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, the consistency of embedding differences before and after transformation remains unaddressed, risking the loss of valuable inductive bias inherent in the embeddings. This inconsistency stems from two problems. First, transformation representations are specified for relations in a disconnected manner, allowing dissimilar transformations and corresponding entity-embeddings for similar relations. Second, a generalized plug-in approach as a SFBR (Semantic Filter Based on Relations) disrupts this consistency through excessive concentration of entity embeddings under entity-based regularization, generating indistinguishable score distributions among relations. In this paper, we introduce a plug-in KGE method, *Relation-Semantics Consistent Filter* (RSCF), containing more consistent entity-transformation 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. To amplify the advantages of consistency that preserve semantics on embeddings, RSCF adds relation transformation and prediction modules for enhancing the semantics. In knowledge graph completion tasks with distance-based and tensor decomposition models, RSCF significantly outperforms state-of-the-art KGE methods, showing robustness across all relations and their frequencies.

1 Introduction

Knowledge graphs (KGs) play crucial roles in a wide area of machine learning and its applications (Zhang et al., 2022b; Zhou et al., 2022; Geng et al., 2022). However, KGs, even on a large scale, still suffer from incompleteness (Dong et al., 2014). This problem has been extensively studied as a task to predict missing entities, known 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., 2020a).

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 (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 (Ge et al., 2023).

ETMs, however, have a limit to learning useful inductive bias that could be obtained in semantically similar relations. For example, 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., 2020a), 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 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 Consis-*

tent Filter (RSCF), that contains more consistent entity transformation (ET). In details, ET incorporates three features: 1) shared affine transformation for consistency mapping of relations to entity-transformations, 2) rooted entity-transformation 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. To amplify the benefit of the consistency, RSCFs adds relation transformation (RT) and relation prediction (RP) (Chen et al., 2021), for inducing useful relation-specific semantics on embeddings.

Our contributions are as follows.

- We raise and clarify two problems in terms of *relation-semantics consistency* in learning useful inductive bias on embeddings.
- We propose a novel and significantly outperforming RSCF as a plug-in KGE method, which induces the consistency and effectively learns useful semantic representations.
- 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 ETs 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 ET 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 ETs and, therefore, overall similar entity embeddings. Two phenomena of losing this inductive bias and their causes are as follows.

Disconnection of Entity-Transformations Disconnected ET loosely use this bias, especially under lack of triplet data. In existing methods, relation-specific ETs use separate parameters such as $\mathsf{h}_r = \mathsf{W}_r\mathsf{h}$ and $\mathsf{t}_r = \mathsf{W}_r\mathsf{t}$, 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 KG introduces a wide variety of possible ETs 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.

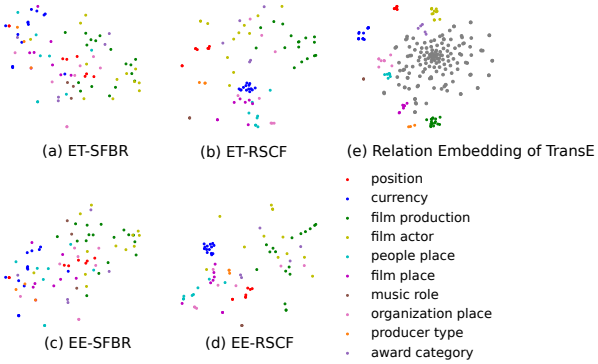


Figure 1: Head entity-transformations and entity embeddings for semantically similar relation groups. (a) and (b) indicate ET of SFBR and RSCF, (c) and (d) indicate EE of SFBR and RSCF. Points in the same color are relations in the same group. Clearly distinct groups are selected from the original TransE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

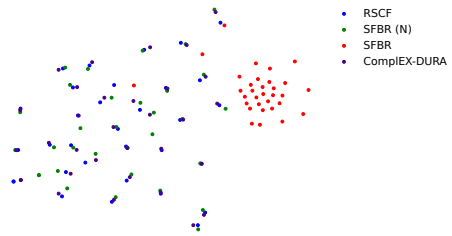
Metric	ET-SFBR	ET-RSCF	EE-SFBR	EE-RSCF
Concentration Score (↑)	0.91	2.15	0.43	1.91
Inter Cluster Distance (↑)	0.27	0.82	0.40	0.85

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.

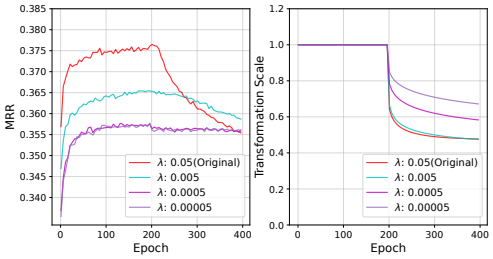
Empirical Evidence for Disconnection Figure 1 shows the impact of the disconnection via T-SNE visualization of ET and corresponding entity embeddings (EE) of TransE-SFBR (SFBR) and TransE-RSCF (RSCF). We split them into relation groups, defined by clearly clustered relation groups in the original TransE presented in Figure 1 (e). An entity for EE is randomly selected from FB15k-237. Compared to RSCF, the ET and EE distribution of SFBR are mostly dispersed between semantically different relation groups. This phenomenon is supported by concentration score and inter cluster distance in Table 1, which implies the limit of SFBR in inducing relation-semantics consistency. [1]

Entity Embedding Concentration In particular, SFBR additionally loses consistency under entity-

[1]For details about the relation group, concentration score and inter cluster distance, please refer to the Appendix A



(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).

based regularization, DURA (Zhang et al., 2020a). In KGE based on TDM, DURA has shown significant 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., 2020a) and DURA in SFBR (below).

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

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 B.1) and this causes indistinguishable score distribution in all score distributions.

Empirical Evidence for Concentration Figure 2 presents a T-SNE visualization of score distributions for selected queries. We selected the relation r_1, which shows significantly low performance in SFBR on FB15k-237, and selected all queries (h, r_1, ?) for this 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 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 ETMs, 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

$$\mathbf{e_r} = (\overbrace{\mathrm{N}_p}^{\text{ⓑ}}(\overbrace{\mathbf{rA_1}}^{\text{ⓐ}})\overbrace{+\mathbf{1}}^{\text{ⓒ}}) \otimes \mathbf{e} \quad (2)$$

where $\mathbf{A_1} \in \mathbf{R}^{n\times n}$ is shared affine transformation across all relations, $\mathbf{r}$ 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 A basic property of affine transformation is to maintain the parallelism of two parallel line segments

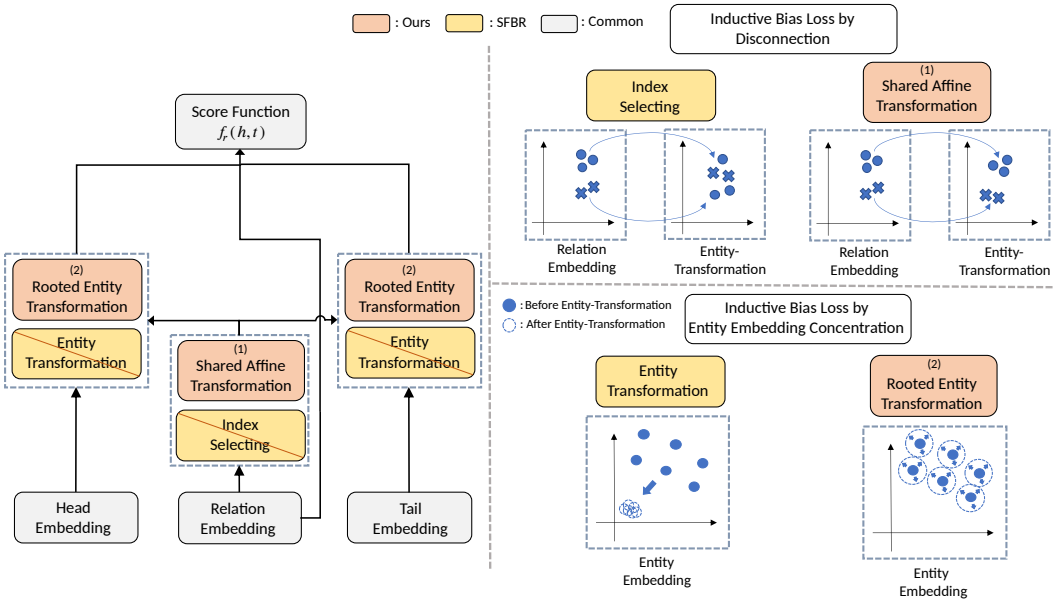


Figure 3: Overview of Relation-Semantics Consistent Filter and Its Effect. Its process (left) is illustrated on SFBR coloring changed modules. The two effects (right) are shown by comparing SFBR and RSCF on ET and entity embeddings.

Method	On a Line	$\frac{\lvert AC\rvert}{\lvert AB\rvert} > 1$	$\frac{\lvert AC\rvert}{\lvert AB\rvert} > 1.01$	$\frac{\lvert AC\rvert}{\lvert AB\rvert} > 1.02$
Transformation (ⓐ)	1.000	.728	.808	.875
Normalization (ⓑ)	.965	.869	.958	.994
Add one (ⓒ)	1.000	1.000	1.000	1.000

Table 2: Consistency of our ET: The numbers represent the success rates of preserving the superiority of distances for 10,000 randomly generated samples of three relation embeddings, notated as A, B, and C. (Line: the rates for samples with elements exactly on a line, the others: the rate for samples not on a line with varying distance rate conditions.)

after the transformation and preserves the ratio of their lengths. This property guarantees consistent mapping of relation embeddings at least on a line to generated vectors (part ⓐ in Equation 2). Even in the case that embeddings are not exactly on a line, the consistency is maintained with high probability as shown in Table 2. It presents the proportion of consistency maintenance rate for each component of RSCF based on Monte Carlo simulations. The result indicates that our ET can preserve consistency in the most cases. (Details about this result are presented in Appendix B.3).

After normalization of the generated vectors (part ⓑ), the consistency still holds in most cases, showing a maximum of over 99% in the table 2. The addition of one vector to the normalized change (part ⓒ) does not alter the inequality of distances, 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 across all relations inevitably reduces the expressiveness of ET compared to entirely separate relation-specific ET such as SFBR. 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. [2] 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 for More Consistent Relation Embedding to its Semantics The inductive bias introduced by the RSCF is dependent on the semantics of relation embeddings. Therefore, directly enhancing these semantics results in the improvement of RSCF performance. An effec-

[2] Refer to Appendix B.2 for details on the bounds of entity changes.

Model	Entity Transformation	Relation Transformation	Training Objective
PairRE	$\mathbf{e_r} = \mathbf{e} \otimes r^e$	✗	$\sum_p \phi(\mathbf{h_r}\|\mathbf{r}, \mathbf{t_r}) + \phi(\mathbf{t_r}\|\mathbf{h_r}, \mathbf{r})$
SFBR	$\mathbf{e_r} = \mathbf{W_r} \cdot \mathbf{e} + \mathbf{b}$	✗	$\sum_p \phi(\mathbf{h_r}\|\mathbf{r}, \mathbf{t_r}) + \phi(\mathbf{t_r}\|\mathbf{h_r}, \mathbf{r})$
CompoundE	$\mathbf{e_r} = \mathbf{T_r} \cdot \mathbf{R_r}(\theta) \cdot \mathbf{S_r} \cdot \mathbf{e}$	✗	$\sum_p \phi(\mathbf{h_r}\|\mathbf{r}, \mathbf{t_r}) + \phi(\mathbf{t_r}\|\mathbf{h_r}, \mathbf{r})$
RSCF	$\mathbf{e_r} = \psi(r) \otimes \mathbf{e}$	$\mathbf{r_{ht}} = \psi(h) \otimes \psi(t) \otimes \mathbf{r}$	$\sum_p \phi(\mathbf{h_r}\|\mathbf{r_{ht}}, \mathbf{t_r}) + \phi(\mathbf{t_r}\|\mathbf{h_r}, \mathbf{r_{ht}}) + \lambda\phi(\mathbf{r}\|\mathbf{h}, \mathbf{t})$ (Chen et al., 2021)

Table 3: Difference between RSCF and ETMs, where $\mathbf{e_r}$ is transformed entity embedding that contains both head entity $\mathbf{h_r}$ and tail entity $\mathbf{t_r}$. ψ is a rooted transformation that is presented in Equations 2, 4 and ϕ is a loss function with a score function that depends on the model. Complexity analysis is presented in Appendix C.5

tive approach is Relation Prediction (RP) (Chen et al., 2021) forming a cluster for semantically similar relations and improving discrimination of dissimilar relations. We add the training objective of RP (Chen et al., 2021) to RSCF as follows:

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

where ϕ is a loss function with a score function and λ is a hyper-parameter that controls the contribution of RP.

Relation Transformation for Relation Embedding of its Fine-Grained Semantics In KGs, some relations have various semantic meanings that can be divided into fine-grained sub-relations according to their semantics (Zhang et al., 2018). Because the semantic meanings of sub-relations are determined by their context, which is defined by head and tail entities (Jain and Krestel, 2022), we propose an entity-specific relation transformation (RT) to split relations into sub-relations, and apply the filter of ET of RSCF for the same purpose. By using Equation 2, we present the RT as follows:

$$\mathbf{r_{ht}} = (\mathrm{N}_p(\mathbf{h}\mathbf{A_2}) + 1) \otimes (\mathrm{N}_p(\mathbf{t}\mathbf{A_3}) + 1) \otimes \mathbf{r} \quad (4)$$

where $\mathbf{A_2} \in \mathbf{R}^{n \times n}$ and $\mathbf{A_3} \in \mathbf{R}^{n \times n}$ are shared affine transformation across all heads and tails. To predict score of given triplet (h, r, t), transformed entities $\mathbf{e_r}$ and relation $\mathbf{r_{hr}}$ are used. The difference of RSCF and ETMs and are summarized in Table 3.

4 Related Works

Knowledge Graph Embedding KGE encodes entities and relations into low-dimensional latent spaces to assess the validity of triples. TransE (Bordes et al., 2013) and RotatE (Sun et al., 2018) describe each relation as a translation and rotation between entities, respectively. DistMult (Yang et al., 2015) regards KGC as a tensor completion problem in euclidean space, and ComplEX (Trouillon et al., 2016) extends it to complex space. TransHRS (Zhang et al., 2018) improves knowledge representation by using the information from the HRS. DURA (Zhang et al., 2020a), AnKGE (Yao et al., 2023), and CompliE (Cui and Zhang, 2024) are methods that can be applied to KGE models to prevent overfitting, provide analogical inference and enable composition reasoning. VLP (Li et al., 2023) presents an explicit copy strategy to allow referring to related factual triples. GreenKGC (Wang et al., 2023) and SpeedE (Pavlović and Sallinger, 2024) propose low-dimensional embedding methods to handle large-scale KGs. WeightE (Zhang et al., 2023) utilizes a reweighting technique to alleviate the data imbalance issue. UniGE (Liu et al., 2024) introduce integration KGE in both euclidean and hyperbolic to capture various relational patterns. However, using only a single embedding for an entity or a relation can restrict the learning of complex relation patterns.

Entity Transformation Models 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. Recently, AutoETER (Niu et al., 2020) learns the type embedding for each entity with relation-specific transformation. 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. ReflectE (Zhang et al., 2022a) introduces relation-specific householder transformation to handle sophisticated relation mapping properties. CIBLE (Cui and Chen, 2022) use relation-aware-transformation for prototype modeling to represent the knowledge graph. CompoundE (Ge et al., 2023) applied compound operation to both head and tail

Knowledge Graph Embedding	WN18RR			FB15k-237			YAGO3-10		
	MRR	H@1	H@10	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	-	-	-
DistMult-HRS (Zhang et al., 2018)	-	-	-	.315	.241	.496	-	-	-
AutoETER (Niu et al., 2020)	-	-	-	.344	.250	.538	.550	.465	.699
ComplEX-DURA (Zhang et al., 2020a)	.491	.449	.571	.371	.276	.560	.584	.511	.713
PairRE (Chao et al., 2021)	-	-	-	.351	.256	.544	-	-	-
CIBLE (Cui and Chen, 2022)	.490	.446	.575	.341	.246	.532	-	-	-
ReflectE (Zhang et al., 2022a)	.488	.450	.559	.358	.263	.546	-	-	-
HAKE-AnKGE (Yao et al., 2023)	.500	.454	.587	.385	.288	.572	-	-	-
CompoundE (Ge et al., 2023)	.491	.450	.576	.357	.264	.545	-	-	-
RotatE-GreenKGC (Wang et al., 2023)	.411	.367	.491	.345	.265	.507	.453	.361	.629
RotatE-VLP (Li et al., 2023)	.498	.455	.582	.362	.271	.542	-	-	-
RotatE-WeightE (Zhang et al., 2023)	.501	.448	**.592**	.371	.281	.557	.580	.504	.713
CompliE-DURA (Cui and Zhang, 2024)	.495	.453	.579	.372	.277	.563	-	-	-
SpeedE (Pavlović and Sallinger, 2024)	.493	.446	-	.320	.227	-	.413	.332	-
UniGE (Liu et al., 2024)	.502	.455	**.592**	.357	.264	.559	.583	.512	.715
TransE-SFBR (Liang et al., 2021)	.242	.028	.548	.338	.240	.538	-	-	-
ComplEX-DURA-SFBR (Liang et al., 2021)	.498	.454	.584	.374	.277	.567	.584	.512	.712
TransE-RSCF (Ours)	.267 ±.001	.066 ±.002	.546 ±.002	.363 ±.001	.264 ±.001	.558 ±.001	-	-	-
ComplEX-DURA-RSCF (Ours)	**.503** ±.001	**.460** ±.001	.588 ±.002	**.388** ±.001	**.295** ±.002	**.573** ±.004	**.589** ±.002	**.516** ±.003	**.718** ±.002

Table 4: Test performance of KGE-based KGC on FB15k-237, WN18RR and YAGO3-10. Bold indicates the best result, and underlined signifies the second best result. ± indicates standard deviation. (The comparison results of RSCF, SFBR, and other KGE models presented in Appendix C.2.)

entities. However, these models have no chance for inductive bias sharing due to the separate parameter of ET, and SFBR (Liang et al., 2021), which can be applied to both DBM and TDM, suffers from indistinguishable score distribution because of the entity embedding concentrations.

5 Experiments

5.1 Settings

Dataset To evaluate our proposed RSCF models, we consider three KGs datasets: WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova and Chen, 2015), and YAGO3-10 (Mahdisoltani et al., 2013). The statistics for the three benchmarks are shown in Appendix C.1.

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, except for the predicted triple. 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 KGE models: TransE, DistMult, ComplEX, RotatE, DistMult-HRS, AutoETER, ComplEX-DURA, PairRE, SFBR, CIBLE, ReflectE, HAKE-AnKGE, CompoundE, RotatE-GreenKGC, RotatE-VLP, RotatE-WeightE, CompliE-DURA, SpeedE, and UniGE.

Because RSCF is a module that is plugged in based on existing models, we use DBM, including TransE, RotatE, and TDM, including CP, RESCAL, and ComplEX as base models. Additionally, following the setting of SFBR, ET is applied to both head and tail entities in DBM, while it is applied only to the head entity in TDM due to computational cost (Liang et al., 2021). For the same reason, both head and tail entities are utilized for RT in DBM, while only the head entity is used in TDM. In the FB15k-237, the entity/relation ratio and the triple/relation ratio are significantly lower than in the other two datasets, limiting the context information available to each relation. This limitation is particularly critical in TDM, which relies solely on the head entity. Therefore, RT is not applied to TDM in the FB15k-237 dataset.

5.2 Performance

Performance on KGC Table 4 shows the performance comparison of the RSCF and other KGE models on WN18RR, FB15k-237 and YAGO3-10. Overall, RSCF shows higher or competitive performance compared to other KGE models. Especially in FB15k-237 and YAGO3-10, RSCF outperforms other state-of-the-art models that include HAKE-AnKGE and CompliE-DURA.

Ablation Study Table 6 presents ablation studies of RSCF to verify the effectiveness of each com-

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**)	**5** / 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**)	**3** / 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**)	**10** / 21

Table 5: 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. TransE is used as baseline

Model	ET		RP	RT	T	C
	ⓐ	ⓑ & ⓒ			MRR	MRR
RSCF	✓	✓	✓	✓	.363	.387
+ w/o ET	✗	✗	✓	✓	.356	.385
+ w/o RP	✓	✓	✗	✓	.358	.374
+ w/o RT	✓	✓	✓	✗	.356	.388
+ w/o RP	✓	✓	✗	✗	.349	.375
+ w/o ET	✗	✗	✓	✗	.338	.385
+ w/o ⓐ	✗	✓	✓	✗	.354	.386
+ w/o ⓑ & ⓒ	✓	✗	✓	✗	.353	.377

Table 6: Results of an ablation study of RSCF on FB15k-237. TransE and ComplEX are used as base models. **T** and **C** indicate the base models of RSCF, which denote TransE and ComplEX, respectively. MRR is used for performance comparison.

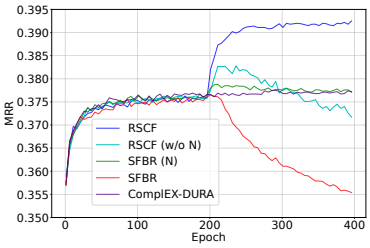


Figure 4: MRR changes over epochs of RSCF, RSCF (w/o N), SFBR (N), SFBR, and ComplEX on FB15k-237.

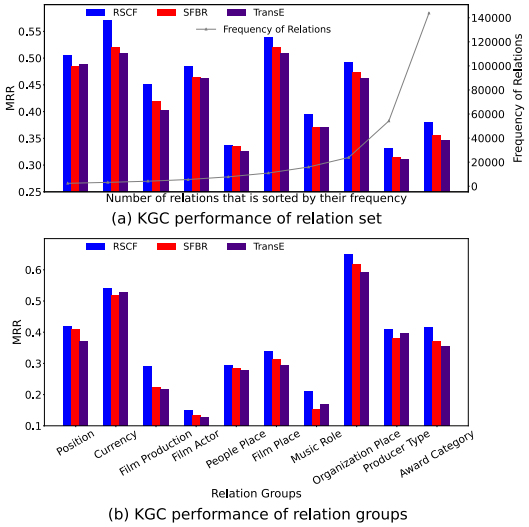


Figure 5: KGC performance of the relation set that is sorted by their frequency (above) and groups of semantically similar relations observed in Figure 1 (e) (below) on FB15k-237

ponent. ⓐ and ⓑ & ⓒ are the components of ET described in Equation 2. In the ablation study, ⓑ & ⓒ are combined because ⓑ & ⓒ should be used simultaneously to maintain the original scale. In Table 6, RSCF shows higher performance compared to the other ablated models in both TransE and ComplEX, suggesting 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] In addition, please note that RT can reduce the performance of ComplEX-RSCF on FB15k-237 because of context information restriction.

[3]Detailed description of SFBR (N) is presented in Appendix C.3.

Performance on Relation Frequency and Semantically Distinguished Relation Groups To demonstrate the generality of applying RSCF regardless of relation frequency, we sort relations by their frequency and divide them into ten sets. Each set has an equal number of relations. Figure 5 above shows the MRR for each set in TransE, RSCF, and SFBR. The results shows that RSCF outperform 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.

Figure 5 below shows the MRR for each relation group as defined in Figure 1 (e). RSCF outperformed SFBR and TransE in all groups, demonstrating that RSCF can be utilized without specific bias to the semantics of relations and that incorporating relation semantics into the transformation function can improve model performance.

Qualitative Example Analysis For qualitative analysis, Table 5 presents sampled queries, their correct answers, related triples with the sample

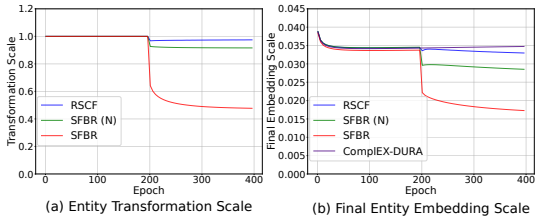


Figure 6: Entity transformation scale (left) and final entity embedding scale (right) of RSCF, SFBR (N), SFBR, and ComplEX-DURA over epochs on FB15k-237. DURA is applied in all epochs and RSCF and SFBR are applied after 200 epochs.

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 5, 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.

Recovery of Embedding Scale and Score Distribution Figure 6 presents transformation scale and final entity embedding scale over epochs on FB15k-237, using ComplEX as the baseline. Following the approach of SFBR, DURA is applied in all epochs, and RSCF and SFBR are plugged in after 200 epochs. The results show that SFBR decreases both transformation scale and final entity embedding scale. In contrast, RSCF and SFBR (N) maintain scales, indicating that normalization helps preserve the embedding scale due to normalization. As shown in Figure 4, MRR decreases for SFBR and RSCF w/o normalization but increases for both RSCF and SFBR (N), implying that entity embedding concentration negatively impacts model performance.

To investigate the detailed change of score distribution, we present the score distribution of randomly sampled queries from Figure 2 (a) in Figure 7. SFBR shows near-zero scores for most entities, with significantly similar distributions across queries. However, by applying normalization or using RSCF, the diversity of scores is recovered as the original base model.

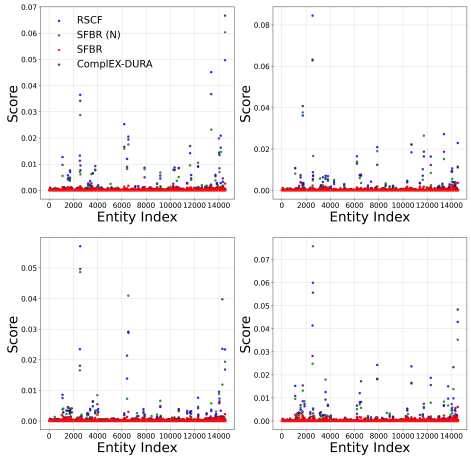


Figure 7: Score distribution of all entities for randomly selected queries from Figure 2 (a)

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

Table 7: KGC performance of all queries associated with the relation that shows strong concentration of entity embedding in SFBR. Concentration presents entity embedding concentration.

Performance Decrease by Entity Embedding Concentration To assess the impact of indistinguishable score distribution, we conducted a performance evaluation for the selected relation that shows critical entity embedding concentration in Figure 2. Table 7 presents the MRR for all queries associated with the selected relation. SFBR shows significantly lower performance than RSCF, SFBR (N), and the ComplEX. This result implies that indistinguishable score distribution strongly affects the prediction of SFBR, and simply applying normalization can recover it.

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

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. Multi-relational poincaré graph embeddings. *Advances in Neural Information Processing Systems*, 32.

Ivana Balažević, Carl Allen, and Timothy Hospedales. 2019. Tucker: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5185–5194.

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.

Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, and Qingming Huang. 2021. Dual quaternion knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6894–6902.

Zongsheng Cao, Qianqian Xu, Zhiyong Yang, and Qingming Huang. 2022. Er: equivariance regularizer for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5512–5520.

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*.

Wanyun Cui and Xingran Chen. 2022. Instance-based learning for knowledge base completion. *Advances in Neural Information Processing Systems*, 35:30744–30755.

Wanyun Cui and Linqiu Zhang. 2024. Modeling knowledge graphs with composite reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8338–8345.

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.

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.

Nitisha Jain and Ralf Krestel. 2022. Discovering fine-grained semantics in knowledge graph relations. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 822–831.

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.

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.

Rui Li, Xu Chen, Chaozhuo Li, Yanming Shen, Jianan Zhao, Yujing Wang, Weihao Han, Hao Sun, Weiwei Deng, Qi Zhang, et al. 2023. To copy rather than memorize: A vertical learning paradigm for knowledge graph completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6335–6347.

Yizhi Li, Wei Fan, Chao Liu, Chenghua Lin, and Jiang Qian. 2022. Transher: Translating knowledge graph

embedding with hyper-ellipsoidal restriction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8517–8528.

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.

Yuhan Liu, Zelin Cao, Xing Gao, Ji Zhang, and Rui Yan. 2024. Bridging the space gap: Unifying geometry knowledge graph embedding with optimal transport. In *Proceedings of the ACM on Web Conference 2024*, pages 2128–2137.

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.

Guanglin Niu, Bo Li, Yongfei Zhang, Shiliang Pu, and Jingyang Li. 2020. Autoeter: Automated entity type representation for knowledge graph embedding. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1172–1181.

Aleksandar Pavlović and Emanuel Sallinger. 2024. Speede: Euclidean geometric knowledge graph embedding strikes back. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 69–92.

Tengwei Song, Jie Luo, and Lei Huang. 2021. Rotpro: Modeling transitivity by projection in knowledge graph embedding. *Advances in Neural Information Processing Systems*, 34:24695–24706.

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.

Yun Cheng Wang, Xiou Ge, Bin Wang, and C-C Jay Kuo. 2023. Greenkgc: A lightweight knowledge graph completion method. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10596–10613.

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.

Zhen Yao, Wen Zhang, Mingyang Chen, Yufeng Huang, Yi Yang, and Huajun Chen. 2023. Analogical inference enhanced knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4801–4808.

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

Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. 2019. Quaternion knowledge graph embeddings. *Advances in neural information processing systems*, 32.

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.

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

Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020b. Learning hierarchy-aware knowledge graph embeddings for link prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3065–3072.

Zhao Zhang, Zhanpeng Guan, Fuwei Zhang, Fuzhen Zhuang, Zhulin An, Fei Wang, and Yongjun Xu. 2023. Weighted knowledge graph embedding. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, pages 867–877.

Zhao Zhang, Fuzhen Zhuang, Meng Qu, Fen Lin, and Qing He. 2018. Knowledge graph embedding with hierarchical relation structure. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3198–3207.

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 Appendix A

A.1 Relation Groups for Entity Transformation

Figure 8 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 14. Note that similar relations belong to the same group.

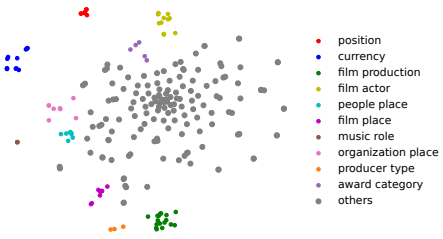


Figure 8: Visualization of relation embeddings of TransE using T-SNE

A.2 Distribution of Tail Entity-Transformations and Corresponding Entity Embedding

Figure 9 presents the T-SNE visualization of tail ET and corresponding EE of RSCF and SFBR. Even in the tail, RSCF shows more concentrated clusters. Also, in Table 8, RSCF exhibits higher concentration scores and inter class distance compared to SFBR.

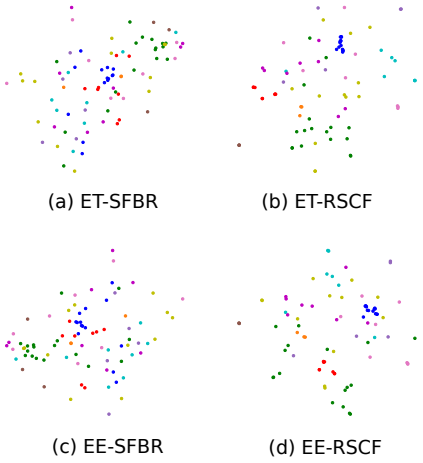


Figure 9: Tail entity-transformations and entity embeddings for semantically similar relation groups. (a) and (b) indicate ET of SFBR and RSCF, (c) and (d) indicate EE of SFBR and RSCF.

Metric	ET-SFBR	ET-RSCF	EE-SFBR	EE-RSCF
Concentration Score (↑)	0.19	1.90	0.34	1.50
Inter Cluster Distance (↑)	0.46	0.70	0.46	0.74

Table 8: Concentration score and inter cluster distance of tail entity transformation and entity embedding of SFBR and RSCF.

A.3 Measurement of Cluster Concentration

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

$$\sum_{k}^{n}\sum_{i}^{m}\frac{||(k_i - C_k)||}{n||C_k||} \tag{5}$$

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:

$$C_k = \frac{\sum_i^m k_i}{m} \tag{6}$$

In the equation 5, the vector norm of C ($||C||$) is used because of the relative concentration score

for clusters. We use the reciprocal of equation 5 as concentration score. Also to evaluate the distance between different clusters we defined inter cluster distance score as follows:

$$\sum_{k}^{n} \frac{||(C_k - C_{kc})||}{\sum_{i}^{m} ||k_i||} \tag{7}$$

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

$$C_{kc} = \min_{k \neq j} \{||C_k - C_j||\} \tag{8}$$

In Equation 7, 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.

B Appendix B

B.1 Proof of scale decrease of ET

Let $\mathsf{W}_{\mathsf{r}_j}$ is the ET of SFBR and $w_{r_j,n}$ is n-th element of $\mathsf{W}_{\mathsf{r}_j}$, than the gradient of $w_{r_j,n}$ in DURA can be calculated as:

$$\begin{aligned} &\sum_{p} \frac{dL}{dw_{r_j,n}} ||\mathsf{w}_{\mathsf{r}_{j,n}} \mathsf{h}_{\mathsf{i},\mathsf{n}} \mathsf{r}_{\mathsf{j},\mathsf{n}}||_2^2 + ||\mathsf{w}_{r_j n} \mathsf{h}_{\mathsf{i},\mathsf{n}}||_2^2 \\ =&\sum_{p} \frac{dL}{dw_{r_j,n}} \mathsf{w}_{\mathsf{r}_{j,n}}^2 (\mathsf{h}_{\mathsf{i},\mathsf{n}} \mathsf{r}_{\mathsf{j},\mathsf{n}})^2 + \mathsf{w}_{r_j n}^2 \mathsf{h}_{\mathsf{i},\mathsf{n}}{}^2 \\ =&\sum_{p} 2\mathsf{w}_{\mathsf{r}_{j,n}} (\mathsf{h}_{\mathsf{i},\mathsf{n}} \mathsf{r}_{\mathsf{j},\mathsf{n}})^2 + 2\mathsf{w}_{r_j n} \mathsf{h}_{\mathsf{i},\mathsf{n}}{}^2 \end{aligned} \tag{9}$$

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.

B.2 Normalization of Change for Reducing Entity Embedding Concentration

The change generated 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

$$\|\alpha \otimes \mathbf{e}\|_p \tag{10}$$

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

$$\|\frac{\mathbf{e}}{\|\mathbf{e}\|_p} \otimes \mathbf{e}\|_p = \|\frac{\mathbf{e}^2}{\|\mathbf{e}\|_p}\|_p = \frac{\|\mathbf{e}^2\|_p}{\|\mathbf{e}\|_p} \tag{11}$$

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.

B.3 Empirical Experiments on Maintaining Consistency through Monte Carlo Simulation

Linear transformation ensures consistency when relation embeddings exist on a line. Furthermore, for relations that do not lie on the line, we can predict that that similar relations will have similar ETs because of the continuous property of linear transformation, i.e. if $r_1 \approx r_2$ then $r_1 A \approx r_2 A$. However, there has been no research on the proportion of these relations that maintain consistency after transformation and after normalization, and it is extremely challenging to determine this proportion through mathematical formulations. Therefore, we conducted an empirical analysis using Monte Carlo simulations to investigate the consistency of points that do not lie on the line. Table 2 presents the proportion of consistency maintained under various conditions based on Monte Carlo simulations. For the experiment, we divided the scenarios into four cases: (1) when three randomly generated points (A, B, C) lie on the same line and the distance between A and C is greater than the distance between A and B (Line), (2) when the three points (A, B, C) do not lie on the line and the distance between A and C is greater than the distance between A and B($\frac{|\overline{AC}|}{|\overline{AB}|} > 1$), (3) when the distance between A and C is at least 1.01 times greater than the distance between A and B ($\frac{|\overline{AC}|}{|\overline{AB}|} > 1.01$), and (4) when the distance between A and C is at least 1.02 times greater than the distance between A and B ($\frac{|\overline{AC}|}{|\overline{AB}|} > 1.02$) and sampling was performed 10,000 times for each condition. Under each condition, we measured the proportion of cases where ($|\overline{AC}| > |\overline{AB}|$) was maintained even after Transformation (ⓐ), Normalization (ⓑ), and the Add one (ⓒ). The results showed that consistency was preserved in most cases, even when the three points did not lie on the same line. Specifically, in condition where ($\frac{|\overline{AC}|}{|\overline{AB}|} > 1$), over 72.8% of the samples maintained consistency in Transformation (ⓐ).

Knowledge Graph Embedding	WN18RR			FB15k-237			YAGO3-10		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
TuckER (Balažević et al., 2019)	.470	.443	.526	.358	.266	.544	-	-	-
QuatE (Zhang et al., 2019)	.488	.438	.582	.348	.248	.550	-	-	-
MuRP (Balazevic et al., 2019)	.481	.440	.566	.335	.243	.518	-	-	-
HAKE (Zhang et al., 2020b)	.497	.452	.582	.346	.250	.542	.545	.462	.694
RoTH (Chami et al., 2020)	.496	.449	.586	.344	.246	.535	.570	.495	.706
DualE (Cao et al., 2021)	.492	.444	.584	.365	.268	.559	-	-	-
FieldE (Nayyeri et al., 2021)	.48	.44	.57	.36	.27	.55	.51	.41	.68
Rot-Pro (Song et al., 2021)	.457	.397	.577	.344	.246	.540	.542	.443	.699
HAKE-CAKE (Niu et al., 2022)	-	-	-	.321	.226	.515	-	-	-
GIE (Yang et al., 2022)	.491	.452	.575	.362	.271	.552	.579	.505	.709
ComplEX-ER (Cao et al., 2022)	.494	.453	.575	.374	.282	.563	.588	.515	**.718**
TranSHER (Li et al., 2022)	-	-	-	.360	.264	.551	-	-	-
STaR-DURA (Li and Yang, 2022)	.497	.452	.583	.368	.273	.557	.585	.513	.713
ComplEX-DURA-RSCF (Ours)	**.503** ±.001	**.460** ±.001	**.588** ±.002	**.388** ±.001	**.295** ±.002	**.573** ±.004	**.589** ±.002	**.516** ±.003	**.718** ±.002

Table 9: Test performance of KGE-based KGC on FB15k-237, WN18RR and YAGO3-10. Bold indicates the best result, and underlined signifies the second best result.

Dataset	Entities	Relations	Entities/Relations	Triples/Relations	Triples		
					Train	Valid	Test
WN18RR	40,943	11	3,722	7,894	86,835	3,034	3,134
FB15k-237	14,541	237	61	1,148	272,115	17,535	20,466
YAGO3-10	123,182	37	3,329	29,163	1,079,040	5,000	5,000

Table 10: Statistics of KGC Benchmark Datasets

Furthermore, in cases where ($\frac{|\overline{AC}|}{|\overline{AB}|} > 1.02$) approximately 87.5% of the samples maintained consistency after Transformation (ⓐ). Also in Normalization (ⓐ), over 86.9% of the samples maintained consistency in ($\frac{|\overline{AC}|}{|\overline{AB}|} > 1$) and 99.4% in ($\frac{|\overline{AC}|}{|\overline{AB}|} > 1.02$). Considering that relations tend to form clusters based on their semantics because of score function and RP (Chen et al., 2021), it is difficult that these conditions are unrealistic, indicating that our method performs robustly across various conditions.

C Appendix C

C.1 Datasets

We evaluate the RSCF using three widely-used datasets: WN18RR, FB15k-237, and YAGO3-10. WN18RR, FB15k-237 and YAGO3-10 are subsets of WN18 (Bordes et al., 2013), FB15k (Bordes et al., 2013), and YAGO3 (Mahdisoltani et al., 2013), respectively, designed to alleviate the test set leakage problem. Statistics of these datasets are shown in Table 10.

C.2 Performance Comparison of RSCF, SFBR, and Other KGE models

Table 9 shows the performance comparison of the RSCF and previous KGE-based models on WN18RR, FB15k-237 and YAGO3-10. Overall, ComplEX-DURA+RSCF shows higher performance than other KGE models in all settings, demonstrating that the effectiveness of the RSCF for the KGC task.

Table 11 shows the performance comparison of the DBM-RSCF and DBM-SFBR on WN18RR and FB15k-237. Overall, DBM-RSCF shows similar or higher performance than DBM-SFBR in most settings. Table 12 shows the performance comparison in TDMs. Compared to TDM-SFBR, TDM-RSCF shows consistent performance improvements in all datasets and settings.

C.3 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:

$$\mathbf{N}_p(\mathbf{W_r}) + 1 \tag{12}$$

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

$$\mathbf{e_r} = (\mathbf{N}_p(\mathbf{W_r}) + 1)\mathbf{e} \tag{13}$$

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

C.4 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-

Distance-Based Model with Entity Transformation	WN18RR			FB15k-237		
	MRR	H@1	H@10	MRR	H@1	H@10
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	.267 ±.001	.063 ±.002	.550 ±.002	**.364** ±.001	.266 ±.001	**.560** ±.001
TransE-RSCF (Linear-2)	.343 ±.009	.232 ±.014	.499 ±.003	.359 ±.001	.262 ±.001	.552 ±.001
RotatE-RSCF	.493 ±.001	.447 ±.001	.584 ±.001	.363 ±.000	**.268** ±.001	.556 ±.001
RotatE-RSCF (Linear-2)	**.495** ±.001	**.452** ±.001	.578 ±.001	**.364** ±.000	**.268** ±.000	.556 ±.001

Table 11: Test performance of DBM-based RSCF and SFBR on FB15k-237 and WN18RR. Bold indicates the best result, and underlined signifies the second best result.

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
CP-DURA + SFBR (Liang et al., 2021)	.485	.447	.561	.370	.274	.563	.582	.510	.711
RESCAL-DURA + SFBR (Liang et al., 2021)	.500	.458	.581	.369	.276	.555	.581	.509	.712
ComplEX-DURA + SFBR (Liang et al., 2021)	.498	.454	.584	.374	.277	.567	.584	.512	.712
CP-DURA + RSCF	.486 ±.001	.447 ±.001	.561 ±.001	.379 ±.000	.287 ±.000	.565 ±.001	.585 ±.000	.514 ±.001	.711 ±.001
RESCAL-DURA + RSCF	**.507** ±.000	**.467** ±.000	.581 ±.000	.381 ±.000	.289 ±.000	.562 ±.000	.584 ±.000	.511 ±.000	.716 ±.000
ComplEX-DURA + RSCF	.503 ±.001	.460 ±.001	**.588** ±.002	**.389** ±.001	**.296** ±.002	**.575** ±.004	**.590** ±.002	**.518** ±.003	**.719** ±.002

Table 12: Test performance of TDM-based RSCF and SFBR on FB15k-237, WN18RR, and YAGO3-10. Bold indicates the best result, and underlined signifies the second best result.

2) can be written as:

$$\mathbf{W_r}^{\text{Linear-2}} = \begin{bmatrix} diag(\mathbf{w}_1) & diag(\mathbf{w}_2) \\ diag(\mathbf{w}_3) & diag(\mathbf{w}_4) \end{bmatrix} \quad (14)$$

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 $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4 \in \mathbf{R}^{n/2}$.

C.5 Complexity Analysis

Table 13 presents the complexity comparison between RSCF and other ETMs. In general, because the number of parameters in KGE methods is significantly influenced by the number of entities and relations, the parameter difference between RSCF and ETMs is marginal. Additionally, although RSCF requires more training time compared to other models, considering that the proposed KGE models assume an offline learning setting and have similar inference times to RSCF, this is not a significant drawback.

C.6 Implementation Details

When training the RSCF, we followed the experimental settings described in the SFBR (Liang et al., 2021). Following the settings of SFBR, RSCF and RSCF (Linear-2) are applied to both head and tail entities in DBM, and RSCF is applied to only the head entity in TDM due to computational costs (Liang et al., 2021). For the same reason, both head and tail entities are used for RT in DBM, whereas only the head entity is used in TDM. In FB15k-237, the entity/relation ratio and the train/relation ratio are significantly lower compared to the other two datasets, which restricts the context information that each relation can obtain. This restriction is more critical in TDM, which uses only the head entity, and thus RT is not applied to TDM on FB15k-237. 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. (2020a). The presented results of RSCF represent mean of the three runs for each model.] Experiments for the DBM were conducted on an NVIDIA 3090 with 24GB of memory, while experiments for the TDM were conducted on an NVIDIA 2080TI with 11GB and an A100 with 40GB of memory was used for both DBM and TDM.

D Appendix D

D.1 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

Model	Training Time		Inference Time		# Params	
	WN18RR	FB15k-237	WN18RR	FB15k-237	WN18RR	FB15k-237
TransE	45m	1h	30s	1m 20s	20.48M	14.78M
PairRE	-	3h	30s	1m 20s	-	22.52M
T-SFBR	1h	1h 15m	30s	1m 30s	20.49M	15.25M
CompoundE	2h 40m	2h 20m	30s	1m 20s	20.5M	9.58M
T-RSCF	3h 10m	5h 30m	30s	1m 50s	21.48M	18.78M

Table 13: Training time, inference time and number of parameters of RSCF and ETMs, T-SFBR (Diag) and T-RSCF indicate TransE-SFBR and TransE-RSCF, respectively. Inference Time denotes inference time on test set. Both training and inference time are measured on RTX3090.

TransE-RSCF can be expressed as:

$$d_r(\mathbf{h}, \mathbf{r}) = \|\mathbf{h_r} + \mathbf{r_{ht}} - \mathbf{t_r}\| \tag{15}$$

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

$$d_r(\mathbf{h}, \mathbf{r}) = \|\mathbf{h_r} \circ \mathbf{r_{ht}} - \mathbf{t_r}\| \tag{16}$$

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

$$d_r(\mathbf{h}, \mathbf{r}) = \|\mathbf{h_r r_h t}\| \tag{17}$$

In TDM, tail embeddings are not transformed according to the settings of SFBR in order to reduce computational costs. For the same reason, only the head entity is used for relation transformation.

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 14: Clearly distinct relation groups that are selected from original TransE