Modeling Knowledge Graphs with Composite Reasoning

Wanyun Cui, Linqiu Zhang

Shanghai University of Finance and Economics
cui.wanyun@sufe.edu.cn, zhang.linqiu@stu.sufe.edu.cn

Abstract

The ability to combine multiple pieces of existing knowledge to infer new knowledge is both crucial and challenging. In this paper, we explore how facts of various entities are combined in the context of knowledge graph completion (KGC). We use composite reasoning to unify the views from different KGC models, including translational models, tensor factorization (TF)-based models, instance-based learning models, and KGC regularizers.

Moreover, our comprehensive examination of composite reasoning revealed an unexpected phenomenon: certain TF-based models learn embeddings with erroneous composite reasoning, which ultimately violates their fundamental collaborative filtering assumption and reduces their effects. This motivates us to reduce their composition error. Empirical evaluations demonstrate that mitigating the composition risk not only enhances the performance of TF-based models across all tested settings, but also surpass or is competitive with the state-of-the-art performance on two out of four benchmarks. Our code, data and supplementary material are available at https://github.com/zlq147/CompilE

1 Introduction

Diverse paradigms have been developed for knowledge graph modeling, including translation models (Bordes et al. 2013; Sun et al. 2019; Zhang et al. 2020; Lin et al. 2015), tensor factorization models (Hitchcock 1927; Trouillon et al. 2016; Yang et al. 2015), instance-based learning (Cui and Chen 2022), and KGC regularizers (Zhang, Cai, and Wang 2020). Given the diversity of different KGC forms, it is crucial to provide a unified understanding for them. Firstly, this aids in a deeper understanding of the principles and application domains of each method. Secondly, it motivates new algorithmic innovations.

To this end, we propose a novel paradigm for representing knowledge graphs: composite reasoning. Our motivation for adopting composite reasoning in knowledge graph modeling is straightforward. We aim to leverage the known facts about other entities to predict the target entity. For example, consider a knowledge graph with composition $\mathrm{Alphabet} = \mathrm{Google} + \mathrm{DeepMind} + \cdots$. If we know the fact $\mathrm{Google}, \mathrm{employee}, \mathrm{Jeff\ Dean})$, we can infer $(\mathrm{Alphabet}, \mathrm{employee}, ?) = \mathrm{Jeff\ Dean}$.

The composite reasoning unifies several existing paradigms for knowledge graph modeling, such as translation models, tensor factorization models, and instance-based learning models; as well as knowledge graph regularization methods like DURA (Zhang, Cai, and Wang 2020). We show how composite reasoning works in Fig. 1. The results provides novel insights to interpreting and comparing different KGC models.

Through a comparative analysis of different KGC models from the viewpoint of composite reasoning, we have discovered *an anomalous characteristic* of tensor factorization (TF) models: a query can be decomposed into several entities that are completely unrelated to the query entity. (see Fig. 1 and Table 1) This finding unveils a fundamental issue with traditional factorization-based approaches, namely, the learned embeddings may violate the collaborative filtering assumption due to erroneous knowledge composition. More details of the comparison can be found in Sec 3.5.

To address the erroneous knowledge composition problem in TF-based models, we propose a measure to mitigate and reduce the caused generalization risk. In this paper, we refer to this risk as composite risk. Measuring and reducing the composite risk pose challenges as obtaining ground truth for knowledge composition is hard. One of our key observation is that we can relax the definition of low-risk entities to *neighbor* entities, thereby obtaining a lower bound for the composite risk. Our experiments demonstrate a strong correlation between prediction quality and the approximated composition risk (see Sec 4.4).

Comparison with other KGC explanations The embedding spaces of many existing KGC models are designed according to how humans explain knowledge. For example, translational models usually explicitly represent inverse/symmetric/transitive relations via embedding translations. Tensor factorization-based models conform to the low-rank assumption of real-world knowledge. However, these explanations are usually only from an intra-triple perspective, i.e. explaining a single triplet fact. The composite reasoning-based explanation provides a novel inter-triplet view to explain the interactions among different facts.

The main contribution of this paper includes: (1) We propose a novel composite reasoning perspective to unify dif-

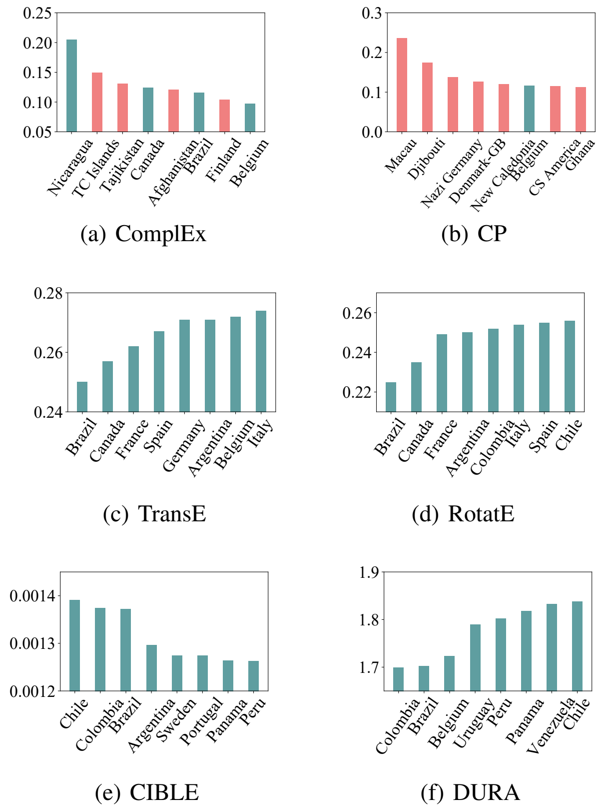


Figure 1: How composite reasoning unifies and interprets different KGC models. For each model, we show top 8 entities from the perspective of composite reasoning for query $(\mathrm{Mexico}, \text{official language}, ?)$ in FB15k-237. For TransE, RotatE, and DURA, less α_i indicates higher composite dependency. For other models, higher α_i indicates higher composite dependency. For ComplEx and CP, entities that are intuitively unrelated for humans are marked in red.

ferent KGC models. (2) We compare different modeling approaches under the framework of composite reasoning and uncover the anomalous knowledge composition in TF-based models. (3) We quantify how errors in tensor factorization model's decomposition affect its generalization capability. We optimize the TF models by approximating and reducing the composite risk.

2 The Composite Reasoning Framework for Knowledge Graph Completion

In this section, we present the formulation of the KGC problem and demonstrate how it can be represented within the composite reasoning framework.

ComplEx	CP	TransE	RotatE	CIBLE	DURA
0.073 ↓	0.084 ↓	0.197	0.207	0.211	0.191

Table 1: Composite rationality of different models.

Knowledge Graph Completion: A knowledge graph is a collection of facts represented as triples in the form of *(head, relation, tail)*, denoted as $\mathrm{KG} = (\mathrm{h_i}, \mathrm{r_i}, \mathrm{t_i})_{\mathrm{i}=1}^{\mathrm{N}}$. As the available facts in the knowledge graph are incomplete, a common task for evaluating knowledge graph representation is knowledge graph completion. In this paper, we approach the task as a link prediction problem, which involves predicting missing values for queries of the form $(\mathrm{h}, \mathrm{r}, ?)$ or $(?, \mathrm{r}, \mathrm{t})$.

The Composite Reasoning Framework In this framework, we utilize the notation $\mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t})$ to represent the plausibility of a triple, such that the prediction to $(\mathrm{h}, \mathrm{r}, ?)$ is the t with highest plausibility.

To illustrate composite reasoning, consider the example of $\mathrm{score}(\mathrm{Alphabet}, \mathrm{employee}, ?) = \mathrm{score}(\mathrm{Google}, \mathrm{employee}, ?) + \mathrm{score}(\mathrm{DeepMind}, \mathrm{employee}, ?) + \cdots$. In order to effectively represent this composite reasoning, it must satisfy the following condition:

$$\forall \mathrm{t}, \mathrm{score}(\mathrm{Alphabet}, \mathrm{employee}, \mathrm{t}) = \mathrm{score}(\mathrm{Google}, \mathrm{employee}, \mathrm{t}) + \mathrm{score}(\mathrm{DeepMind}, \mathrm{employee}, \mathrm{t}) \tag{1}$$

Building upon this example, we formally define composite reasoning as the process of combining known facts about other entities to model the target entity. Specifically, given a query $(\mathrm{h}, \mathrm{r}, ?)$, the composite reasoning framework is formulated as:

$$\forall \mathrm{t}, \mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \sum_{(\mathrm{h_i}, \mathrm{r}, \mathrm{t}) \in \mathrm{KG}} \alpha_{\mathrm{i}} \cdot \mathrm{score}(\mathrm{h_i}, \mathrm{r}, \mathrm{t}) \tag{2}$$

Here, α_i represents the weight assigned to the i-th entity h_i, and the constraint $(\mathrm{h_i}, \mathrm{r}, \mathrm{t}) \in \mathrm{KG}$ ensures that the prediction relies on known facts. In Sec 3, we will demonstrate how different models can be explained using different α_i values within this framework.

3 Unifying KGC via Composite Reasoning

In this section, we explain how to use the composite reasoning framework to unify different KGC models, including TF-based models (Sec 3.1), translational models (Sec 3.2), instance-based learning models (Sec 3.3), and the DURA regularizer (Sec 3.4).

3.1 Explaining TF Models

Tensor Factorization (TF)-based models is a widely studied class of knowledge graph embedding models. The basic idea is to represent a triplet as a high-dimensional tensor. TF models approximate the tensor by decomposing it into the product of tensors corresponding to entities and relations. More formally, a triple $(\mathrm{h}, \mathrm{r}, \mathrm{t})$ is encoded into $\mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t}) \in \mathbb{R}^{\mathrm{d}}$ using:

$$\mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \mathrm{h} \otimes \mathrm{r} \otimes \mathrm{t} \tag{3}$$

where $\mathrm{h}, \mathrm{r}, \mathrm{t} \in \mathbb{R}^{\mathrm{d}}$ represent the tensors for the corresponding head, relation, and tail, and $\otimes$ denotes the product in the Euclidean space (CP(Hitchcock 1927), DistMult(Yang et al. 2015)) or the complex space (ComplEx(Trouillon et al. 2016)). The plausibility of a fact is modeled as the sum of values across all its dimensions:

$$\mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \sum_{\mathrm{i}=1}^{\mathrm{d}} \mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t})_{\mathrm{i}} \tag{4}$$

Compositional View We use the composite reasoning framework to represent TF models based on its linearity. Specifically, for a given entity h in the knowledge graph, we represent it as a linear combination of other entities:

$$\mathrm{h} = \sum_{\mathrm{i}} \mathbf{a}_{\mathrm{i}} \mathrm{h}_{\mathrm{i}} + \Delta \tag{5}$$

where $\mathbf{a}_i$ is the weight of h_{i}, and Δ is the residual.

Since TF is linear, the linear decomposition of h in Eq.(5) also determines its generalization to unknown relations. This allows us to model the relationship between entity composition and model generalization. Specifically, we use this composition to transform the model's prediction of $(\mathrm{h}, \mathrm{r}, ?)$ into a combination of known facts from the knowledge graph:

$$\mathrm{h} = \sum_{\mathrm{h}_{\mathrm{i}} \in \mathrm{KG}(\mathrm{r})} \mathbf{a}_{\mathrm{i}} \mathrm{h}_{\mathrm{i}} + \Delta \tag{6}$$

where $\mathrm{KG}(\mathrm{r})$ denotes the set of entities whose relation r is known in the knowledge graph, i.e., $\mathrm{KG}(\mathrm{r}) = \{\mathrm{h}_{\mathrm{i}} | \exists \mathrm{t}, (\mathrm{h}_{\mathrm{i}}, \mathrm{r}, \mathrm{t}) \in \mathrm{KG}\}$.

Then, we can transform the representation of $(\mathrm{h}, \mathrm{r}, ?)$ into a combination of known facts from the knowledge graph:

$$\forall \mathrm{t}, \mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \sum_{\mathrm{h}_{\mathrm{i}} \in \mathrm{KG}(\mathrm{r})} \mathbf{a}_{\mathrm{i}} \mathbf{e}(\mathrm{h}_{\mathrm{i}}, \mathrm{r}, \mathrm{t}) + \mathbf{e}(\Delta, \mathrm{r}, \mathrm{t}) \tag{7}$$

When explaining TF under the composite reasoning framework, we have:

$$\alpha_{\mathrm{i}}^{(\mathrm{TF})} = \mathbf{a}_{\mathrm{i}} \text{ s.t. } \mathrm{h} = \sum_{\mathrm{h}_{\mathrm{i}} \in \mathrm{KG}(\mathrm{r})} \mathbf{a}_{\mathrm{i}} \mathrm{h}_{\mathrm{i}} + \Delta \tag{8}$$

Representation Capability of the Composition The ability to connect the composition of entities with model generalization is that any query $(\mathrm{h}, \mathrm{r}, ?)$ can be represented by the facts of known entities. To establish such connections, we want to minimize the impact of the residual term $\mathbf{e}(\Delta, \mathrm{r}, \mathrm{t})$. We measure the capability of the entity composition by the residual ratio:

$$\text{residual ratio} = \min_{\mathbf{a}} \frac{||\mathbf{e}(\Delta, \mathrm{r}, \mathrm{t})||}{||\mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t})||} \tag{9}$$

In large-scale knowledge graphs, the number of entities for a given relation is always greater than the dimension of entity embeddings (i.e. $|\mathrm{KG}(\mathrm{r})| > \mathrm{d}$). For example, in WN18RR, the mean of $|\mathrm{KG}(\mathrm{r})|$ is 3722, while d is usually set to 500 or 2000. This means that we can always find a decomposition **a** with residual ratio = 0 for large-scale knowledge graphs. For smaller datasets, the effect of residual is more significant. We will empirically analyze the residual ratio in Sec 5.3.

3.2 Explaining Translational Models

Translational Models treat r as a translation in the entity embedding space. The score function is defined as

$$\mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = ||\mathrm{trans}(\mathrm{h}, \mathrm{r}) - \mathrm{t}|| \tag{10}$$

where $\mathrm{trans}(\mathrm{h}, \mathrm{r})$ is the translated embedding of h for relation r. For example, TransE (Bordes et al. 2013) defines the translation function as $\mathrm{trans}_{\mathrm{TransE}}(\mathrm{h}, \mathrm{r}) = \mathrm{h} + \mathrm{r}$. RotatE (Sun et al. 2019) is another well-known translational model, which consider the translation as a rotate in the complex space $\mathrm{trans}_{\mathrm{RotatE}}(\mathrm{h}, \mathrm{r}) = \mathrm{h} \circ \mathrm{r}$.

Compositional View For the query $(\mathrm{h}, \mathrm{r}, ?)$, we assume that at least one entity already contain the target t of relation r. That is, $\exists \mathrm{h}_{\mathrm{i}}, (\mathrm{h}_{\mathrm{i}}, \mathrm{r}, \mathrm{t}) \in \mathrm{KG}$. For example, when predicting $(\mathrm{Alphabet}, \mathrm{employee}, ?) =$ Jeff Dean, we assume that a known fact about employee-Jeff Dean is already in the training knowledge graph (e.g. (Google, employee, Jeff Dean)). It is noteworthy that one-to-one relations cannot be represented under such assumption.

We also assume that the high expressiveness of high-dimensional neural networks leads to very low training loss:

$$\forall (\mathrm{h}, \mathrm{r}, \mathrm{t}) \in \mathrm{KG}, ||\mathrm{trans}(\mathrm{h}, \mathrm{r}) - \mathrm{t}|| = 0 \tag{11}$$

Given the aforementioned assumptions, for any query $(\mathrm{h}, \mathrm{r}, ?)$, we can establish that for all candidate answer t, there exists $(\mathrm{h}_{\mathrm{i}}, \mathrm{r}, \mathrm{t}) \in \mathrm{KG}$ such that $||\mathrm{trans}(\mathrm{h}_{\mathrm{i}}, \mathrm{r}) - \mathrm{t}|| = 0$. Therefore, we have:

$$\mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = ||\mathrm{trans}(\mathrm{h}, \mathrm{r}) - \mathrm{trans}(\mathrm{h}_{\mathrm{i}}, \mathrm{r})|| \tag{12}$$

Taking it further, we use h_{i} to express the prediction results of the translational model. According to Eq. (11) and Eq. (12), the top-k tail entities can be represented by:

$$\begin{aligned} &\mathrm{topk}_{\mathrm{t}} \mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \\ &\arg\min \mathrm{k}_{\mathrm{t}} ||\mathrm{trans}(\mathrm{h}, \mathrm{r}) - \mathrm{trans}(\mathbf{h}(\mathrm{r}, \mathrm{t}), \mathrm{r})|| \end{aligned} \tag{13}$$

where $\mathbf{h}(\mathrm{r}, \mathrm{t})$ denotes the head entity h whose relation r is t in the known KG, i.e., $(\mathbf{h}(\mathrm{r}, \mathrm{t}), \mathrm{r}, \mathrm{t}) \in \mathrm{KG}$.

Based on Eq. (13), we use $||\mathrm{trans}(\mathrm{h}, \mathrm{r}) - \mathrm{trans}(\mathbf{h}(\mathrm{r}, \mathrm{t}), \mathrm{r})||$ to align translational models with the composite reasoning framework:

$$\alpha_{\mathrm{i}}^{(\mathrm{TRANS})} = ||\mathrm{trans}(\mathrm{h}, \mathrm{r}) - \mathrm{trans}(\mathrm{h}_{\mathrm{i}}, \mathrm{r})|| \tag{14}$$

3.3 Explaining Instance-based Learning Models

CIBLE (Cui and Chen 2022) is a recently proposed knowledge graph completion model based on instance-based learning. This model utilizes prototypes modeling to represent the knowledge graph. Its scoring function for $(\mathrm{h}, \mathrm{r}, ?)$ can be formulated by:

$$\mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \beta \sum_{(\mathrm{p}, \mathrm{r}, \mathrm{t}) \in \mathrm{KB}} \mathrm{f}_{\mathrm{hr}}(\mathrm{p}) \tag{15}$$

where β is a coefficient to normalize the score, $\mathrm{f}_{\mathrm{hr}}(\mathrm{p})$ denotes the plausibility of a candidate prototype p:

$$\mathrm{f}_{\mathrm{hr}}(\mathrm{p}) = \max(\gamma - ||\mathrm{trans}_{\mathrm{r}}(\mathrm{emb}(\mathrm{h})) - \mathrm{trans}_{\mathrm{r}}(\mathrm{emb}(\mathrm{p}))||, 0) \tag{16}$$

When explaining CIBLE with composite reasoning, we have:

$$\alpha_{\mathrm{i}}^{(\mathrm{CIBLE})} = \mathrm{f}_{\mathrm{hr}}(\mathrm{h}_{\mathrm{i}}) \tag{17}$$

3.4 Explaining the DURA Regularizer

DURA is a recently proposed effective and widely-applicable KGC regularizer. Its basic form is:

$$\mathrm{score}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = ||\mathrm{h} \otimes \mathrm{r} - \mathrm{t}|| \tag{18}$$

We noticed that its form is compatible with the translational model in Eq. (10). Thus, similar to Eq. (14), we represent DURA under the composite reasoning framework:

$$\alpha_{\mathrm{i}}^{(\mathrm{DURA})} = ||\mathrm{h} \otimes \mathrm{r} - \mathrm{h}_{\mathrm{i}} \otimes \mathrm{r}|| \tag{19}$$

3.5 Understanding and Comparing the Composite Reasoning of KGC Models

In the preceding discussion, we employed composite reasoning to elucidate various KGC models. In this subsection, we provide a more direct understanding of composite reasoning by visualizing how different KGC models combine facts from diverse entities. To illustrate this, we consider the query $(\mathrm{Mexico}, \mathrm{official\ language}, ?)$ from the FB15k-237 dataset and present the top eight entities ranked by their corresponding α_i values.

The visualization results show that the composite reasoning framework provides a convincing explanation for the behavior of KGC models. In the majority of cases, the top entities identified by the framework align closely with human intuition. For instance, the TransE model leverages facts about Brazil and Canada, which are highly associated with Mexico, as well as Spain, which shares the same official language as Mexico. These findings demonstrate the effectiveness of the composite reasoning framework in capturing meaningful relationships between entities.

However, the composition results obtained by two TF-based models, CP and ComplEx, yielded unexpected outcomes. The top entities identified exhibited both low relevance to Mexico and different tail entities, such as Turks and Caicos Islands (TC Islands) and Macau. Intuitively, these entities are unlikely to contribute to accurate predictions.

This phenomenon is not a mere coincidence. To further investigate it, we computed the average composite rationality between the top 8 decomposed entities and the query entity for all queries in the test set in FB15k-237. The composite rationality between two entities was measured using the Jaccard coefficient of their corresponding triplets.

Table 1 presents the results obtained for different models. Notably, the tensor factorization-based CP and ComplEx models displayed significantly lower average relevance values compared to the other models. In Sec 4, we will delve into the causes behind this phenomenon, discuss its experimental implications, and propose solutions.

4 Modeling and Alleviating Composition Risk for TF-based Models

4.1 Measuring Erroneous Knowledge Composition via Composition Risk

Under the composite reasoning framework, the prediction regarding h is an aggregation of the known facts of other entities. As a result, decomposing into certain entities is more likely to result in generalization errors than others.

For instance, if the model decomposes Mexico as $\mathrm{Mexico} = \mathbf{a}_1\mathrm{Panama} + \mathbf{a}_2\mathrm{Macau}$, the predictions for Mexico's official language will use Macau's facts, which is obviously riskier than using Panama's facts. Therefore, we aim to identify and mitigate the impact of entities with higher risk to generalization errors.

Furthermore, unlike the decomposition into $\mathrm{h_j}$, the term $\mathbf{e}(\Delta, \mathrm{headquarters}, \mathrm{t})$ cannot be represented by facts of existing entities. We posit that this residual term is also riskier.

Motivated by this, we propose the concept of *composition risk* for TF models, which refers to the risk of generalization errors caused by decomposing into riskier entities or the residual. More formally, when representing the composition of h, we divide the entities into two categories: reliable entities and risky entities. For example, Panama is a reliable entity for Mexico, while Macau is a risky entity. We want the composite reasoning to rely on reliable entities. This is illustrated in Fig. 2. Suppose for $(\mathrm{h}, \mathrm{r}, ?)$, the composition of h is:

$$\mathrm{h} = \sum_{\mathrm{h_i} \in \mathrm{reliable(h)} \cap \mathrm{KG(r)}} \mathbf{a}_\mathrm{i}\mathrm{h_i} + \sum_{\mathrm{h_j} \in \mathrm{risky(h)} \cap \mathrm{KG(r)}} \mathbf{a}_\mathrm{j}\mathrm{h_j} + \Delta \tag{20}$$

According to Eq. (7), to make the model's behavior be more consistent with entities in $\mathrm{reliable(h)}$, we expect $\sum_{\mathrm{h_i} \in \mathrm{reliable(h)}} \mathbf{a}_\mathrm{i}\mathbf{e}(\mathrm{h_i}, \mathrm{r}, \mathrm{t})$ to be close to $\mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t})$ and $\sum_{\mathrm{h_j} \in \mathrm{risky(h)}} \mathbf{a}_\mathrm{j}\mathbf{e}(\mathrm{h_j}, \mathrm{r}, \mathrm{t}) + \mathbf{e}(\Delta, \mathrm{r}, \mathrm{t})$ to be close to zero. We formulate the composition risk formally as the ratio associated with the risky composition and the residual:

$$\mathrm{cr}_\mathbf{a}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \frac{||\mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t}) - \sum_{\mathrm{h_i} \in \mathrm{reliable(h)} \cap \mathrm{KG(r)}} \mathbf{a}_\mathrm{i}\mathbf{e}(\mathrm{h_i}, \mathrm{r}, \mathrm{t})||}{||\mathbf{e}(\mathrm{h}, \mathrm{r}, \mathrm{t})||} \tag{21}$$

By minimizing this ratio, we effectively reduce the impact of risky decompositions and the residual.

It should be noted that for a fixed TF model, there are multiple compositions $\mathbf{a}$ for h. As long as there exists an $\mathbf{a}$ such that $\mathrm{cr}_\mathbf{a}(\mathrm{h}, \mathrm{r}, \mathrm{t})$ is minimized, the model's prediction for $(\mathrm{h}, \mathrm{r}, \mathrm{t})$ will depend maximally only on the entities in $\mathrm{reliable(h)}$, which is the desired outcome. Therefore, we take the $\mathbf{a}$ that minimizes the $\mathrm{cf}_\mathbf{a}(\mathrm{h}, \mathrm{r}, \mathrm{t})$ to define the composition risk.

Definition 4.1 (Composition risk)**.** Composition risk w.r.t. $(\mathrm{h}, \mathrm{r}, \mathrm{t})$ is defined as:

$$\mathrm{cr}(\mathrm{h}, \mathrm{r}, \mathrm{t}) = \min_{\mathbf{a}} \mathrm{cr}_\mathbf{a}(\mathrm{h}, \mathrm{r}, \mathrm{t}) \tag{22}$$

4.2 Composition Risk Leads to the Violation of Collaborative Filtering Assumption

The concept of using tensor factorization is based on the principle of collaborative filtering (Koren, Bell, and Volinsky 2009). One of the central assumption of collaborative filtering in knowledge graphs (KGs) is that entities that share similar relationships are likely to have similar characteristics in other relationships as well. For example, Alphabet and Google share the same CEO, so they are likely to have the same headquarters.

However, we found that traditional TF models can easily fit the training data while violating the collaborative filtering assumption. The learned embeddings of similar entities are not necessarily similar and may even be orthogonal. This phenomenon has already been reported in (Zhang, Cai, and Wang 2020). In this paper, we aim to further explain how this phenomenon leads to generalization errors from the perspective of composite reasoning.

We illustrate this by the example in Table 2. Despite fitting all the training data, the TF model does not adhere to the collaborative filtering assumption. Although Google and

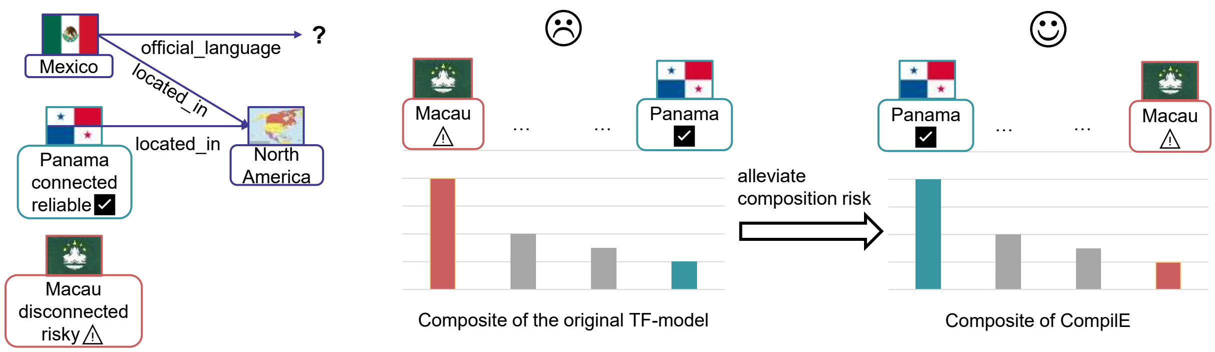


Figure 2: Motivation of modeling and alleviating composition risk. The composite of the original TF-model may rely on risky entities (e.g. Macau for Mexico since they are disconnected). By alleviating composite risk, we encourage the composite to rely on reliable entities (e.g. Panama for Mexico since they are connected.)

	Google [1,0]	Alphabet [0,1]
CEO, Sundar Pichai [1,1]	1 ($\mathrm{cr} = 1$)	1 ($\mathrm{cr} = 1$)
headquarters, Mountain View [1,0]	1 ($\mathrm{cr} = 1$)	**pred = 0**

Table 2: An example of how TF models can violate the collaborative filtering assumption and result in incorrect predictions. The goal is to predict the value in the bottom right corner. The values in square brackets represent the corresponding tensors.

Alphabet have the same CEO, their embeddings are orthogonal. This results in the model not being able to predict that Alphabet's headquarters is in Mountain View using the knowledge of Google's headquarters.

We link the violation of the collaborative filtering assumption to composition risk. The high expressive capacity of high-dimensional TF models can cause the model to neglect learning effective entity compositions. We show the composition risk of the facts in Table 2. Even though the model fits the training data perfectly, it still has a high composition risk because Alphabet and Google are connected. Reducing the composition risk encourages the model to learn the association between Google and Alphabet, and thus make accurate predictions.

4.3 Approximating and Minimizing the Lower Bound of Composition Risk

Minimizing composition risk requires accurate estimation of $\mathrm{reliable}(h_i)$ and $\mathrm{risky}(h_i)$. In this subsection, we will explain how to estimate and optimize the lower bound of the composition risk as an alternative to directly optimizing it.

$\mathrm{reliable}(h)$ is a set of entities that have highly consistent facts with h and can be used for prediction. It is reasonable to assume that these entities have at least one identical fact with h from the KG.

$$\mathrm{connected}(h) = \{h_i | h \neq h_i, \exists r_1, r_2, t, (h, r_1, t) \in \mathrm{KG}, (h_i, r_2, t) \in \mathrm{KG}\} \tag{23}$$

Based the linearity of TF models, the lower bound of the composition risk can be calculated in Theorem 4.2.

Theorem 4.2 (Lower bound of composition risk)**.** *Assuming that* $\mathrm{connected}(h)$ *is a weaker restriction of* $\mathrm{reliable}(h)$*, i.e.* $\mathrm{reliable}(h) \subseteq \mathrm{connected}(h)$*, we have:*

$$\mathrm{cr}(h, r, t) \geq \min_{\mathbf{a}} \frac{||\mathbf{e}(h, r, t) - \sum_{h_i \in \mathrm{connected}(h) \cap \mathrm{KG}(r)} \mathbf{a}_i \mathbf{e}(h_i, r, t)||}{||\mathbf{e}(h, r, t)||} \tag{24}$$

We use the lower bound as the approximated composition risk, denoted as $\hat{\mathrm{cr}}(h, r, t)$. See the proof in the supplementary material.

4.4 The Impact of (Approximated) Composition Risk on Generalization Errors

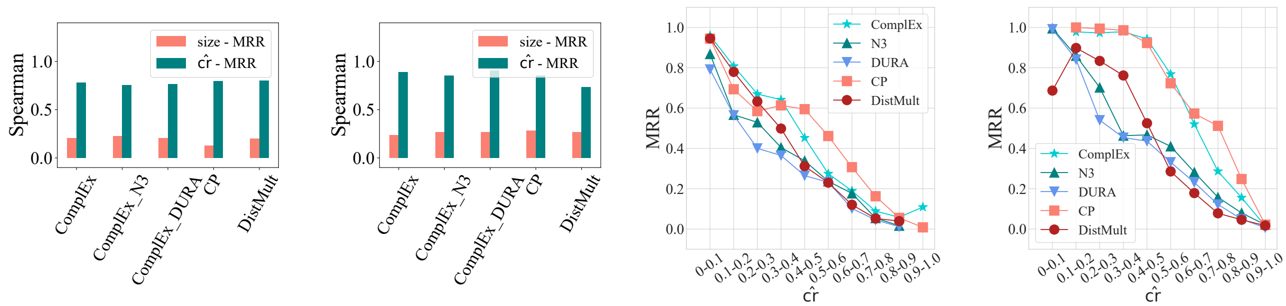
To demonstrate the relationship between composition risk and generalization errors, we examined the correlation between the approximated composition risk and the accuracy of predictions for entities in real-world datasets.

Specifically, we investigate the relationship between the model's prediction quality, as measured by the mean reciprocal rank (MRR), and the composition risk (CR) of queries in the test set. We use Spearman's rank correlation coefficient to quantify the correlation, with a stronger correlation indicating a greater impact of $\hat{\mathrm{cr}}$ on the model's generalization ability. Additionally, we compare this correlation to the relationship between the frequency of an entity in the knowledge graph and the MRR, as a baseline. This is because the predictions for more frequent entities tend to be easier. The results are presented in Fig. 3(a) 3(b). We also plot the direct impact of $\hat{\mathrm{cr}}$ on MRR in Fig. 3(c) 3(d).

It can be seen that the correlation of the approximated composition risk $\hat{\mathrm{cr}}$ is significantly stronger. This verifies $\hat{\mathrm{cr}}$ brings generalization errors. Since $\hat{\mathrm{cr}}$ is a metric that can be optimized, this motivates us to decrease it during training.

4.5 Alleviating Composition Risk in Training

Incorporating Composition Risk into TF Models To minimize the composition risk in TF models, we incorporate it



(a) FB15k-237 (b) WN18RR (c) FB15k-237 (d) WN18RR

Figure 3: Correlation between composition risk and prediction performance for fact in test sets. For N3 and DURA, we use ComplEx as their base models.

as a penalty term in the training loss. Specifically, we use the following loss function:

$$\mathcal{L} = \mathcal{L}_{\text{origin}} + \beta \sum_{(\mathrm{h},\mathrm{r},\mathrm{t})\in \mathrm{KG}} \hat{\mathrm{cr}}(\mathrm{h},\mathrm{r},\mathrm{t}) \tag{25}$$

where $\mathcal{L}_{origin}$ is the original loss function for the TF model(Zhang, Cai, and Wang 2020; Lacroix, Usunier, and Obozinski 2018), and β is the weight for the composition risk term. We denote this model as CompilE (composition risk alleviation)

Finding the Optimal Composition In Eq. (24), we need to compute $\mathbf{a}$ to minimize $\hat{\mathrm{cr}}$. This can be done by solving a least squares problem, as the equation is a classical linear regression problem.

5 Effect of Reducing Composition Risk in TF models

5.1 Setup

Baselines We compare our proposed method with several state-of-the-art models and regularization techniques as baselines. These include classic tensor factorization models such as ComplEx(Trouillon et al. 2016), DistMult(Yang et al. 2015), and CP(Hitchcock 1927), regularization methods like N3(Lacroix, Usunier, and Obozinski 2018) and DURA(Zhang, Cai, and Wang 2020), and other state-of-the-art KGC models like TransE(Bordes et al. 2013), RotatE(Sun et al. 2019), NeuralLP(Yang, Yang, and Cohen 2017), RNNLogic(Qu et al. 2020), CIBLE(Cui and Chen 2022), and NBFNet(Zhu et al. 2021). We use ComplEx as our default model and also incorporate traditional regularization techniques to reduce parameter complexity. We refer to our model with N3 regularization as CompilE$_N$ and with DURA regularization as CompilE$_D$.

Dataests We use four datasets of different scales, including two larger datasets (FB15k-237 and WN18RR), and two smaller datastes (UMLS and Kinship).

Evaluation We use standard evaluation metrics commonly used in KGC, including Mean Rank (MR), Mean Reciprocal Rank (MRR), and Hits@k under the filtered setting.

5.2 Main Results

The main results for the four benchmarks are presented in Table 3 and Table 4. It can be observed that CompilE outperforms all other baselines on smaller datasets. On larger datasets, it also achieves better performance than other baselines, except for the GNN-based NBFNet. This confirms the effectiveness of our approach.

Effect improvement across different datasets and baselines Our method shows improvement over both DURA and N3 on all four datasets. This suggests that traditional TF-based models need to optimize their knowledge composition in addition to using state-of-the-art regularizers. This is also supported by the results shown in Fig.3.

Effect on knowledge-sparse datasets Our method demonstrates higher effectiveness on small-scale datasets. For example, on Kinship, the MRR of CompilE$_N$ improved by 3.2% over other TF-based models. We believe this is because overfitting is more likely to occur on smaller datasets, making effective composition more crucial. This supports the value of our proposed method in knowledge-sparse scenarios.

5.3 Capabilities of the Composite Reasoning

In Sec 3.1, we explained that the effectiveness of the entity decomposition framework can be assessed using the residual ratio. We plot the residual ratios of various models on different datasets in Fig. 4. Consistent with our analysis in Sec 3.1, the residual ratios are close to zero on large-scale knowledge graphs, which suggests that entity decomposition is more effective in these cases. Even on small-scale knowledge graphs, CompilE effectively reduces the residual ratios and thus improves the capability of entity decomposition.

6 Related Work

Researchers have discovered that the representation of knowledge graphs can be improved by optimizing the way different facts are composited. Prior studies have implicitly optimized the compositionality between entities by decreasing model complexity (Lacroix, Usunier, and Obozinski 2018). More recent efforts, however, have focused on directly optimizing specific compositions between facts, such

	FB15k-237				WN18RR			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE	0.294	-	-	0.465	0.226	-	-	0.501
RotatE†	0.338	0.241	0.375	0.533	0.476	0.428	0.492	0.571
NeuralLp†	0.237	0.173	0.259	0.361	0.381	0.368	0.386	0.408
RNNLogic+†	0.349	0.258	0.385	0.533	0.513	0.471	0.532	0.579
CIBLE†	0.341	0.246	0.378	0.532	0.490	0.446	0.507	0.575
NBFNet†	**0.415**	**0.321**	**0.454**	**0.599**	**0.551**	**0.497**	**0.573**	**0.666**
TF-based models								
DistMult	0.343	0.251	0.376	0.525	0.440	0.410	0.451	0.499
CP	0.332	0.244	0.364	0.509	0.438	0.416	0.444	0.482
ComplEx	0.350	0.259	0.386	0.531	0.460	0.429	0.471	0.521
DURA	0.371	0.276	-	0.560	0.491	0.449	-	0.571
N3	0.367	0.271	0.403	0.558	0.488	0.441	0.503	0.581
CompilE$_D$	**0.372**	**0.277**	**0.408**	**0.563**	**0.495**	**0.453**	**0.510**	0.579
CompilE$_N$	0.368	0.272	0.404	0.559	0.492	0.447	0.506	**0.582**

Table 3: Effect on larger benchmarks.†: the results are from (Cui and Chen 2022).

	UMLS				KINSHIP			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
RotatE†	0.744	0.636	0.822	0.939	0.651	0.504	0.755	0.932
NeuralLP†	0.483	0.332	0.563	0.775	0.302	0.167	0.339	0.596
RNNLogic†	0.842	0.772	0.891	0.965	0.722	0.598	0.814	0.949
CIBLE†	0.856	0.787	0.916	0.970	**0.728**	**0.603**	0.820	0.956
NBFNet†	0.778	0.688	0.840	0.938	0.606	0.435	0.725	0.937
TF-based-models								
DistMult	0.725	0.615	0.788	0.954	0.456	0.270	0.537	0.892
CP	0.819	0.718	0.910	0.964	0.653	0.507	0.755	0.937
ComplEx	0.840	0.765	0.902	0.968	0.660	0.513	0.762	0.938
DURA	0.841	0.767	0.900	0.966	0.670	0.526	0.773	0.941
N3	0.842	0.767	0.905	0.969	0.697	0.560	0.796	0.953
CompilE$_D$	0.861	0.792	0.920	0.972	**0.724**	**0.593**	**0.830**	**0.962**
CompilE$_N$	**0.868**	**0.802**	**0.924**	**0.973**	0.713	0.579	0.813	0.955

Table 4: Effect on smaller benchmarks. The improvement brings by CompilE is more significant.

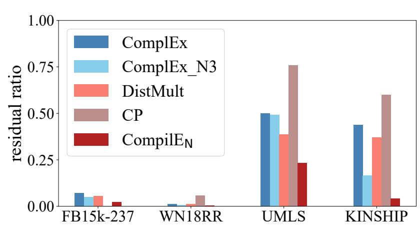


Figure 4: Representation capabilities of the entity decomposition for model generalization.

as equal and inverse relations (Minervini et al. 2017), compositions between entities of the same category (Guo et al. 2015; Cao et al. 2022), and compositions between entities under the same head-relation (Zhang, Cai, and Wang 2020). However, these works lack a general framework to model one-to-many fact composition and do not accurately depict the connection between composition regularization and model generalization.

7 Conclusion

This study provides a comprehensive understanding of composite reasoning for KGC models, including TF-based models, translational models, instance-based learning models, and KGC regularizers. We take advantage of the composite reasoning to uncovers a novel issue with TF-based models where irrelevant entities can be incorporated into the inference process, causing generalization errors. This issue is rooted in the models' violation of the low-rank assumption due to inaccurate composite learning. We propose to mitigate this composition risk, effectively enhancing the performance of these models.

References

Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *NeurIPS*.

Cao, Z.; Xu, Q.; Yang, Z.; and Huang, Q. 2022. ER: Equivariance Regularizer for Knowledge Graph Completion. In *AAAI*.

Cui, W.; and Chen, X. 2022. Instance-based Learning for Knowledge Base Completion. In *Advances in Neural Information Processing Systems*.

Guo, S.; Wang, Q.; Wang, B.; Wang, L.; and Guo, L. 2015. Semantically smooth knowledge graph embedding. 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)*, 84–94.

Hitchcock, F. L. 1927. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4): 164–189.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37.

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

Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; and Zhu, X. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

Minervini, P.; Costabello, L.; Muñoz, E.; Nováček, V.; and Vandenbussche, P.-Y. 2017. Regularizing knowledge graph embeddings via equivalence and inversion axioms. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 668–683. Springer.

Qu, M.; Chen, J.; Xhonneux, L.-P.; Bengio, Y.; and Tang, J. 2020. RNNLogic: Learning Logic Rules for Reasoning on Knowledge Graphs. In *ICLR*.

Sun, Z.; Deng, Z.-H.; Nie, J.-Y.; and Tang, J. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In *ICLR*.

Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; and Bouchard, G. 2016. Complex embeddings for simple link prediction. In *ICML*, 2071–2080. PMLR.

Yang, B.; Yih, W.-t.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*.

Yang, F.; Yang, Z.; and Cohen, W. W. 2017. Differentiable learning of logical rules for knowledge base reasoning. In *NeurIPS*.

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

Zhang, Z.; Cai, J.; Zhang, Y.; and Wang, J. 2020. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*, 3065–3072. AAAI Press.

Zhu, Z.; Zhang, Z.; Xhonneux, L.-P.; and Tang, J. 2021. Neural bellman-ford networks: A general graph neural network framework for link prediction. *NeurIPS*.