Threshold η is chosen based on the relation prediction performance on the validation set and varies from datasets to datasets. On each dataset, we manually select the optimal threshold η from $\{0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9\}$ that achieves the best performance(Hits@1) on the validation set. On FB15K and YAGO26K-906, $\eta = 0.1$. On DB111K-174, $\eta = 0$. Relation prediction performance of the prior model with different thresholds η on the validation set is shown in Table 1, Table 2 and Table 3 for FB15K, YAGO26K-906 and DB111K-174 respectively.

Table 1: Effectiveness of the threshold η on FB15K

Threshold	Ave.$\lvert T_{r,head}\rvert$	Ave.$\lvert T_{r,tail}\rvert$	MR	Hits@1	Hits@10
0.0	46	43	3.47	52.2	95.6
0.1	20	19	2.78	**64.2**	96.6
0.2	14	14	**2.68**	60.8	**97.2**
0.4	7	7	3.56	57.8	**97.2**
0.6	5	5	5.11	50.4	96.6
0.8	4	4	6.12	46.8	97.0
0.9	3	3	9.04	42.2	96.8

Table 2: Effectiveness of the threshold η on YAGO26K-906

Threshold	Ave.$\lvert T_{r,head}\rvert$	Ave.$\lvert T_{r,tail}\rvert$	MR	Hits@1	Hits@10
0.0	55	59	**3.27**	79.3	**89.2**
0.1	6	5	3.34	**79.6**	88.6
0.2	4	3	5.24	43.2	81.1
0.4	2	1	6.92	40.3	80.6
0.6	1	1	8.86	15.1	80.7
0.8	1	1	8.90	15.1	80.4
0.9	1	1	9.52	9.2	80.4

B Type collection for YAGO26K-906 and DB111K-174

Both YAGO26K-906 and DB111K-174 contains two KGs: instance KG $\mathcal{G}^{inst} = \{\mathcal{E}^{inst}, \mathcal{R}^{inst}\}$ and ontology KG $\mathcal{G}^{onto} = \{\mathcal{T}^{inst}, \mathcal{R}^{onto}\}$. Two KGs are connected to each other through type links. Type links and ontology KG are collected as prior type information. In particular, for each

Table 3: Effectiveness of the threshold η on DB111K-174

Threshold	Ave.$\lvert T_{r,head}\rvert$	Ave.$\lvert T_{r,tail}\rvert$	MR	Hits@1	Hits@10
0.0	7	12	**2.64**	**54.9**	**96.6**
0.1	2	3	13.76	51.5	89.1
0.2	2	2	25.77	50.6	81.1
0.4	1	2	28.30	47.8	71.2
0.6	1	1	28.60	21.3	56.6
0.8	1	1	30.14	21.5	56.3
0.9	1	1	30.74	21.5	58.1

entity $e \in \mathcal{E}^{inst}$, if there is a type link connecting entity e to type $t \in \mathcal{T}^{onto}$, we have the type t as the most specific type regarding to entity e. Given type $t \in \mathcal{E}^{onto}$, if there exists a triple $(t, is\_a, t') \in \mathcal{G}^{onto}$, where $t' \in \mathcal{T}^{onto}$ and $is\_a \in \mathcal{R}^{onto}$. we take t' as the parent of type t, which is of one level lower than type t within the hierarchy. We then repeat the procedure finding the parent of type t', and we stop when we reach the root type that doesn't have any parent in $\mathcal{G}^{onto}$. We denote the root type as t^{root} and we collect the types with its hierarchy as $H = \{t^{root}/.../t'/t\}$ regarding to entity $e \in \mathcal{E}^{inst}$.

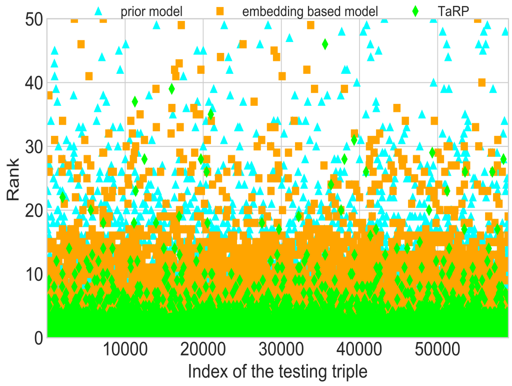
C Evaluation of the TaRP model

We evaluate the TaRP model by comparing it to three baseline embedding based models. In addition to the results shown in the paper(Table 5), we provide more insights of the results by visualizing the performance of the prior model, embedding based model and the TaRP model. We report the performance with RotatE being the baseline embedding based model and visualize the performance on FB15K and DB111K-174. In Figure 1, we plot the rank of all the testing triples and in Figure 2, we visualize the number of triples within different rank ranges. As we can see from Figure 1 and Figure 2, the ranks of testing triples produced by the TaRP are concentrated within in rank 1- rank 10. The number of triples that are of rank 10 or higher are significantly reduced by combing the prior model with the embedding based model.

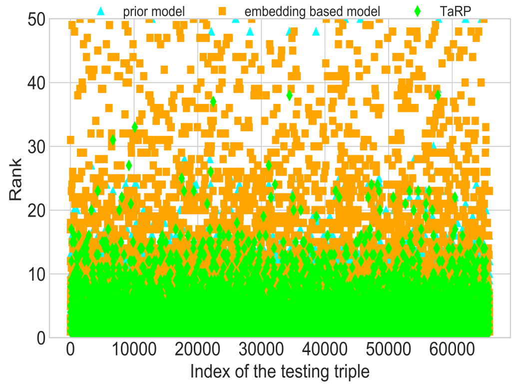
In addition, we report some specific examples to better illustrate the effectiveness of the proposed FaRP model. The examples are from FB15K. RotatE is applied as the embedding based model. The cases are listed in Table 4.

Table 4: Case Studies.

Rank		
Prior model	Embedding-based model	TaRP
(`Meet Dave`, `/people/person/gender`, `Male`)		
5	1169	10
(`Police`, `/gardening_hint/split_to`, `Police procedural`)		
158	1333	469
(`Jean-Claude Van Damme`, `/martial_artist/martial_art`, `Taekwondo`)		
1	1345	6

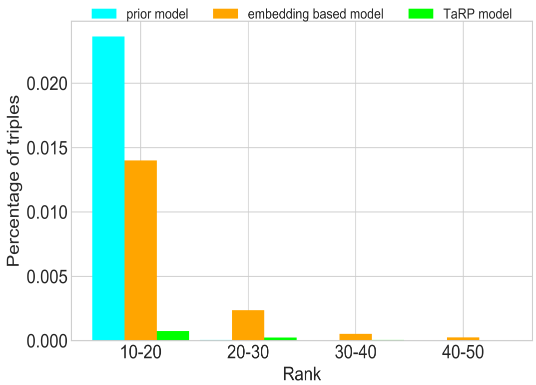


(a) FB15K

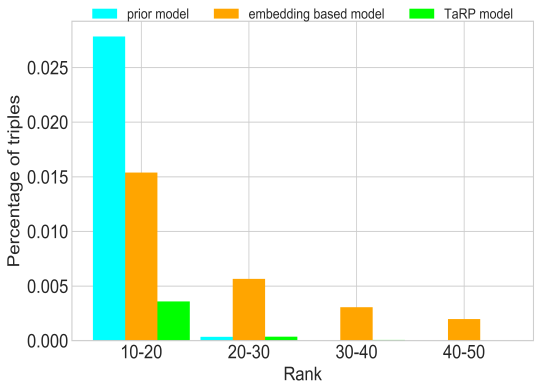


(b) DB111K-174

Figure 1: Distribution of ranks of testing triples



(a) FB15K



(b) DB111K-174

Figure 2: Histograms of ranks of testing triples