A PROOF FOR THEOREM 1

First, we can easily prove the cross-entropy loss is L_1-Smooth. Then we prove the $\mathcal{L}_{\text{rentention}}$ is L_2-Smooth. To simplify, we set $k=1$. Then we have $\Delta_y x(y)=\frac{1}{1-y}+\frac{1}{2}\frac{1-2y}{y-y^2}$. Notice that we set $0<\|y\|<1$, hence we can easily prove that $|\Delta_{y_1}-\Delta_{y_2}|\leq L_1\|y_1-y_2\|$. In this way, we can have that the loss function in Eq.(9) is $H/2$-smooth loss function (we can let $H/2=\alpha L_1+\beta L_2$), and it can be easily shown that $\mathcal{F}$ is H-smooth.

(1) For any $k\in[0,K]$, we can have

$$\begin{aligned}\mathcal{F}\left(\boldsymbol{w}_{k+1}\right) &\leq \mathcal{F}\left(\boldsymbol{w}_k\right)+\nabla\mathcal{F}\left(\boldsymbol{w}_k\right)^T\left(\boldsymbol{w}_{k+1}-\boldsymbol{w}_k\right)+\frac{H}{2}\left\|\boldsymbol{w}_{k+1}-\boldsymbol{w}_k\right\|^2\\ &=\mathcal{F}\left(\boldsymbol{w}_k\right)+\left(\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)+\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right)^T\left(-\alpha\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right)+\frac{\alpha^2H}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\\ &=\mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\alpha-\frac{\alpha^2H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\alpha\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle\end{aligned}\tag{11}$$

For the term $\langle g_1(w_k),g_2(w_k)\rangle$, it follows that

$$\begin{aligned}&\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle\\ =&\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)+\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle\\ =&\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle+\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle\\ =&\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle+\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle+\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle\end{aligned}\tag{12}$$

Notice that

$$\begin{aligned}&2\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle+\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2+\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\\ =&\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)+\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\geq 0\end{aligned}\tag{13}$$

We have

$$\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle\geq-\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\tag{14}$$

Following the same line, it can be shown that

$$\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle\geq-\frac{1}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2\tag{15}$$

Combining Eq.(12), Eq.(14) and Eq.(15) gives a lower bound on $\boldsymbol{g}_1(\boldsymbol{w}_k),\boldsymbol{g}_2(\boldsymbol{w}_k)$, *i.e.*,

$$\begin{aligned}&\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right),\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle\\ \geq&-\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\\ &-\frac{1}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2+\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle\\ \geq&-\frac{H^2}{8}\left\|\boldsymbol{w}_k-\boldsymbol{w}_0\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\\ &-\frac{H^2}{8}\left\|\boldsymbol{w}_k-\boldsymbol{w}_0\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2+\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle\\ =&-\frac{H^2}{4}\left\|\boldsymbol{w}_k-\boldsymbol{w}_0\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2+\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right),\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle,\end{aligned}\tag{16}$$

where the second inequality is true because of the smoothness of the loss function. Based on the update formulation, it can be seen that

$$\boldsymbol{w}_k=\boldsymbol{w}_0-\alpha\sum_{i=0}^{k-1}\boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\tag{17}$$

Therefore, continuing with Eq.(11), we can have

$$
\begin{aligned}
&\mathcal{F}\left(\boldsymbol{w}_{k+1}\right) \\
&\leq \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\alpha-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\alpha\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right), \boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle
\end{aligned} \tag{18}
$$

Then we have:

$$
\begin{aligned}
&\mathcal{F}\left(\boldsymbol{w}_{k+1}\right) \\
\leq &\mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\alpha-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2+\frac{\alpha^3 H^2}{4}\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2+\frac{\alpha}{2}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2 \\
&+\frac{\alpha}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\alpha\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right), \boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle \\
= &\mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2+\frac{\alpha^3 H^2}{4}\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2+\frac{\alpha}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\alpha\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_0\right), \boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\rangle \\
\leq &\mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2+\frac{\alpha^3 H^2}{4}\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2+\frac{\alpha}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2 \\
&-\alpha \epsilon_1\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\|,
\end{aligned} \tag{19}
$$

where the last inequality is based on $\left\langle\nabla \mathcal{L}_j\left(\boldsymbol{w}^j\right), \nabla \mathcal{L}_{j-1}\left(\boldsymbol{w}^{j-1}\right)\right\rangle \geq \epsilon_1\left\|\nabla \mathcal{L}_j\left(\boldsymbol{w}^j\right)\right\|_2\left\|\nabla \mathcal{L}_{j-1}\left(\boldsymbol{w}^{j-1}\right)\right\|_2$. Next, it can be shown that

$$
\alpha \leq \frac{\gamma\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|}{H B K} \leq \frac{\gamma\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|}{H\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|} \tag{20}
$$

It then follows that

$$
\begin{aligned}
&\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2+\frac{\alpha^2 H^2}{4}\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2 \\
\leq &\frac{1}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2+\frac{\gamma^2\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2}{4 H^2\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2} H^2\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2 \\
= &\frac{2+\gamma^2}{4}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2 .
\end{aligned} \tag{21}
$$

Therefore, we can obtain that

$$
\begin{aligned}
\mathcal{F}\left(\boldsymbol{w}_{k+1}\right) & \leq \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2+\frac{\alpha\left(2+\gamma^2\right)}{4}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\alpha \epsilon_1\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_0\right)\right\| \\
& \leq \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2 \\
& <\mathcal{F}\left(\boldsymbol{w}_k\right)
\end{aligned} \tag{22}
$$

where the second inequality is true because $\epsilon_1 \geqslant \frac{(2+\gamma^2)\|g_1(w_0)\|}{4\|g_2(w_0)\|}$. This sufficient decrease of the objective function value indicates that the optimal $\mathcal{F}(w*)$ can be obtained eventually for convex loss functions.

(2) For a non-convex loss function $\mathcal{L}$, we can have the following as in Eq.(11):

$$
\begin{aligned}
\mathcal{F}\left(w_{k+1}^{r}\right) \leq & \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\alpha-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\alpha\left\langle\boldsymbol{g}_1\left(\boldsymbol{w}_k\right), \boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\rangle \\
& \stackrel{(a)}{=} \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\alpha-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\frac{\alpha}{2}\left[\left\|\nabla \mathcal{F}\left(\boldsymbol{w}_k\right)\right\|^2-\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)\right\|^2-\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\right] \\
& = \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\frac{\alpha}{2}\left\|\nabla \mathcal{F}\left(\boldsymbol{w}_k\right)\right\|^2+\frac{\alpha}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)\right\|^2 \\
& = \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\frac{\alpha}{2}\left\|\nabla \mathcal{F}\left(\boldsymbol{w}_k\right)\right\|^2+\frac{\alpha}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)+\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2 \\
& \leq \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\frac{\alpha}{2}\left\|\nabla \mathcal{F}\left(\boldsymbol{w}_k\right)\right\|^2+\alpha\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_k\right)-\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2 \\
& \quad +\alpha\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2 \\
& \stackrel{(b)}{\leq} \mathcal{F}\left(\boldsymbol{w}_k\right)-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2-\frac{\alpha}{2}\left\|\nabla \mathcal{F}\left(\boldsymbol{w}_k\right)\right\|^2+\frac{H^2 \alpha^3}{4}\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2+\alpha\|\boldsymbol{g}_1(\boldsymbol{w}_0)\|^2
\end{aligned} \tag{23}
$$

where (a) is because $\delta\mathcal{F}(w_k) = \boldsymbol{g}_1(\boldsymbol{w}_k) + \boldsymbol{g}_2(\boldsymbol{w}_k)$, and (b) is because of the smoothness of $\mathcal{L}$ and Eq.(17). Therefore,

$$
\begin{aligned}
& \min_k \left\|\nabla \mathcal{F}\left(\boldsymbol{w}_k\right)\right\|^2 \\
\leq & \frac{1}{K} \sum_{k=0}^{K-1}\left\|\nabla \mathcal{F}\left(\boldsymbol{w}_k\right)\right\|^2 \\
\leq & \frac{2}{\alpha K} \sum_{k=0}^{K-1}\left[\mathcal{F}\left(\boldsymbol{w}_k\right)-\mathcal{F}\left(\boldsymbol{w}_{k+1}\right)+\frac{H^2 \alpha^3}{4}\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2+\alpha\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\left[\frac{\alpha}{2}-\frac{\alpha^2 H}{2}\right]\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2\right] \\
\leq & \frac{2}{\alpha K}\left[\mathcal{F}\left(\boldsymbol{w}_0\right)-\mathcal{F}\left(\boldsymbol{w}_K\right)\right]+\frac{H^2 \alpha^2}{2(K-1)} \sum_{k=1}^{K-1}\left\|\sum_{i=0}^{k-1} \boldsymbol{g}_2\left(\boldsymbol{w}_i\right)\right\|^2+2\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\frac{1-\alpha H}{K} \sum_{k=0}^{K-1}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2 \\
\stackrel{(a)}{\leq} & \frac{2}{\alpha K}\left[\mathcal{F}\left(\boldsymbol{w}_0\right)-\mathcal{F}\left(\boldsymbol{w}_K\right)\right]+\frac{\gamma^2}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2+2\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\frac{1-\alpha H}{K} \sum_{k=0}^{K-1}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2 \\
\leq & \frac{2}{\alpha K}\left[\mathcal{F}\left(\boldsymbol{w}_0\right)-\mathcal{F}\left(\boldsymbol{w}^*\right)\right]+\frac{4+\gamma^2}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2-\frac{1-\alpha H}{K} \sum_{k=0}^{K-1}\left\|\boldsymbol{g}_2\left(\boldsymbol{w}_k\right)\right\|^2 \\
\leq & \frac{2}{\alpha K}\left[\mathcal{F}\left(\boldsymbol{w}_0\right)-\mathcal{F}\left(\boldsymbol{w}^*\right)\right]+\frac{4+\gamma^2}{2}\left\|\boldsymbol{g}_1\left(\boldsymbol{w}_0\right)\right\|^2
\end{aligned} \tag{24}
$$

where (a) holds due to $\mathcal{F}(w*) \leq \mathcal{F}(\boldsymbol{w}_K)$ and $\|\sum_{i=0}^{k-1}\boldsymbol{g}_2(\boldsymbol{w}_i)\|^2 \leq \frac{\gamma^2}{\alpha^2 H^2}\|\boldsymbol{g}_1(\boldsymbol{w}_0)\|^2$ based on Eq.(21).

B Details for CLKGE

Abalation Study for g in Eq.(1). However, new entities may obey different distributions, hence we need to alleviate the distribution gaps by a function g. To achieve this, as stated in (Yang et al., 2022), some networks such as MLP can eliminate the difference in distribution by constructing an implicit space. In this way, for the sake of simplicity, we specify g as an MLP. Furthermore, we conduct the experiments on the ENTITY dataset as shown in Table 4, where CLKGE w/ g denotes CLKGE utilizing the and CLKGE w/o g denotes CLKGE remove g. One can observe that CLKGE w/ g can be superior to CLKGE w/o g significantly, which demonstrates the effectiveness of g.

Abalation Study for different KGE methods. Different KGE methods have somewhat impact on performance. To demonstrate this point, we conduct the experiments on ENTITY datasets utilizing

Table 4: Abalation Study for g in Eq.(1)

Models	MRR	H@1	H@3	H@10
CLKGE w g	$.233_{\pm.002}$	$.138_{\pm.001}$	$.245_{\pm.002}$	$.398_{\pm.002}$
CLKGE w/o g	$.211_{\pm.003}$	$.120_{\pm.002}$	$.233_{\pm.002}$	$.387_{\pm.003}$

Table 5: Abalation Study for different KGE methods

Models	MRR	H@1	H@3	H@10
$CLKGE_C$	$.248_{\pm.001}$	$.144_{\pm.001}$	$.278_{\pm.002}$	$.436_{\pm.002}$
$CLKGE_R$	$.250_{\pm.002}$	$.145_{\pm.002}$	$.279_{\pm.002}$	$.439_{\pm.003}$
$CLKGE_T$	$.245_{\pm.003}$	$.143_{\pm.002}$	$.274_{\pm.003}$	$.435_{\pm.003}$

the KGE method ComplEx, RotatE, and TransE as $CLKGE_C$, $CLKGE_R$, and $CLKGE_T$, respectively. As shown in the table 5, one can observe that the difference of performance for various KGE method is not big, which also show the robustness of our method.

Experiments for WN18RR-5-LS dataset. We conduct experiments on the WN18RR-5-LS dataset. Specifically, WN18RR-5-LS is divided into several snapshots. As shown in the table below, one can notice that CLKGE superiors other methods, and the overall performance shows the effectiveness of components including knowledge transfer and knowledge retention.

C Experimental Details

Learning Efficiency. In this part, we conduct experiments to compare the training time. We report the total time cost on FACT, which is easier for comparison. Table 8 shows the results. Unsurprisingly, one can observe that re-training is the most time-consuming. By contrast, our model is the most efficient, and it can converge to the optimal embedding in a fast way. The reason is two-fold. On one hand, the new knowledge can learn the representations via knowledge transfer by old knowledge effectively. On the other hand, old knowledge can alleviate catastrophic forgetting via knowledge retention with new knowledge significantly. In total, these components work together thus reducing the training time, which demonstrates the effectiveness of our method.

Ablation Study for Knowledge Retention To demonstrate the effectiveness of utilizing the energy-based manifold, we compare the performance of two versions. The first is the original version (denoted as CLKGE_1) while the retention loss in the second version (denoted as CLKGE_2) is

$$\mathcal{L}_{\text{retention}} = \sum_{e \in \mathcal{E}_{i-1}} \omega(e) \left\| \mathbf{e}_i - \mathbf{e}_{i-1} \right\|_2^2 + \sum_{r \in \mathcal{R}_{i-1}} \omega(r) \left\| \mathbf{r}_i - \mathbf{r}_{i-1} \right\|_2^2, \tag{25}$$

where where $omega(x)$ is the regularization weight for x. The results are shown in Table 9. One can observe that the performance of CLKGE_1 is superior to CLKGE_2. In this way, it validates the rationality of our method and the effectiveness of CLKGE to model the association between true embedding and the obtained embedding for knowledge.

Table 6: Experiments for WN18RR-5-LS dataset.

Models	MRR	H@1	H@3	H@10
EMR	$.351_{\pm.002}$	$.232_{\pm.002}$	$.317_{\pm.003}$	$.380_{\pm.001}$
DiCGRL	$.365_{\pm.002}$	$.244_{\pm.003}$	$.325_{\pm.002}$	$.392_{\pm.002}$
LKGE	$.372_{\pm.002}$	$.251_{\pm.002}$	$.337_{\pm.003}$	$.401_{\pm.001}$
CLKGE	$.384_{\pm.002}$	$.260_{\pm.002}$	$.347_{\pm.003}$	$.415_{\pm.002}$

Table 7: Statistical data of the four constructed growing KG datasets. For the i-th snapshot, $\mathcal{T}_{\Delta_i}$ denotes the set of new facts in this snapshot, and $\mathcal{E}_i$, $\mathcal{R}_i$ denote the sets of cumulative entities and relations in the first i snapshots, respectively.

Datasets	Snapshot 1			Snapshot 2			Snapshot 3			Snapshot 4			Snapshot 5		
	$\lvert\mathcal{T}_{\Delta_1}\rvert$	$\lvert\mathcal{E}_1\rvert$	$\lvert\mathcal{R}_1\rvert$	$\lvert\mathcal{T}_{\Delta_2}\rvert$	$\lvert\mathcal{E}_2\rvert$	$\lvert\mathcal{R}_2\rvert$	$\lvert\mathcal{T}_{\Delta_3}\rvert$	$\lvert\mathcal{E}_3\rvert$	$\lvert\mathcal{R}_3\rvert$	$\lvert\mathcal{T}_{\Delta_4}\rvert$	$\lvert\mathcal{E}_4\rvert$	$\lvert\mathcal{R}_4\rvert$	$\lvert\mathcal{T}_{\Delta_5}\rvert$	$\lvert\mathcal{E}_5\rvert$	$\lvert\mathcal{R}_5\rvert$
ENTITY	46,388	2,909	233	72,111	5,817	236	73,785	8,275	236	70,506	11,633	237	47,326	14,541	237
RELATION	98,819	11,560	48	93,535	13,343	96	66,136	13,754	143	30,032	14,387	190	21,594	14,541	237
FACT	62,024	10,513	237	62,023	12,779	237	62,023	13,586	237	62,023	13,894	237	62,023	14,541	237
HYBRID	57,561	8,628	86	20,873	10,040	102	88,017	12,779	151	103,339	14,393	209	40,326	14,541	237

Table 8: Cumulative time (seconds) cost on FACT during 5 snapshots.

Models	Re-taring	CWE	DiCGRL	PNN	GEM	EMR	EWC	SI	LKGE	Fine-tuning	CLKGE
	5000	3100	2150	1900	1750	1540	1500	1400	1300	1350	1200

Table 9: Ablation Study for Knowledge Retention.

Model	ENTITY				RELATION			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
CLKGE_1	0.248	0.144	0.278	0.436	0.203	0.115	0.226	0.379
CLKGE_2	0.234	0.136	0.271	0.429	0.196	0.107	0.219	0.372