A Dataset Statistics.

We show the statistics for the benchmarks used in this paper in Table 1.

Dataset	#Entities	#Relations	#Train	#Validation	#Test
FB15K-237	14541	237	272115	17535	20466
WN18RR	40943	11	86835	3034	3134
Kinship	104	25	3206	2137	5343
UMLS	135	46	1959	1306	3264

Table 1: Dataset statistics

B Hyper-parameters and Random Seed

In our framework, we set the dimension d to 2000, the batch size to 100, and perform a search for the learning rate in the range $\{0.05, 0.1\}$. Additionally, we search for the optimal value of β in the range $\{0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0\}$. For the ComplEx and CP models on the FB15k-237 and WN18RR datasets, we use the same hyperparameters as those used in the DURA (Zhang, Cai, and Wang 2020) and N3 (Lacroix, Usunier, and Obozinski 2018) models. For the remaining datasets and models, we use wandb to perform a grid search for the coefficient of DURA and N3 in the range $\{0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$. We simply use 0, 1, 2 as three different seeds. We run our methods in the main experiment with these seeds and report the mean and standard deviation.

C Correlation between $\hat{\mathrm{cr}}$ and MRR on UMLS and Kinship

The correlation between $\hat{\mathrm{cr}}$ and MRR on UMLS and Kinship are shown in Fig.1. Their correlations are still strong.

D Computational Complexity

The original tensor factorization model has $\mathrm{O(nd)}$ complexity for one training sample. To calculate the composite risk in Eq. (25), an extra least squares problem is solved as in Sec 4.5, with $\mathrm{O(c^2d)}$ complexity, where $\mathrm{c} = |\mathrm{connected(h)}|$. Due to the sparsity of KG, c^2 is usually much smaller than n, ensuring computational efficiency.

E Effect on CP and DistMult

We show the effect of CP and DistMult over WN18RR and UMLS in Table 2. It can be observed that CompilE consistently improves the performance. Notably, on UMLS, CompilE$_D$ improved MRR by 5.9% over previous DistMult-based models. This supports the claim that reducing composition risk is effective for different tensor factorization models.

F Effect on Different Relation Type

We compare the prediction performance on different types of relations in WN18RR, the results are shown in Table 3. We refer to (Lacroix, Usunier, and Obozinski 2018) to categorize relation types. The performance of CompilE$_N$ is consistent with the need of compositional reasoning ability. 1-1 relations do not require compositional reasoning. So after enhancing the compositional reasoning ability, the performance of CompilE$_N$ slightly decreases. For 1-m, m-1, and m-m relations, compositional reasoning is more important, so the performance of CompilE$_N$ is improved or remains unchanged.

	WN18RR		UMLS	
	MRR	H@10	MRR	H@10
CP	0.438	0.482	0.819	0.964
+DURA	0.477	0.550	0.836	0.964
+N3	0.470	0.545	0.841	0.971
+CompilE$_D$	**0.478**	**0.552**	0.840	0.968
+CompilE$_N$	0.472	0.546	**0.844**	**0.971**
DistMult	0.440	0.499	0.725	0.954
+DURA	0.460	0.547	0.746	0.939
+N3	0.452	0.537	0.745	0.934
+CompilE$_D$	**0.461**	**0.548**	**0.805**	**0.962**
+CompilE$_N$	0.456	0.545	0.790	0.947

Table 2: Effect on CP and DistMult.

	1-1	1-m	m-1	m-m
ComplEx	**0.976**	0.169	0.186	0.943
+N3	0.964	0.204	0.250	0.943
+CompilE$_N$	0.952	**0.208**	**0.272**	**0.943**

Table 3: Effect on different relation type

G Proxies for Composition Risk Estimation

We explore using entity similarity to estimate the composite risk and found comparable results. More specifically, we defined $\mathrm{reliable(h)}$ as entities similar to h based on Jaccard coefficients of their neighborhood distribution. Table 4 compares the results of connectivity and Jaccard, demonstrating comparable performance.

	WN18RR		UMLS	
	MRR	H@10	MRR	H@10
$\mathrm{CompilE_D(connectivity)}$	0.494	0.580	**0.867**	**0.973**
$\mathrm{CompilE_D(Jaccard)}$	**0.495**	**0.581**	0.852	0.970
$\mathrm{CompilE_N(connectivity)}$	**0.493**	**0.588**	0.854	0.967
$\mathrm{CompilE_N(Jaccard)}$	0.492	0.582	**0.862**	**0.969**

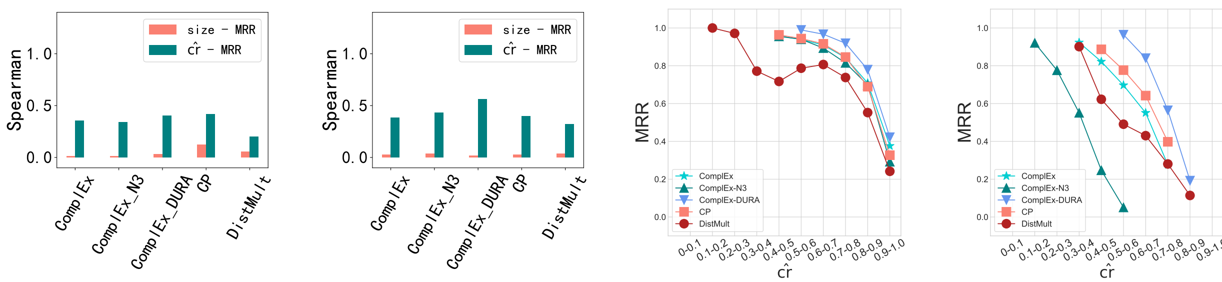
Table 4: Connectivity vs Jaccard.

H Ablation Study for Regularizer-free Setting

We compare the results of original ComplEx model and the ComplEx with only composite risk. In this regularizer-free scenario, the performances are still improved with composite risk optimization on FB15k-237 and WN18RR as shown in Table 5

	FB15k-237				WN18RR			
Method	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
ComplEx	0.350	0.259	0.386	0.531	0.460	0.429	0.471	0.521
+composite risk	**0.353**	**0.261**	**0.389**	**0.533**	**0.463**	**0.431**	**0.475**	**0.527**

Table 5: Ablation study for regularizer-free setting



(a) UMLS (b) Kinship (c) UMLS (d) Kinship

Figure 1: Correlation between the approximated composition risk and MRR on UMLS and Kinship.

I Proof of Theorem 4.2

Proof. From the definition, we have:

$$\hat{\mathrm{cr}}(\mathrm{h},\mathrm{r},\mathrm{t}) = \min_{\mathbf{a}} \frac{||\mathbf{e}(\mathrm{h},\mathrm{r},\mathrm{t}) - \sum_{\mathrm{h_i} \in \mathrm{connected(h)} \cap \mathrm{KG(r)}} \mathbf{a}_\mathrm{i} \mathbf{e}(\mathrm{h_i},\mathrm{r},\mathrm{t})||}{||\mathbf{e}(\mathrm{h},\mathrm{r},\mathrm{t})||} \tag{1}$$

Note that $\mathrm{reliable(h)} \subseteq \mathrm{connected(h)}$, so we have

$$\min_{\mathbf{a}} \frac{||\mathbf{e}(\mathrm{h},\mathrm{r},\mathrm{t}) - \sum_{\mathrm{h_i} \in \mathrm{reliable(h)} \cap \mathrm{KG(r)}} \mathbf{a}_\mathrm{i} \mathbf{e}(\mathrm{h_i},\mathrm{r},\mathrm{t})||}{||\mathbf{e}(\mathrm{h},\mathrm{r},\mathrm{t})||} \geq \min_{\mathbf{a}} \frac{||\mathbf{e}(\mathrm{h},\mathrm{r},\mathrm{t}) - \sum_{\mathrm{h_i} \in \mathrm{connected(h)} \cap \mathrm{KG(r)}} \mathbf{a}_\mathrm{i} \mathbf{e}(\mathrm{h_i},\mathrm{r},\mathrm{t})||}{||\mathbf{e}(\mathrm{h},\mathrm{r},\mathrm{t})||} \tag{2}$$

Thus according to the definition of $\mathrm{cr}(\mathrm{h},\mathrm{r},\mathrm{t})$ and $\hat{\mathrm{cr}}(\mathrm{h},\mathrm{r},\mathrm{t})$, we have $\mathrm{cr}(\mathrm{h},\mathrm{r},\mathrm{t}) \geq \hat{\mathrm{cr}}(\mathrm{h},\mathrm{r},\mathrm{t})$, which means $\hat{\mathrm{cr}}(\mathrm{h},\mathrm{r},\mathrm{t})$ is a lower bound. □

References

Lacroix, T.; Usunier, N.; and Obozinski, G. 2018. Canonical Tensor Decomposition for Knowledge Base Completion. In *ICML*.

Zhang, Z.; Cai, J.; and Wang, J. 2020. Duality-induced regularizer for tensor factorization based knowledge graph completion. *Advances in Neural Information Processing Systems*, 33: 21604–21615.