RSCF: Relation-Semantics Consistent Filter for Entity Embedding of Knowledge Graph

Anonymous ACL submission

Abstract

In knowledge graph embedding, leveraging relation-specific entity-transformation has markedly enhanced performance. However, this approach lacks assurance for consistent changes in relation and entity embeddings due to the disconnected entity-transformation representation, missing valuable inductive bias among semantically similar relations. Furthermore, a generalized plug-in approach as a SFBR disrupts this consistency through excessive concentration of entity embeddings under entity-based regularization, generating indistinguishable score distributions among relations. To tackle these challenges, we introduce *Relation-Semantics Consistent Filter* (RSCF), characterized by three features: 1) shared affine transformation of relation embeddings across all relations, 2) rooted entity-transformation that adds an entity embedding to its change represented by the transformed vector, and 3) normalization of the change to prevent scale reduction. In knowledge graph completion tasks with distance-based and tensor decomposition models, RSCF notably enhances performance across all relations, regardless of their frequency.

1 Introduction

Knowledge graphs (KGs) play crucial roles in a wide area of machine learning and its applications (Liu et al., 2021; Zhang et al., 2022b; Zhou et al., 2022; Fang et al., 2017; Gao et al., 2019; Cao et al., 2019; Geng et al., 2022; Wang et al., 2019). KGs, even on a large scale, still suffer from a lack of data. For example, 71% of people in Freebase have no birthplace information, and 75% have no nationality information (Dong et al., 2014). This incompleteness issue has been extensively studied as a task to predict missing entity information, called as knowledge graph completion (KGC). An effective approach for KGC is knowledge graph embedding (KGE) that learns vectors to represent entities and relations in a low dimensional space to measure the validity of triples. Two primary approaches to determine the validity are distance-based model (DBM) using the Minkowski distance and tensor decomposition model (TDM) regarding KGC as a tensor completion problem (Zhang et al., 2020b). A recently tackled issue of the models is to learn only single embedding for an entity, which is insufficient to express its various attributes in complex relation patterns such as 1-N, N-1 and N-N (Wang et al., 2014; Chao et al., 2021; Ge et al., 2023). A proposed and effective approach for this issue is entity-transformation based model (ETM) that uses relation-specific transformations to generate different entity embeddings for relations from their original embedding, enabling more complex entity and relation learning (Liang et al., 2021; Wang et al., 2014; Chao et al., 2021; Ge et al., 2023).

ETMs, however, have a limit to learning useful inductive bias that could be obtained in semantically similar relations. For example, Semantic Filter Based on Relations (SFBR), a recently proposed method plugged in to various KGE models (Liang et al., 2021), assigns mutually disconnected relation-specific transformation to each relation. Furthermore, under a significantly useful regularizer such as DURA (Zhang et al., 2020b), especially on TDM, the method critically concentrates entity embeddings, including unobserved entities and generates indistinguishable score distributions across relations. Both issues are interpreted as limited learning an important and implicit inductive bias that semantically similar relation embeddings have similar relation-specific entity-transformation, called *relation-semantics consistency* in this paper.

To alleviate the issues, we present a simple and effective method *Relation-Semantically Consistent Filter* (RSCF), incorporating three features: 1) shared affine transformation for consistency mapping of relations to entity-transformations, 2) rooted entity-transformation representation using the affine transformation to generate only the change of an entity-embedding subsequently added by this embedding and 3) normalization of the change for preventing critical scale reduction breaking consistency. Our contributions are as follows.

- We raise and clarify two problems of entity-transformation models in learning inductive bias in terms of *relation-semantics consistency*.
- We propose a novel and significantly outperforming relation-semantics consistent filter (RSCF) to induce the consistency as a plug-in method to KGE models.
- We provide experimental results on common benchmarks of KGC, and in-depth analysis to verify the causes and derived effects.

2 Loss of Useful Inductive Bias

Because semantically similar relations have similar embedding (Yang et al., 2015), we define that mapping relation embeddings to entity-transformations is *relation-semantically consistent* if and only if any relation pairs (r_1, r_2) and shorter pair (r_1, r_3) for a given r_1 are mapped to entity-transformation pair (T_1, T_2) and shorter pair (T_1, T_3), respectively. This consistency serves as an inductive bias implying that semantically similar relations have similar entity-transformations and, therefore, overall similar entity embeddings. Two phenomena of losing this inductive bias and their causes are as follows.

Disconnection of Entity-Transformations Disconnected entity-transforamtions loosely use this bias, especially under lack of triplet data. In existing methods, relation-specific entity-transformations use separate parameters such as $\mathsf{h}_\mathsf{r} = \mathsf{W}_r\mathsf{h}$ and $\mathsf{t}_\mathsf{r} = \mathsf{W}_r\mathsf{t}$ (Liang et al., 2021), where h, t, are head and tail entity embedding, and W_r is a relation-specific transformation. Despite the disconnection, the methods can still learn similar W_r for given two similar relation embeddings if their desirable entity ranks are similar. However, limited observation of entities due to sparse triplet data introduces a wide variety of possible entity-transformations and their corresponding embedding distributions, thereby diluting consistency. In this environment, the disconnected representation without any specific training and initialization process aiming to foster the consistency is exposed to the loss of useful inductive bias of similar relations.

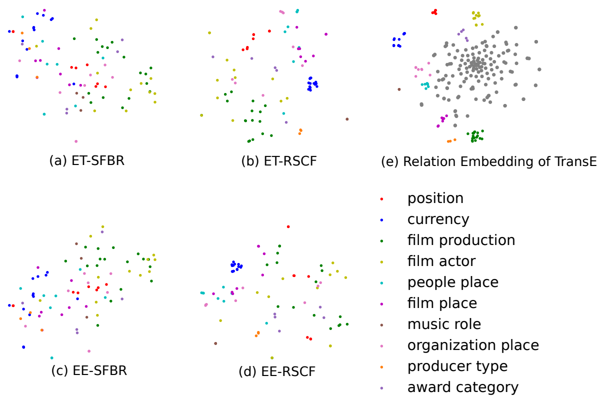


Figure 1: Head entity-transformations and entity embeddings for semantically similar relation groups. (a) and (b) indicate ET of SFBR and RSCF, (c) and (d) indicate EE of SFBR and RSCF. Points in the same color are relations in the same group. Clearly distinct groups are selected from the original TransE (e)

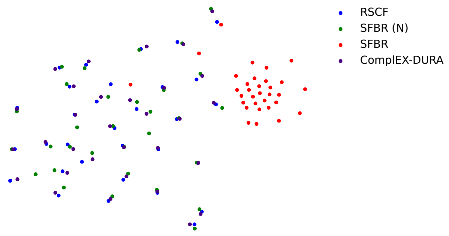
Metric	ET-SFBR	EE-SFBR	ET-RSCF	EE-RSCF
Concentration Score (↑)	0.91	0.43	1.71	1.22
Inter Cluster Distance (↑)	0.27	0.40	0.99	0.75

Table 1: Concentration score and inter cluster distance of SFBR and RSCF. Concentration score shows numerical results of their in-cluster concentration and inter cluster distance presents numerical results of the distance between different clusters.

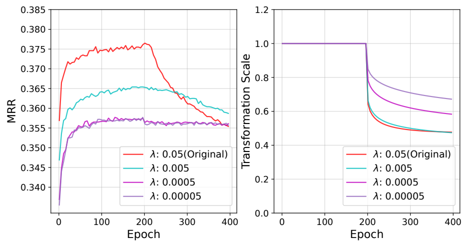
Empirical Evidence for Disconnection Figure 1 shows the impact of the disconnection via T-SNE visualization of head entity-transformations (ET) and corresponding entity embeddings (EE) of TransE-SFBR (SFBR) and TransE-RSCF (RSCF). We split them into relation groups, defined by clearly clustered relation groups in original TransE that are presented in Figure 1 (e)[1]. An entity for EE is randomly selected on FB15k-237. The ET and EE distribution of SFBR are mostly dispersed between semantically different relation groups, which implies the limit in inducting relation-semantics consistency, while RSCF methods show more in-cluster concentration. Also, the higher concentration score and inter cluster distance of RSCF in Table 1 supports this phenomenon. See more detailed distributions in Appendix J, and concentration score and inter cluster distance are in Appendix E.

Entity Embedding Concentration In particular, SFBR additionally loses consistency under entity-based regularization, DURA (Zhang et al., 2020b). In KGE based on TDM, DURA has shown signif-

[1] For details about the relation group, please refer to the Appendix G.



(a) Score Distribution



(b) MRR of validation set (left) and Transformation Scale (right)

Figure 2: Result of entity embedding concentration, and performance and scale decrease in training. The results are collected from ComplEX with DURA regularization. DURA is applied in all epochs and SFBR is applied after 200 epochs (λ: regularization weight).

icant improvement enough to be inevitable. However, ComplEX-SFBR with DURA reduces the scale of ET, causing a strong concentration of entire entity embeddings. Observed entities are relatively safe because the score distribution is continuously adjusted to predict correct triples, but unobserved entities are critically vulnerable to the concentration causing indistinguishable score distributions for semantically different relations, implying critically broken consistency. This cause of this phenomenon is simply derived in the following equations of DURA in the original (above) (Zhang et al., 2020b) and DURA in SFBR (below).

$$\begin{array}{ll} \sum_p & ||\mathsf{h_i}\overline{\mathsf{R_j}}||_2^2+||\mathsf{h_i}||_2^2+||\mathsf{t_k}||_2^2+||\mathsf{t_k}\overline{\mathsf{R_j}}^\top||_2^2 \\ \sum_p & ||\mathsf{W}_{r_j}\mathsf{h_i}\overline{\mathsf{R_j}}||_2^2+||\mathsf{W}_{r_j}\mathsf{h_i}||_2^2+||\mathsf{t_k}||_2^2+||\mathsf{t_k}\overline{\mathsf{R_j}}^\top||_2^2 \end{array} \quad (1)$$

where $p = (h_i, r_j, t_k) \in S$ for total training data S, $\mathbf{h_i}$ and $\mathbf{t_k}$ are head and tail embeddings with indices and $\overline{\mathbf{R_j}}$ is a matrix representing relation r_j. In the equation 1, to minimize DURA loss, model always decreases the scale of ET (simple proof in Appendix D) and this causes indistinguishable score distribution in all score distributions.

Empirical Evidence for Concentration Figure 2 presents T-SNE visualization of score distributions for selected queries. We select the relation r_1 showing significantly low performance in SFBR on FB15k-237, and select all queries (h, r_1, ?) for the relation r_1 in the validation set. We then generate score distribution for each query using ComplEX-RSCF, ComplEX-SFBR, ComplEX-SFBR with normalization (SFBR (N)), and the ComplEX-DURA. The results show that SFBR concentrates embeddings into a small and clear cluster, while the other methods are diversely dispersed.

Do We Need to Use DURA regularizer? Generating indistinguishable score distributions cannot be merely resolved by handling the regularization weight. Figure 2 (b) shows the valid MRR (left) of SFBR and transformation scale (right) according to the regularizer weight λ. In training until 200 epochs, largely weighted DURA shows significant performance, but applying SFBR starts to decrease MRR and the transformation scale. The results imply that integrating SFBR with DURA causes performance degradation with scale decrease ending up in the entity embedding concentration. Also, the result of SFBR with a small weighted DURA indicates that simply excluding DURA on the TDM will critically decrease the performance.

3 Method

Overview In this section, we propose *Relation-Semantics Consistent Filter* (RSCF) to address the consistency issues. In Figure 3, the overall filtering process of RSCF, distinguished features compared to SFBR, and their intended effects are illustrated. RSCF represents the ET as an addition of original embedding and its relation-specific change (ⓒ). The change is generated by an affine transformation from relation embedding (ⓐ), and then normalized (ⓑ), described as

$$\mathbf{e_r} = \underbrace{(\overbrace{\mathrm{N}_p(\overbrace{\mathbf{rA}}^{a})}^{b} + \mathbf{1})}_{c} \otimes \mathbf{e} \quad (2)$$

where $\mathbf{A} \in \mathbf{R}^{n \times n}$ is shared affine transformation, $\mathbf{r}$ and $\mathbf{e} \in \mathbf{R}^n$ are relation and entity embedding. $\mathrm{N}_p(\mathbf{rA}) = \frac{\mathbf{rA}}{\|\mathbf{rA}\|_p}$, and $\otimes$ is an elementwise product. Detailed motivation and effects are as follows.

Shared Affine Transformation for Consistency Affine transformation maintains the parallelism of two parallel line segments after the transformation and preserves the ratio of their lengths. This

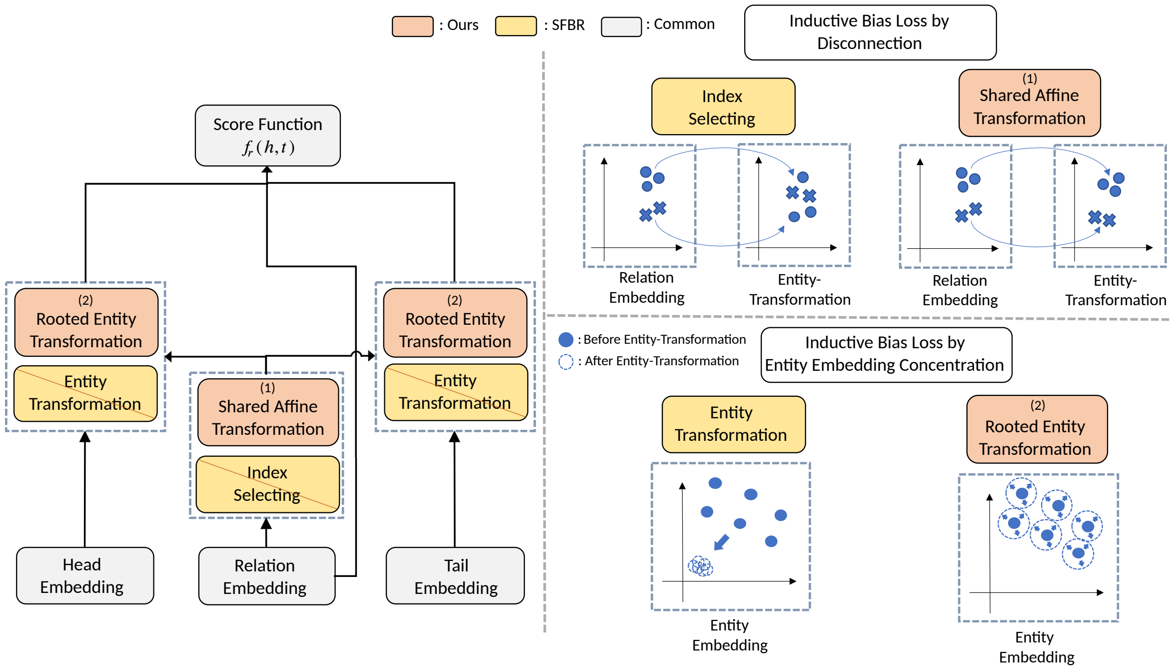


Figure 3: Overview of Relation-Semantics Consistent Filter and Its Effect. Its process (left) is illustrated on SFBR coloring changed modules. The two effects (right) are shown by comparing SFBR and RSCF on ET and entity embeddings.

property guarantees consistent mapping of relation embeddings at least on a line to generated vectors (part ⓐ in Equation 2). After normalization of the generated vectors (part ⓑ), the consistency still holds because the monotonic increase of distance between pairs is guaranteed even if the rate of lengths is not equally maintained (simple proof in Appendix C). The addition of one vector to the normalized change (part ⓒ) does not alter the inequality of distances of lines, so the consistency is again maintained. Overall, by applying the affine transformation, we can maintain the consistency between relation embedding and its ET. To implement the affine transformation shared across relations, we simply adopt a linear transformation for $\mathbf{A}$.

Rooted Entity Transformation Sharing an affine transformation overall relations inevitably reduces the expressiveness of ET compared to entirely separate relation-specific transformations such as SFBR that has shown to yield positive results (Liang et al., 2021). To mitigate the negative effects from this reduction, we decrease required expressiveness by learning only the changes in entity embeddings, rather than learning their diverse positions. Moreover, this rooted ET representation enables safely bounding changes via normalization without altering original entity embeddings. To implement it, we add one to the normalized change $\mathbf{N}_p(\mathbf{rA})$ and multiply it to the original entity-embedding (part ⓒ).

Relation Prediction to Facilitate Relation Semantic Reflection to Relation Embeddings Because relation prediction makes semantically similar relations form a cluster and can improve discrimination of dissimilar relations (Zhang et al., 2021; Chen et al., 2021), we add RP (Chen et al., 2021) to RSCF, to facilitate relation semantic reflection to relation embeddings. Therefore, training objective of RSCF can be written as:

$$\mathcal{L} = \sum_{p} \phi(\mathbf{h_r}|\mathbf{r}, \mathbf{t_r}) + \phi(\mathbf{t_r}|\mathbf{h_r}, \mathbf{r}) + \lambda\phi(\mathbf{r}|\mathbf{h}, \mathbf{t}) \quad (3)$$

where ϕ is a loss function with a score function and λ is a hyper-parameter that controls the contribution of RP. In Table 2, difference between RSCF and SFBR is presented. RSCF uses shared affine transformation to reflect relation semantics to ET, and RP (Chen et al., 2021) is only applied to RSCF.

Normalization of Change for Reducing Entity Embedding Concentration The change gener-

Model	Entity Transformation	Training Objective
SFBR	$\mathbf{e_r} = \mathbf{W_r e} + \mathbf{b}$	$\sum_p \phi(\mathbf{h_r}\vert\mathbf{r}, \mathbf{t_r}) + \phi(\mathbf{t_r}\vert\mathbf{h_r}, \mathbf{r})$
RSCF	$\mathbf{e_r} = (\mathrm{N}_p(\mathbf{rA}) + 1) \otimes \mathbf{e}$	$\sum_p \phi(\mathbf{h_r}\vert\mathbf{r}, \mathbf{t_r}) + \phi(\mathbf{t_r}\vert\mathbf{h_r}, \mathbf{r}) + \lambda\phi(\mathbf{r}\vert\mathbf{h}, \mathbf{t})$ (Chen et al., 2021)

Table 2: Difference between RSCF and SFBR, where $\mathbf{e_r}$ is transformed entity embedding that contains both head entity $\mathbf{h_r}$ and tail entity $\mathbf{t_r}$. ϕ is a loss function with a score function that depends on the model.

ated from the affine transformation is normalized by its length, expressed as $\mathbf{N}_p(\mathbf{rA})$ in the part ⓑ. This normalization alleviates critical entity embedding concentration via reducing scale decrease of transformed entity embeddings $\mathbf{e_r}$ in DURA regularization. In our relation-specific rooted ET, the change of $\mathbf{e_r}$ is simply written as

$$\|\alpha \otimes \mathbf{e}\|_p \tag{4}$$

where $\alpha = \mathbf{N}_p(\mathbf{rA})$. This value has a maximum when α has the same direction to $\mathbf{e}$. Since α is a unit vector in p-norm, $\alpha = \mathbf{e}/\|\mathbf{e}\|_p$. Then, the maximum change is

$$\|\frac{\mathbf{e}}{\|\mathbf{e}\|_p} \otimes \mathbf{e}\|_p = \|\frac{\mathbf{e}^2}{\|\mathbf{e}\|_p}\|_p = \frac{\|\mathbf{e}^2\|_p}{\|\mathbf{e}\|_p} \tag{5}$$

In practice, the elements of embedding vectors are much less than 1 in most cases. Therefore, the maximum change $\|\mathbf{e}^2\|_p/\|\mathbf{e}\|_p$ is significantly lower than the unrestricted scale change in SFBR.

Extension of RSCF The shared affine transformation can be easily extended to $Linear-2$ that is introduced in SFBR by extending shared affine transformation $\mathbf{W_e} \in \mathbf{R}^{n\times n}$ to $\mathbf{W_e} \in \mathbf{R}^{n\times 2n}$. Therefore, RSCF (Linear-2) can be written as:

$$\mathbf{W_r}^{\text{Linear-2}} = \begin{bmatrix} diag(\mathbf{w}_1) & diag(\mathbf{w}_2) \\ diag(\mathbf{w}_3) & diag(\mathbf{w}_4) \end{bmatrix} \tag{6}$$

where $\mathbf{W_r}^{\text{Linear-2}} \in \mathbf{R}^{n\times n}$ is ET built from the relation-specific change vector $\mathbf{N}_p(\mathbf{rA}) + 1$ of RSCF that is notated as concatenation of diagonal values of $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4 \in \mathbf{R}^{n/2}$.

4 Related Works

Disconnection of Relation-Specific Transformation in ETM ETM is a model that uses relation-specific ET to model various attributes of an entity. Models such as TransH (Wang et al., 2014), TransR (Lin et al., 2015), and TransD (Ji et al., 2015) are variants of TransE (Bordes et al., 2013), designed to handle complex relations by employing hyperplanes, projection matrices, and dynamic mapping matrices for their transformation functions, respectively. To address the heterogeneity and imbalance presented in TransE and its variants, TransSparse (Ji et al., 2016) utilizes adaptive sparse matrices to model different types of relations. PairRE (Chao et al., 2021) performs a scaling operation through the Hadamard product to the head and tail entities. SFBR (Liang et al., 2021) and AT (Yang et al., 2021) present a universal entity transformation applicable to both DBM and TDM. To handle complex relations in TDM, STaR (Li and Yang, 2022) integrates scaling, translation, and rotation operation for semantic matching scoring functions. ReflectE (Zhang et al., 2022a) models the transformation function using relation-specific dynamic reflection hyperplanes. CompoundE (Ge et al., 2023) applied compound operation to both head and tail entities. However, these models have no chance for inductive bias sharing due to the separate parameter of ET.

Entity Embedding Concentration in ETM SFBR (Liang et al., 2021) applies ET to both DBM and TDM. However, it also suffers from inductive bias loss due to the separate parameter of ET and indistinguishable score distribution because of the entity embedding concentration.

Dataset	Entities	Relations	Triples		
			Train	Valid	Test
WN18RR	40,943	11	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466
YAGO3-10	123,182	37	1,079,040	5,000	5,000

Table 3: Statistics of KGC Benchmark Datasets

5 Experiments

5.1 Settings

Dataset To evaluate our proposed RSCF models, we consider three KGs datasets: WN18RR (Toutanova and Chen, 2015), FB15k-237 (Dettmers et al., 2018), and YAGO3-10 (Mahdisoltani et al., 2013). The statistics for the three benchmark datasets are shown in Table 3.

Evaluation Protocol We evaluated the performance of KGC following the filtered setting (Bordes et al., 2013). The filtered setting removes all valid triples from the candidate set when evaluating the test set, except for the predicted test triple.

Distance-Based Model with Entity Transformation	WN18RR			FB15k-237		
	MRR	H@1	H@10	MRR	H@1	H@10
TransH (Wang et al., 2014) †	.220	.042	.495	.299	.201	.488
TransR (Lin et al., 2015) †	.219	.050	.498	.309	.220	.489
TransD (Ji et al., 2015) †	.211	.087	.505	.306	.218	.486
TransE-AT (Yang et al., 2021)	.479	.434	.571	.351	.257	.538
RotatE-AT (Yang et al., 2021)	.488	.438	.583	.348	.253	.537
PairRE (Chao et al., 2021)	-	-	-	.351	.256	.544
ReflectE (Zhang et al., 2022a)	.488	.450	.559	.358	.263	.546
CompoundE (Ge et al., 2023)	.491	.450	.576	.357	.264	.545
TransE-SFBR (Diag) (Liang et al., 2021)	.242	.028	.548	.338	.240	.538
TransE-SFBR (Linear-2) (Liang et al., 2021)	.263	.110	.495	.354	.258	.545
RotatE-SFBR (Diag) (Liang et al., 2021)	.489	.437	**.593**	.351	.254	.549
RotatE-SFBR (Linear-2) (Liang et al., 2021)	.490	.447	.576	.355	.258	.553
TransE-RSCF	.256	.050	.551	.356	.258	.552
TransE-RSCF (Linear-2)	.436	.378	.531	.359	.264	.549
RotatE-RSCF	.492	.447	.582	.360	.264	**.555**
RotatE-RSCF (Linear-2)	**.496**	**.455**	.581	**.362**	**.267**	.554

Table 4: Test performance of DBM-based KGC on FB15k-237 and WN18RR. Bold indicates the best result, and underlined signifies the second best result. (†: reproduced result from Zhang et al. (2020a)).

Tensor Decomposition Model with Eentity Transformation	WN18RR			FB15k-237			YAGO3-10		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
ComplEX-N3 + AT (Yang et al., 2021)	.500	.455	**.592**	.369	.273	.559	.582	.507	.712
STaR-DURA (Li and Yang, 2022)	.497	.452	.583	.368	.273	.557	.585	.513	.713
CP-DURA + SFBR †	.479	.441	.555	.368	.275	.557	.581	.510	.707
RESCAL-DURA + SFBR †	.497	.454	.576	.369	.277	.550	.578	.503	.712
ComplEX-DURA + SFBR †	.491	.450	.571	.373	.277	.563	.587	.517	.715
CP-DURA + RSCF	.482	.445	.556	.379	.287	.564	.585	.515	.711
RESCAL-DURA + RSCF	**.501**	**.460**	.581	.381	.290	.563	.582	.509	.714
ComplEX-DURA + RSCF	.497	.454	.581	**.389**	**.296**	**.575**	**.589**	**.518**	**.717**

Table 5: Test performance of TDM-based KGC on FB15k-237, WN18RR, and YAGO3-10. Bold indicates the best result, and underlined signifies the second best result. (†: reproduced results whose reference results in Appendix I).

We adopt the MRR and Hits@N to compare the performance of different KGE models. MRR is the average of the inverse mean rank of the entities and Hits@N is the proportion of correct entities ranked within top k.

Baselines and Training Protocol We compare the performance of RSCF with the two categories of KGE models: 1) DBM with entity transformation including TransH (Wang et al., 2014), TransR (Lin et al., 2015), TransD (Ji et al., 2015), PairRE (Chao et al., 2021), AT (Yang et al., 2021), SFBR (Liang et al., 2021), ReflectE (Zhang et al., 2022a) and CompoundE (Ge et al., 2023), 2) TDM with entity transformation including STaR (Li and Yang, 2022), AT (Yang et al., 2021) and SFBR (Liang et al., 2021) Because RSCF is a module that is plugged in based on existing models, we used DBM, including TransE, RotatE, and TDM, including CP, RESCAL, and ComplEX as base models. Examples of applying RSCF to based models are given in Appendix A[2].

[2] Implementation details are given in Appendix F

5.2 Performance

Performance on Distance-Based Model Table 4 shows the performance comparison of the RSCF and DBMs with ET on WN18RR and FB15k-237. Overall, RSCF shows similar or higher performance than SFBR in most settings. Even in comparison with the other DBMs, RotatE-RSCF significantly outperforms CompoundE, the state-of-the-art DBM method.

Performance on Tensor Decomposition Model Table 5 shows the performance comparison in TDMs. Compared to SFBR, RSCF shows consistent performance improvements in all datasets and settings. Furthermore, in performance comparison with the other models, ComplEX-DURA-RSCF outperforms STaR and ComplEX-N3+AT in FB15k-237 and YAGO3-10.

Ablation Study To verify the effectiveness of each component of RSCF, we conduct ablation studies on the RSCF in Table 6. In this table, RSCF shows the best performance compared with other four ablated models in both TransE and ComplEX,

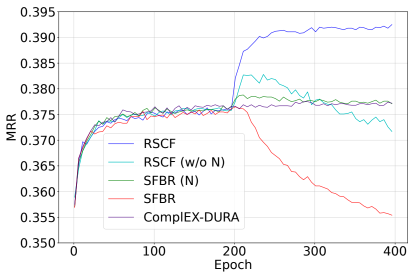


Figure 4: MRR changes over epochs of RSCF, RSCF (w/o N), SFBR (N), SFBR, and ComplEX on FB15k-237.

Model	TransE-RSCF		ComplEX-RSCF	
	MRR	H@10	MRR	H@10
w/o Affine Transformation	.354	.548	.386	.573
w/o Normalize	.353	.541	.377	.567
w/o RP (Chen et al., 2021)	.349	.549	.375	.565
w/o Affine & Normalize	.338	.531	.385	.571
RSCF	**.356**	**.552**	**.389**	**.575**

Table 6: Results of an Ablation study of RSCF on FB15k-237 datasets, TransE and ComplEX are used as base models. MRR and H@10 are used for performance comparison.

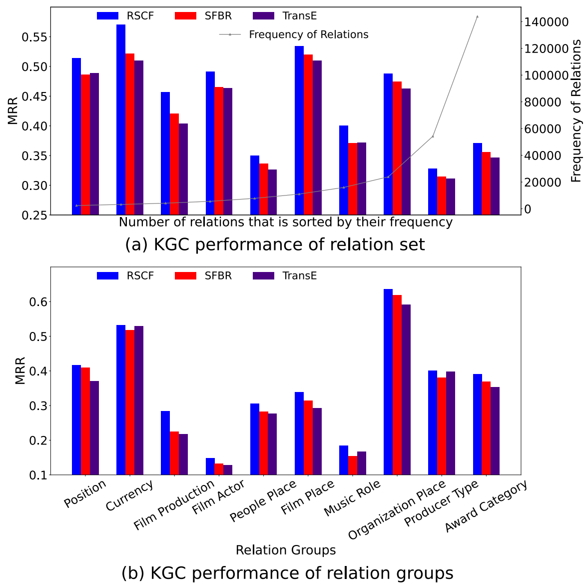


Figure 5: KGC performance of relation set that is sorted by their frequency (above) and groups of semantically similar relations observed in Figure 1 (e) (below) on FB15k-237

which suggests that each component of RSCF contributes to the effectiveness of RSCF. Especially, Figure 4 shows that w/o normalization can significantly reduce model performance and w/ normalization maintain model performance in both RSCF and SFBR in ComplEX, indicating that normalization is necessary to maintain the performance of models that use DURA regularizer[3].

Relation-Wise Performance on Relation Frequency To demonstrate the generality of applying the RSCF regardless of relation frequency, we sorted relations by their frequency in the training set and divided them into ten sets. Each set has the same number of relations. Figure 5 above shows the MRR of the validation set for each set in TransE, RSCF and SFBR. The results showed that RSCF outperformed SFBR and TransE in all sets, demonstrating the robustness of RSCF to relation frequency and showing that RSCF can be applied without trade-off between high and low frequency of relations.

Performance on Semantically Distinguished Relation Groups Figure 5 below presents the validation MRR for each relation group, defined in Figure 1 (e). RSCF outperformed SFBR and TransE in all groups. These results show that RSCF can be utilized without specific bias to the semantics of relations and indicate that reflecting the relation semantics into the transformation function can improve model performance.

Qualitative Example Analysis For qualitative analysis, Table 7 presents sampled queries, their correct answers, related triples with the sample queries, and the ranks obtained by RSCF and SFBR. Relations in sample queries and related triples belong to the same relation group (people place). In Table 7, RSCF shows enhanced performance compared to SFBR, indicating that RSCF can use trained bias between semantically similar relations.

5.3 In-Depth Analysis

Relation-Semantics Consistency of ET and EE Figure 1 shows ET and their corresponding EE of SFBR and RSCF via T-SNE. RSCF represents a more concentrated cluster compared to SFBR, which indicates that similar relations have similar ET and EE in RSCF; in other words, RSCF satisfies relation-semantic consistency.

Embedding Scale and Score Distribution Recovery Figure 6 presents transformation scale and final entity embedding scale over epochs on FB15k-237. ComplEX is used as base model. Following the approach of SFBR (Liang et al., 2021), we applied the DURA regularizer in all epochs, and RSCF, SFBR, and SFBR (N) are plugged in after 200 epochs. In the results, SFBR shows a decrease

[3]Detailed description of SFBR (N) is given in Appendix B

Query (h, r, ?)	Correct Answer	Related Triples in Training Set	Rank(R/S)
(Guillermo del Toro, /people/person/place_of_birth, ?)	Guadalajara	(Guillermo del Toro, /people/person/places_lived./people/place_lived/location, Jalisco) (Guillermo del Toro, /people/person/nationality, Mexico)	**14** / 35
(Shawn Pyfrom, /people/person/places_lived./people/place_lived/location, ?)	Florida	(Shawn Pyfrom, /people/person/place_of_birth, Tampa) (Shawn Pyfrom, /people/person/nationality, United States of America)	**4** / 32
(Walt Whitman, /people/person/places_lived./people/place_lived/location, ?)	New York	(Walt Whitman, /people/deceased_person/place_of_death, Camden) (Walt Whitman, /people/person/nationality, United States of America)	**9** / 21

Table 7: Example KGC results of RSCF compared to SFBR (R: rank of RSCF, S: rank of SFBR). Related triples show that similar relations to the queries have similar entities to the correct answers in the training set.

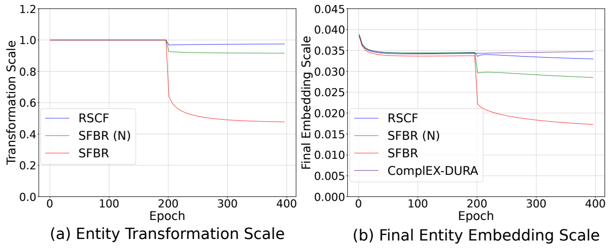


Figure 6: Entity transformation scale (left) and final entity embedding scale (right) of RSCF, SFBR (N), SFBR, and ComplEX-DURA over epochs on FB15k-237 DURA is applied in all epochs and SFBR and RSCF is applied after 200 epochs.

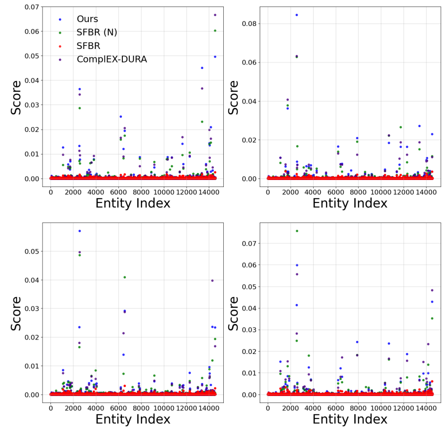


Figure 7: Score distribution of all entities for randomly selected queries from Figure 2 (a)

in both transformation scale and final entity embedding scale. In contrast, RSCF and SFBR (N) show almost no decrease in the transformation scale, and the final entity embedding scale is maintained, indicating that both RSCF and SFBR (N) can maintain the embedding scale due to normalization. Furthermore, in Figure 4, MRR decreases in SFBR and RSCF w/o normalize, while it increases in both RSCF and SFBR (N). This result implies that entity embedding concentration negatively affects to model performance.

To investigate the detailed change of score distribution that directly affects performance, we randomly sample four queries and present the score distribution of selected queries in Figure 7. In the results, SFBR shows near zero scores for most entities, and distributions for the queries are significantly similar. Applying normalization or RSCF, the diversity of scores is recovered as the original base model.

Impact on Over-Smoothed Queries To assess the impact of indistinguishable score distribution, we conducted a performance evaluation for a selected relation that shows critical entity embedding concentration in Figure 2. Table 8 presents the validation performance for all queries associated with the selected relation. SFBR shows significantly lower performance than RSCF, SFBR (N), and the ComplEX-DURA. This result implies that indistinguishable score distribution strongly affects to the accurate prediction of SFBR, simply applying normalization and RSCF gradually recovers it.

Model	MRR	H@10	Concentration
ComplEX-RSCF	**.375**	**.609**	✗
ComplEX-SFBR (N)	.366	.587	✗
ComplEX-SFBR	.267	.522	✓
ComplEX-DURA	.347	.609	✗

Table 8: KGC performance of all queries associated with the relation that shows strong concentration of entity embedding in SFBR. Concentration presents entity embedding concentration.

6 Conclusion

In this paper, we address the limit in inducing relation-semantics consistency, implying that semantically similar relations have similar entity transformation, on entity transformation models for KGC, especially SFBR. We clarify two causes, disconnected entity transformation representation and entity embedding concentration, and provide a novel relation-semantics consistent filter (RSCF) method using shared affine transform to generate the change of entity embedding, normalize it and add it to the embedding. This method significantly improves the performance of KGC compared to state-of-the-art KGE methods for overall relations.

7 Limitations

RSCF uses the simplest form of affine transformation, but it has a limit of expressing all changes across all embeddings, which requires more advanced approach. Future work should extend the method to additional KGE models to enhance generality.

References

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. 2019. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In *The world wide web conference*, pages 151–161.

Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. 2020. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6901–6914.

Linlin Chao, Jianshan He, Taifeng Wang, and Wei Chu. 2021. Pairre: Knowledge graph embeddings via paired relation vectors. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4360–4369.

Yihong Chen, Pasquale Minervini, Sebastian Riedel, and Pontus Stenetorp. 2021. Relation prediction as an auxiliary training objective for improving multi-relational graph representations. In *3rd Conference on Automated Knowledge Base Construction*.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 601–610.

Yuan Fang, Kingsley Kuan, Jie Lin, Cheston Tan, and Vijay Chandrasekhar. 2017. Object detection meets knowledge graphs. International Joint Conferences on Artificial Intelligence.

Junyu Gao, Tianzhu Zhang, and Changsheng Xu. 2019. I know the relationships: Zero-shot action recognition via two-stream graph convolutional networks and knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8303–8311.

Xiou Ge, Yun Cheng Wang, Bin Wang, and C-C Jay Kuo. 2023. Compounding geometric operations for knowledge graph completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6947–6965.

Shijie Geng, Zuohui Fu, Juntao Tan, Yingqiang Ge, Gerard De Melo, and Yongfeng Zhang. 2022. Path language modeling over knowledge graphsfor explainable recommendation. In *Proceedings of the ACM Web Conference 2022*, pages 946–955.

Cosimo Gregucci, Mojtaba Nayyeri, Daniel Hernández, and Steffen Staab. 2023. Link prediction with attention applied on multiple knowledge graph embedding models. In *Proceedings of the ACM Web Conference 2023*, pages 2600–2610.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*, pages 687–696.

Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. 2016. Knowledge graph completion with adaptive sparse transfer matrix. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Jiayi Li and Yujiu Yang. 2022. Star: Knowledge graph embedding by scaling, translation and rotation. In *International Conference on AI and Mobile Services*, pages 31–45. Springer.

Zongwei Liang, Junan Yang, Hui Liu, and Keju Huang. 2021. A semantic filter based on relations for knowledge graph completion. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7920–7929.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

Ye Liu, Yao Wan, Lifang He, Hao Peng, and S Yu Philip. 2021. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6418–6425.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.

Mojtaba Nayyeri, Chengjin Xu, Franca Hoffmann, Mirza Mohtashim Alam, Jens Lehmann, and Sahar Vahdati. 2021. Knowledge graph representation learning using ordinary differential equations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9529–9548.

Guanglin Niu, Bo Li, Yongfei Zhang, and Shiliang Pu. 2022. Cake: A scalable commonsense-aware framework for multi-view knowledge graph completion. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2867–2877.

Aleksandar Pavlović and Emanuel Sallinger. 2022. Expressive: A spatio-functional embedding for knowledge graph completion. In *The Eleventh International Conference on Learning Representations*.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2018. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd workshop on continuous vector space models and their compositionality*, pages 57–66.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR.

Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019. Multi-task feature learning for knowledge graph enhanced recommendation. In *The world wide web conference*, pages 2000–2010.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28.

Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*.

Jinfa Yang, Yongjie Shi, Xin Tong, Robin Wang, Taiyan Chen, and Xianghua Ying. 2021. Improving knowledge graph embedding using affine transformations of entities corresponding to each relation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 508–517.

Jinfa Yang, Xianghua Ying, Yongjie Shi, Xin Tong, Ruibin Wang, Taiyan Chen, and Bowei Xing. 2022. Knowledge graph embedding by adaptive limit scoring loss using dynamic weighting strategy. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1153–1163.

Qianjin Zhang, Ronggui Wang, Juan Yang, and Lixia Xue. 2022a. Knowledge graph embedding by reflection transformation. *Knowledge-Based Systems*, 238:107861.

Siheng Zhang, Zhengya Sun, and Wensheng Zhang. 2020a. Improve the translational distance models for knowledge graph embedding. *Journal of Intelligent Information Systems*, 55:445–467.

Wenqian Zhang, Shangbin Feng, Zilong Chen, Zhenyu Lei, Jundong Li, and Minnan Luo. 2022b. Kcd: Knowledge walks and textual cues enhanced political perspective detection in news media. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4129–4140.

Yao Zhang, Xu Zhang, Jun Wang, Hongru Liang, Wenqiang Lei, Zhe Sun, Adam Jatowt, and Zhenglu Yang. 2021. Generalized relation learning with semantic correlation awareness for link prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4679–4687.

Yongqi Zhang, Zhanke Zhou, Quanming Yao, and Yong Li. 2022c. Efficient hyper-parameter search for knowledge graph embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2715–2735.

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

Yucheng Zhou, Xiubo Geng, Tao Shen, Guodong Long, and Daxin Jiang. 2022. Eventbert: A pre-trained model for event correlation reasoning. In *Proceedings of the ACM Web Conference 2022*, pages 850–859.

A Special Cases with RSCF

Let $\mathbf{h_r}$, $\mathbf{t_r}$ are transformed head and tail embedding by RSCF, then the score function $d_r(\mathbf{h}, \mathbf{r})$ of TransE-RSCF can be expressed as:

$$d_r(\mathbf{h}, \mathbf{r}) = \|\mathbf{h_r} + \mathbf{r} - \mathbf{t_r}\| \quad (7)$$

The score function $d_r(\mathbf{h}, \mathbf{r})$ of RotatE-RSCF can be expressed as:

$$d_r(\mathbf{h}, \mathbf{r}) = \|\mathbf{h_r} \circ \mathbf{r} - \mathbf{t_r}\| \quad (8)$$

The score function $d_r(\mathbf{h}, \mathbf{r})$ of RESCAL-RSCF can be expressed as:

$$d_r(\mathbf{h}, \mathbf{r}) = \|\mathbf{h_r} \mathbf{r} \mathbf{t}\| \quad (9)$$

In TDM, tail embeddings are not transformed according to the settings of SFBR in order to reduce computational costs.

B SFBR with Normalization

To prevent entity embedding concentration, We apply normalization to SFBR that is presented as SFBR (N). Let $\mathbf{W_r}$ is relation-specific ET using separate parameters, then SFBR with normalization can be written as:

$$\mathbf{N}_p(\mathbf{W_r}) + 1 \tag{10}$$

where $\mathrm{N}_p(\mathbf{W_r}) = \frac{\mathbf{W_r}}{\|\mathbf{W_r}\|_p}$. Additionally, transformed entity embedding can be described as:

$$\mathbf{e_r} = (\mathbf{N}_p(\mathbf{W_r}) + 1)\mathbf{e} \tag{11}$$

where $\mathbf{e}$ is a original entity embedding.

C Proof of Consistency of Normalized Change

For any relation embedding r_1, r_2, r_3 on a line and their mapped ET T_1, T_2, T_3 by an affine transform, then the consistency holds. Let T_2 is on $\overline{T_1T_3}$, and $\overline{r_1r_2}$ is shorter $\overline{r_2r_3}$. Then, $\overline{T_1T_2} < \overline{T_2T_3}$ by the properties of afffine transformation. Because T_2 is an interpolated point of T_1 and T_3, $\angle T_1OT_2 < \angle T_2OT_3$. After normalization, let ETs projected on T_1', T_2', and T_3'. $\overline{T_1'T_2'} = \sqrt{2 - 2\cos\angle T_1'OT_2'}$, and $\overline{T_2'T_3'} = \sqrt{2 - 2\cos\angle T_2'OT_3'}$ by simple cosine rule. Cosine function is monotonically decreasing for angles less than π. Therefore, $\overline{T_1'T_2'} < \overline{T_2'T_3'}$

D Proof of scale decrease of ET

The gradient of $w_{r_j,n}$ (n-th element of w_{r_j}) in DURA can be calculated as:

$$\begin{aligned}
&\sum_p \frac{dL}{dw_{r_j,n}} ||\mathsf{w}_{\mathsf{r_j,n}}\mathsf{h}_{\mathsf{i,n}}\mathsf{r}_{\mathsf{j,n}}||_2^2 + ||\mathsf{w}_{r_jn}\mathsf{h}_{\mathsf{i,n}}||_2^2 \\
&= \sum_p \frac{dL}{dw_{r_j,n}} \mathsf{w}_{\mathsf{r_j,n}}^2(\mathsf{h}_{\mathsf{i,n}}\mathsf{r}_{\mathsf{j,n}})^2 + \mathsf{w}_{r_jn}^2 {\mathsf{h}_{\mathsf{i,n}}}^2 \\
&= \sum_p 2\mathsf{w}_{\mathsf{r_j,n}}(\mathsf{h}_{\mathsf{i,n}}\mathsf{r}_{\mathsf{j,n}})^2 + 2\mathsf{w}_{r_jn}{\mathsf{h}_{\mathsf{i,n}}}^2
\end{aligned} \tag{12}$$

The gradient of ET shows that the gradient of $w_{r_j,n}$ has always same sign with $w_{r_j,n}$ parameters. Therefore, gradient descent always reduces the scale of the parameters regardless of their sign.

E Measurement of Cluster Concentration

To measure cluster concentration, we defined concentration score as follows:

$$\sum_k^n \sum_i^m \frac{||(k_i - C_k)||}{n||C_k||} \tag{13}$$

where k is clear and mutually decoupled clusters and k_i is i-th vector embedding of ET in group k and C_k is centroid of cluster k that can be calculated as:

$$C_k = \frac{\sum_i^m k_i}{m} \tag{14}$$

In the equation 13, the vector norm of C ($||C||$) is used because of the relative concentration score for clusters. We use the reciprocal of equation 13 as concentration score. Also to evaluate the distance between different clusters we defined inter cluster distance score as follows:

$$\sum_k^n \frac{||(C_k - C_{kc})||}{\sum_i^m ||k_i||} \tag{15}$$

where C_{kc} represent the centroid that is closest to C_k, and it can be written as:

$$C_{kc} = \min_{k \neq j}\{||C_k - C_j||\} \tag{16}$$

In Equation 15, the norm of the cluster, which is calculated as the sum of the elements in the cluster, is used for the relative inter cluster distance.

F Implementation Details

When training the RSCF, we followed the experimental settings described in the SFBR (Liang et al., 2021). Following the setting of SFBR, RSCF and RSCF (Linear-2) are applied to both head and tail entities in DBM, and RSCF is applied to the only head entity in TDM. The hyper-parameters in DBM are consistent with the hyper-parameters in Sun et al. (2018), and hyper-parameters of TDM are consistent with the hyper-parameters in Zhang et al. (2020b). The presented results of RSCF represent the best performance among the three runs for each model. Experiments for the DBM were conducted on an NVIDIA 3090 GPU with 24GB of memory, while experiments for the TDM were conducted on an NVIDIA 2080TI with 11GB.

G Relation Groups for Entity Transformation

Figure 1 (e) illustrates the relation embedding of TransE. We select ten relation groups whose relation embeddings build clear and mutually decoupled clusters, which implies semantically distinguished relation groups. The other relations are plotted as grey points. The relations corresponding to each group are listed in Table 12. Note that similar relations belong to the same group.

Embedding based Model	WN18RR			FB15k-237		
	MRR	H@1	H@10	MRR	H@1	H@10
TransE (Bordes et al., 2013)	.226	-	.501	.294	-	.465
DistMult (Yang et al., 2015)	.430	.390	.490	.241	.155	.419
ComplEX (Trouillon et al., 2016)	.440	.410	.510	.247	.158	.428
RotatE (Sun et al., 2018)	.476	.428	.571	.338	.241	.533
ROTH (Chami et al., 2020)	.496	.449	.586	.348	.252	.540
ComplEX-DURA (Zhang et al., 2020b)	.491	.449	.571	<u>.371</u>	<u>.276</u>	<u>.560</u>
FieldE (Nayyeri et al., 2021)	.48	.44	.57	.36	.27	.55
KGTuner (Zhang et al., 2022c)	.484	.440	.562	.352	.263	.530
RotatE-IAS (Yang et al., 2022)	.483	**.467**	.570	.339	.242	.532
CAKE (Niu et al., 2022)	-	-	-	.321	.227	.515
STaR-DURA (Li and Yang, 2022)	<u>.497</u>	.452	.583	.368	.273	.557
ExpressivE (Pavlović and Sallinger, 2022)	.482	.407	**.619**	.350	.256	.535
SEPA (Gregucci et al., 2023)	**.500**	.454	<u>.591</u>	.360	.264	.549
CompoundE (Ge et al., 2023)	.491	.450	.576	.357	.264	.545
ComplEX-DURA + RSCF (Ours)	<u>.497</u>	<u>.454</u>	.581	**.389**	**.296**	**.575**

Table 9: Test performance in broader approaches with different constraints based on embedding based KGC on FB15k-237 and WN18RR. Bold indicates the best result, and underline indicates the second best result.

Tensor Decomposition Model	WN18RR			FB15k-237			YAGO3-10		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
CP-DURA + SFBR(R)	.479	.441	.555	.368	**.275**	.557	.581	**.510**	.707
CP-DURA + SFBR (Liang et al., 2021)	**.485**	**.447**	**.561**	**.370**	.274	**.563**	**.582**	**.510**	**.711**
RESCAL-DURA + SFBR(R)	.497	.454	.576	**.369**	.277	.550	.578	.503	.712
RESCAL-DURA + SFBR (Liang et al., 2021)	**.500**	**.458**	**.581**	**.369**	**.276**	**.555**	**.581**	**.509**	**.712**
ComplEX-DURA + SFBR(R)	.491	.450	.571	.373	**.277**	.563	**.587**	**.517**	**.715**
ComplEX-DURA + SFBR (Liang et al., 2021)	**.498**	**.454**	**.584**	**.374**	**.277**	**.567**	.584	.512	.712

Table 10: Comparison of reproduced SFBR and SFBR reported in Liang et al. (2021)

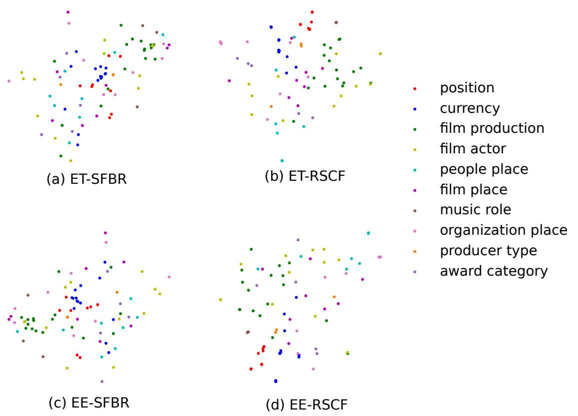


Figure 8: Tail entity-transformations and entity embeddings for semantically similar relation groups. (a) and (b) indicate ET of SFBR and RSCF, (c) and (d) indicate EE of SFBR and RSCF.

H Performance Comparison of RSCF with Other Knowledge Graph Embedding Models

Table 9 shows the comparison of the test performance of the RSCF and embedding based model on WN18RR and FB15k-237. ComplEX-DURA + RSCF outperforms all other models in FB15k-237 and shows competitive results in WN18RR.

Metric	ET-SFBR	EE-SFBR	ET-RSCF	EE-RSCF
Concentration Score (↑)	0.19	0.34	1.01	1.63
Inter Cluster Distance (↑)	0.46	0.46	0.69	0.52

Table 11: Concentration score and inter cluster distance of tail entity transformation and entity embedding of SFBR and RSCF.

I Reproduce of SFBR in TDM

We attempted to reproduce SFBR. During the reproduction process, we designed and executed the experiments based on the information provided in the SFBR and the publicly available datasets. Furthermore, in an attempt to clarify unclear aspects, we tried to communicate with the authors through multiple emails. However, the performance reported in the paper was not achieved. Table 10 shows reproduced SFBR and SFBR that is reported in Liang et al. (2021).

J Distribution of Tail Entity-Transformations and Corresponding Entity Embedding

Figure 8 presents the T-SNE visualization of tail ET and corresponding EE. Even in the tail, RSCF shows more in-cluster concentration. Also, in Table 11, RSCF exhibits higher concentration scores and inter class distance compared to SFBR.

Relation Group	Relations
position	/sports/sports_team/roster./basketball/basketball_roster_position/position /soccer/football_team/current_roster./soccer/football_roster_position/position /ice_hockey/hockey_team/current_roster./sports/sports_team_roster/position /sports/sports_team/roster./american_football/football_historical_roster_position/position_s /sports/sports_team/roster./baseball/baseball_roster_position/position /sports/sports_team/roster./american_football/football_roster_position/position /american_football/football_team/current_roster./sports/sports_team_roster/position /soccer/football_team/current_roster./sports/sports_team_roster/position
currency	/location/statistical_region/gdp_nominal_per_capita./measurement_unit/dated_money_value/currency /film/film/estimated_budget./measurement_unit/dated_money_value/currency /business/business_operation/operating_income./measurement_unit/dated_money_value/currency /organization/endowed_organization/endowment./measurement_unit/dated_money_value/currency /business/business_operation/revenue./measurement_unit/dated_money_value/currency /business/business_operation/assets./measurement_unit/dated_money_value/currency /location/statistical_region/rent50_2./measurement_unit/dated_money_value/currency /education/university/local_tuition./measurement_unit/dated_money_value/currency /location/statistical_region/gdp_real./measurement_unit/adjusted_money_value/adjustment_currency /education/university/domestic_tuition./measurement_unit/dated_money_value/currency /education/university/international_tuition./measurement_unit/dated_money_value/currency /location/statistical_region/gdp_nominal./measurement_unit/dated_money_value/currency /location/statistical_region/gni_per_capita_in_ppp_dollars./measurement_unit/dated_money_value/currency /base/schemastaging/person_extra/net_worth./measurement_unit/dated_money_value/currency
film production	/film/film/costume_design_by /film/film/executive_produced_by /award/award_winning_work/awards_won./award/award_honor/award_winner /tv/tv_program/program_creator /film/film/film_art_direction_by /film/film/music /film/film/film_production_design_by /film/film/other_crew./film/film_crew_gig/crewmember /film/film/produced_by /tv/tv_program/regular_cast./tv/regular_tv_appearance/actor /film/film/edited_by /film/film/written_by /film/film/personal_appearances./film/personal_film_appearance/person /film/film/story_by /film/film/cinematography /film/film/dubbing_performances./film/dubbing_performance/actor /film/film/production_companies
film actor	/award/award_nominee/award_nominations./award/award_nomination/nominated_for /tv/tv_network/programs./tv/tv_network_duration/program /film/special_film_performance_type/film_performance_type./film/performance/film /film/director/film /tv/tv_personality/tv_regular_appearances./tv/tv_regular_personal_appearance/program /film/film_set_designer/film_sets_designed /tv/tv_writer/tv_programs./tv/tv_program_writer_relationship/tv_program /film/actor/film./film/performance/film /tv/tv_producer/programs_produced./tv/tv_producer_term/program /media_common/netflix_genre/titles /film/film_distributor/films_distributed./film/film_film_distributor_relationship/film

	/film/film_subject/films
people place	/music/artist/origin /people/person/places_lived./people/place_lived/location /people/person/place_of_birth /government/politician/government_positions_held./government/government_position_held/jurisdiction_of_office /people/deceased_person/place_of_death /people/person/nationality /people/deceased_person/place_of_burial /people/person/spouse_s./people/marriage/location_of_ceremony
film place	/film/film/distributors./film/film_film_distributor_relationship/region /film/film/featured_film_locations /film/film/release_date_s./film/film_regional_release_date/film_release_region /film/film/release_date_s./film/film_regional_release_date/film_regional_debut_venue /film/film/country /film/film/runtime./film/film_cut/film_release_region /tv/tv_program/country_of_origin /film/film/film_festivals
music role	/music/group_member/membership./music/group_membership/role /music/artist/track_contributions./music/track_contribution/role /music/artist/contribution./music/recording_contribution/performance_role
organization place	/organization/organization/headquarters./location/mailing_address/state_province_region /organization/organization/place_founded /user/ktrueman/default_domain/international_organization/member_states /organization/organization/headquarters./location/mailing_address/country /people/marriage_union_type/unions_of_this_type./people/marriage/location_of_ceremony /base/schemastaging/organization_extra/phone_number./base/schemastaging/phone_sandbox/service_location /government/legislative_session/members./government/government_position_held/district_represented /organization/organization/headquarters./location/mailing_address/citytown
producer type	/tv/tv_producer/programs_produced./tv/tv_producer_term/producer_type /film/film/other_crew./film/film_crew_gig/film_crew_role /tv/tv_program/tv_producer./tv/tv_producer_term/producer_type
award category	/award/award_category/winners./award/award_honor/award_winner /award/award_category/winners./award/award_honor/ceremony /award/award_category/category_of /award/award_category/nominees./award/award_nomination/nominated_for /award/award_category/disciplines_or_subjects

Table 12: Clearly distinct relation groups that are selected from original TransE