

# Node-based Knowledge Graph Contrastive Learning for Medical Relationship Prediction

## ABSTRACT

The embedding of Biomedical Knowledge Graphs (BKGs) generates robust representations, valuable for a variety of artificial intelligence applications, including predicting drug combinations and reasoning disease-drug relationships. Meanwhile, contrastive learning (CL) is widely employed to enhance the distinctiveness of these representations. However, constructing suitable contrastive pairs for CL, especially within Knowledge Graphs (KGs), has been challenging. In this paper, we proposed a novel node-based contrastive learning method for knowledge graph embedding, NC-KGE. NC-KGE enhances knowledge extraction in embeddings and speeds up training convergence by constructing appropriate contrastive node pairs on KGs. This scheme can be easily integrated with other knowledge graph embedding (KGE) methods. For downstream task such as biochemical relationship prediction, we have incorporated a relation-aware attention mechanism into NC-KGE, focusing on the semantic relationships and node interactions. Extensive experiments show that NC-KGE performs competitively with state-of-the-art models on public datasets like FB15k-237 and WN18RR. Particularly in biomedical relationship prediction tasks, NC-KGE outperforms all baselines on datasets such as PharmKG8k-28, DRKG17k-21, and BioKG72k-14, especially in predicting drug combination relationships. We release our code at <https://github.com/DeepLearningPS/NC-KGE>.

## CCS CONCEPTS

- Data Mining → Link Prediction.

## KEYWORDS

Contrastive Learning, Graph Neural Network, Medical Relationship Prediction, Biomedical Knowledge Graph

## ACM Reference Format:

. 2024. Node-based Knowledge Graph Contrastive Learning for Medical Relationship Prediction. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 13 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 INTRODUCTION

Knowledge graphs (KGs) are intricate networks, representing the human knowledge, where nodes denote meaningful entities and

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX, June 03–05, 2024, Woodstock, NY

© 2024 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/XXXXXXX.XXXXXXX>

edges signify their relationships. Specifically, Biomedical Knowledge Graphs (BKGs), such as TTD, PharmGKB, OMIM and DrugBank, start from the high-quality manually curated biomedicine knowledge and capture the intricate interactions among biomedical entities like genes, chemicals, and diseases. They play a pivotal role in many artificial intelligence applications including drug repositioning, adverse drug reaction analysis, and proteomics data analysis. Recently, Knowledge graph embedding methods were developed to generate compact, distributed representations of entities and relations for performance improvements. Once learned, these embeddings can be analyzed using diverse scoring techniques to provide probability scores for triplets, valuable for downstream tasks such as prediction, clustering, and visualization. Especially, biomedical relationship prediction, uncovers hidden associations between biological entities, enhancing disease understanding and biomarker discovery, illustrating biological pathways.

Recent KGE methods can be roughly categorized into structure-based embeddings and enhanced knowledge graph embeddings. The structure-based embeddings focus on the usage of knowledge graph facts and can be further categorized into 3 types. (1) translational distance models such as TransE [3], RotatE [39], QuatE [58], DualE [4] and HAKE [59]. (2) semantic matching methods including RESCAL [33], TATEC [12], DistMult [53], HolE [32], ComplEx [42], ANALOGY [26] and SimplE [20]. (3) neural-based methods, especially GNN-based methods including RGCN [37], SACN [38], KBGAT [31], A2N [2], CompGCN [44] and SE-GNN [21]. The enhanced knowledge graph embeddings aim to use additional information for a better representation, including (1) text-enhanced knowledge embedding such as TEKE [49], DKRL [52], KG-BERT [54], (2) logic-enhanced embedding like KALE [13], (3) image-enhanced embedding and etc. Although these methods have gained improvements in powerful representation of entities in knowledge graphs, it is still challenging for biomedical relationship predictions due to knowledge bases from the biomedical domain are usually sparse, redundant and incomplete.

Recently, self-supervised learning has gradually returned to people's field of vision due to the rise of contrastive learning. Contrastive learning methods [8, 15] have become one of the most popular self-supervised methods for learning representation [7, 40]. In computer vision, unsupervised contrastive learning methods are even better than supervised pre-training methods in object detection and segmentation tasks [16, 30]. Contrastive learning relies on two key elements: similar (positive)  $(x, x^+)$  and dissimilar (negative)  $(x, x^-)$  data point pairs. The training target, usually noise contrast estimation [14], guides the learned representation  $f$  to map the positive pair to a nearby location, and the negative pair to a farther location; in addition, other targets are also considered [7]. The classic contrastive learning form [16] is as follows:

$$\mathcal{L}_q = -\log \frac{\exp(q \cdot k_+ / \tau)}{\exp(q \cdot k_+ / \tau) + \sum_{i=1}^K \exp(q \cdot k_i / \tau)}$$

Where  $\tau$  is a temperature hyper-parameter according to [51]. The total sample size includes 1 positive sample and  $K$  negative samples. Intuitively, this loss is a log loss based on the softmax classifier that attempts to classify  $q$  as  $k_+$ , where the number of classifications is  $(K + 1)$ . The contrastive loss function can also be based on other forms [17, 36, 48, 51], such as interval-based loss and variants of NCE loss.

In the past, graph contrastive learning, a self-supervised method, has achieved significant success in generating generalized, transferable, and robust representations for graph structured data. This success has illuminated the path for learning knowledge graph embeddings. Essentially, contrastive learning aims to extract hidden information between samples by bringing similar samples closer and pushing dissimilar ones apart in latent space. Its core objective is to distinguish between a pair of representations from the two augmentations of the same sample (positives) apart from the  $k$  pairs of representation from the other (negative) samples. Constructing highly confident contrastive pairs is crucial for the discriminative power of contrastive learning models. However, due to the intricate structures within knowledge graphs, defining these contrastive pairs is challenging. Consequently, there are only limited attempts to integrate contrastive learning strategies with Knowledge Graph Embedding (KGE) methods.

SimKGC [47], for instance, creates contrastive pairs using semantic similarity through language models, deviating from the previous graph contrastive learning models that fully mined information underlying graph structures. However, the effectiveness of contrastive KGE methods based on semantic similarity heavily relies on the specific language models used. Another approach, KGE-SymCL [24], utilizes the semantic similarity of entities in relation-symmetrical positions to construct positive contrastive samples. However, extracting the structure of symmetric relations is a tedious process and there is no significant improvement compared to SimKGC. In biomedical knowledge graphs, entities represent gene codes, targets, or chemical compounds. The features generated by language models may lead to inaccurate semantic estimation in this context. Therefore, there is a pressing need for a more stable and universally applicable contrastive learning criterion tailored for biomedical knowledge graphs.

In our study, we introduced a new and versatile node-based contrastive method for knowledge graph embeddings, termed NC-KGE, specifically designed for predicting biomedical relationships. For a given fact triplet in knowledge graph, NC-KGE identifies triplets in the knowledge graph where one entity and the relation type remain the same as positive samples, while all other triplets are considered negative samples. By maximizing the similarity score between positive samples and minimizing it between negative ones using a modified classic contrastive learning loss, NC-KGE enhances the convergence speed during training and improve the performance of non-contrastive methods, such as CompGCN and SE-GNN. Additionally, we incorporated a relation-aware multi-head attention (RAMHA) mechanism into NC-KGE to enhance the utilization of relation semantics and interactions among relations and entities. We evaluated NC-KGE's performance in relation prediction on both general public datasets (FB15k-237 and WN18RR) and biomedical-focused knowledge graphs (PharmKG8k-28, DRKG17k-21, and BioKG72k-14). Our extensive experiments revealed that

NC-KGE competes effectively with state-of-the-art models on general knowledge graphs, and surpasses all baselines on biomedical knowledge graphs, particularly excelling in predicting drug combination relationships.

## 2 RELATED WORK

Knowledge Graph Embedding (KGE) aims to encode the entities and relations to the low dimensional vector or matrix space while maximally preserving its topological properties. Recent existing KGE models can be roughly categorized into structure based embedding and enhanced knowledge embeddings, reviewed by Singh et al. [9] and MINERVINI et al. [34]. In this work, structure-based knowledge graph embedding methods are more related than enhanced knowledge graph embeddings.

### 2.1 Structure-based Knowledge graph embedding methods

**Translation distance models** interpret relations as translations operation from a head node to an tail node in latent space, e.g., TransE [3], TransH [50], TransR [25] and etc. TransE was among the initial attempts to represent relations as addition operations between entities. TransH projects entities onto relation-specific hyperplanes, allowing entities to play different roles in various relations. TransR maps nodes and relations into distinct entities spaces and relation-specific spaces. RotaE [39] treats the relation as a rotation operation. PairRE [5] can encode complex relationships and multiple relationship patterns at the same time. Moreover, HousE [22] involves a novel parameterization based on the designed Householder transformations for rotation and projection.

**Semantic matching models**, including RESCAL [33], DistMult [53], ComplEx [42], ConvE [11], SimplE [20], CrossE [18], QuatE [58], DualE [4], are developed based on similarity scoring functions. RESCAL utilizes a bilinear similarity function to compute the scores of knowledge triples and assumes that positive triples have higher scores than negative ones. DisMult simplifies the bilinear similarity function through using diagonal matrix. ComplEx further generalized DisMult by using complex embeddings and Hermitian dot products. Besides, the advantage of quaternion representations is leveraged by QuatE to enrich the correlation information between head and tail entities based on relational rotation quaternions. Inspired by it, DualE is proposed to gain a better expressive ability by projecting the embeddings in dual quaternion space.

**Neural-based methods**, including ConvE [11], RGCN [37], SACN [38], KBGAT [31], A2N [2], CompGCN [44] and SE-GNN [21]. For example, ConvE introduces the use of convolutional layers to extract information. RGCN introduces a relation-specific transformation to integrate relation information with message aggregation. RGHAT [60] incorporates a two-level attention mechanism, addressing relations and entities separately. KE-GCN [57] introduces a joint propagation method to update node and edge embeddings simultaneously. CompGCN proposes various composition operations for neighbor aggregation to model the structure pattern of multi-relational graph. RAGAT [27] constructs separate message functions for different relations, which aims at exploiting the heterogeneous characteristics of knowledge graphs. SE-GNN [21], with its three levels of semantic evidence, achieves in-depth

knowledge representation by meticulously merging these layers through multi-layer aggregation, leading to highly extrapolative knowledge representations.

## 2.2 Contrastive Learning on Knowledge Graph

Graph Contrastive learning (GCL) operates by mining the hidden information in intra-data in a self-supervised manner. These methods, such as GRACE [62], GraphCL [56], AutoGCL [55] and iGCL [23] etc, has been proven highly successful in both nodes representing learning, relation prediction, classification, graph generation and anomaly detections. Recently, only a few researchers attempt to extend graph contrastive learning into knowledge graph embedding learning. SimKGC [47] tend to combine samples with high semantic similarity as positive pairs. This method heavily rely on semantic similarity predicted by the specific language models. However, the language model may estimate sematic similarity inaccurately for the biomedical entities including gene codes, targets, or chemical compounds, making these methods fails for biomedical knowledge graphs. KGE-SymCL [24] utilizes the semantic similarity of entities in relation-symmetrical positions to construct positive contrastive samples. However, extracting the structure of symmetric relations is a tedious process and there is no significant improvement compared to SimKGC.

## 3 PRELIMINARY

Knowledge Graph (KG) is composed of the fact triplets, denoted as  $\mathcal{G} = \{(e_h, r, e_t) \mid e_h, e_t \in \mathcal{E}, r \in \mathcal{R}\}$ , where  $\mathcal{E}$  is the set of entities (i.e., nodes),  $\mathcal{R}$  is the set of relations (i.e., edge types),  $e_h, e_t$  represents the head and tail entity, respectively, and  $r$  represents the relation between them. Relation Prediction, also known as Link Prediction or Knowledge Completion, involves predicting missing links or relationships in a knowledge graph. For a given head entity  $e_h \in \mathcal{E}$  and a relation  $r \in \mathcal{R}$ , the objective is to identify the most suitable tail entity  $e_t \in \mathcal{E}$ , forming a new plausible triple  $(e_h, r, e_t)$  within  $\mathcal{G}$ . Here, we approach this task by scoring all the candidates  $\{(e_h, r, e'_t) \mid e'_t \in \mathcal{E}\}$ , maximizing the scores for genuine triples and minimizing the scores for all the other candidates.

## 4 NC-KGE METHODS

As shown in Figure 1, we show the overall process of NC-KGE. For a given fact triple  $(e_h, r, e_t)$  in biomedical knowledge graph  $\mathcal{G}$ , NC-KGE first constructs positive and negative augmented samples for contrastive learning. Then, a learnable KGE model was used to generate the embeddings for entities and relations in  $\mathcal{G}$ . Thirdly, a similarity function  $S$  is used to score the triple embeddings  $S(z_h, x_r, z_t^*)$  of both positive and negative samples, and calculates the contrastive loss to optimize KGE model by maximizing the scores of genuine triples and minimizing the scores for all the other candidates. During inference, NC-KGE measures the scores of triple embeddings  $S(z_h, x_r, z_t^*)$  for all candidates  $e_t^*$  when provided  $(e_h, r, ?)$ , and the candidate  $e_t$  with highest scores will form a plausible triple with  $e_h$  and  $r$  in  $\mathcal{G}$ . Additionally, finding the head entity  $e_h$  for the provided  $(?, r, e_t)$  can be effortlessly transformed into a similar process.

## 4.1 Construction of node-based contrastive pairs

In biomedical knowledge graph  $\mathcal{G}$ , the entity  $e_h$  can form the same relationship with multiple other entities  $\{e_t \mid e_t \in \mathcal{E}\}$ . For example, a target entity can correspond to multiple inhibitors, and one disease may be associated with multiple genes. For a given triple  $(e_h, r, e_t)$ , supporting another entity  $e_t^+$  forms the same relation  $r$  with the head entity  $e_h$ ,  $e_t^+$  is defined as a positive entity for  $e_h$  and  $r$ ,  $(e_h, r, e_t^+)$  is a positive triple pair for  $(e_h, r, e_t)$ . Additionally, the triple  $(e_h, r, e_t)$  is also regarded as a positive pair for itself. All the other entities  $\{e_t^-\}$  which don't form the relation  $r$  with  $e_h$  are defined as negative entity of  $e_h$ , and  $(e_h, r, e_t^-)$  is also regarded as a negative pair for  $(e_h, r, e_t)$ .

## 4.2 Relation-aware multi-head attention based KGE model

Here, we proposed a relation-aware multi-head attention mechanism (RAMHA) by integrating the relations between entities into attention computations to enhance the utilization of relation semantics and interactions among relations and entities in NC-KGE. As shown in Figure 2, RAMHA mainly includes three phases: attention computing, message passing and information aggregating. By stacking multiple layers of RAMHA, NC-KGE generates the knowledge graph embeddings for both entities and relations.

Supporting the embedding of entity  $e_u$  is denoted as  $z_u^l$  and  $x_r^l$  represents the embedding of relation type  $r$  between  $e_u$  and  $e_v$  in  $l$ th layer of RAMHA, the next layer will aggregate information to  $e_u$  from its neighbor  $e_v$  according to their relations  $r$  in an attention-based message passing process as shown in Equation 1.

$$z_u^{l+1} = \left\|_{c=1}^C \left( \sum_{v \in N(u)} \sum_{r \in \mathcal{R}(u,v)} \alpha_{u,r,v}^{l+1,c} * Message_{u,r,v}^{l+1,c} \right) \right\| \quad (1)$$

where  $\alpha_{u,r,v}^{l+1,c}$  is the attention weight for the embedding triple  $(z_u^l, x_r^l, z_v^l)$  for  $c$ th head in multi-head attentions,  $Message_{u,r,v}$  denotes the information aggregated from  $e_v$  to  $e_u$  with relation  $r$ ,  $\|\cdot\|$  represents the concatenation operation between the total  $C$  heads attentions. The embedding of relation  $r$  is updated in  $l + 1$  layer with Equation 2 and 3.

$$x_r^{l+1} = \left\|_{c=1}^C (x_r^{l+1,c}) \right\| \quad (2)$$

$$x_r^{l+1,c} = MLP_r^{l,c} (x_r^l) \quad (3)$$

where  $MLP_r^{l,c}$  is a relation-type specific MLP network in  $l$ th layer for  $c$ th heads.

The attention weight  $\alpha_{u,r,v}^{l,c}$  is as shown in Equation 4 ~ 7

$$\alpha_{u,r,v}^{l+1,c} = \frac{\langle q_u^{l,c}, MLP_{r,1}^{l,c} (k_v^{l,c} \star x_r^{l+1}) \rangle}{\sum_{w \in N(u)} \sum_{r' \in \mathcal{R}(u,w)} \langle q_u^{l,c}, MLP_{r',1}^{l,c} (k_w^{l,c} \star x_{r'}^{l+1,c}) \rangle} \quad (4)$$

$$q_u^{l,c} = MLP_q^{l,c} (z_u^l) \quad (5)$$

$$k_u^{l,c} = MLP_k^{l,c} (z_u^l) \quad (6)$$

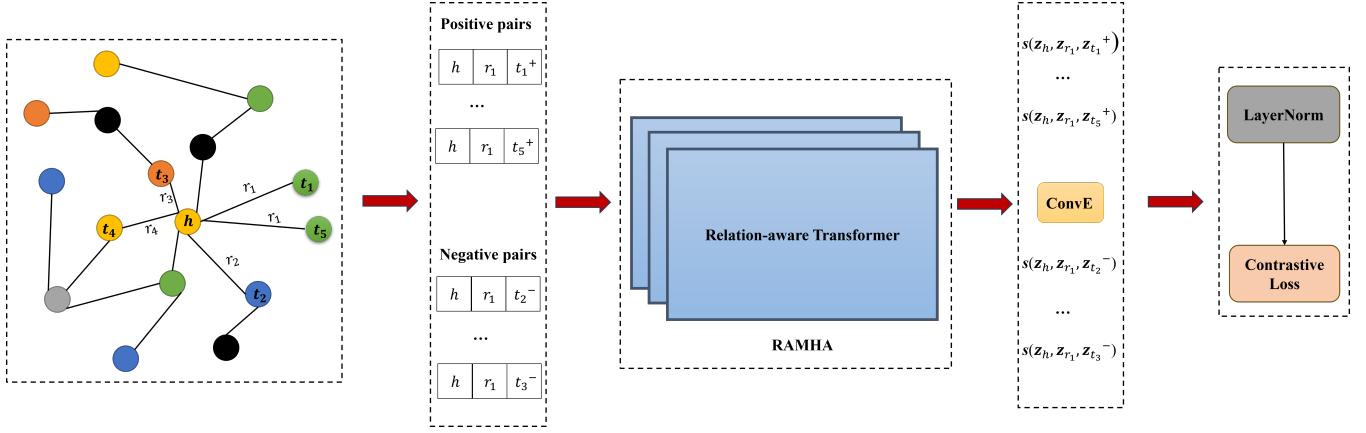


Figure 1: The overall framework of NC-KGE.

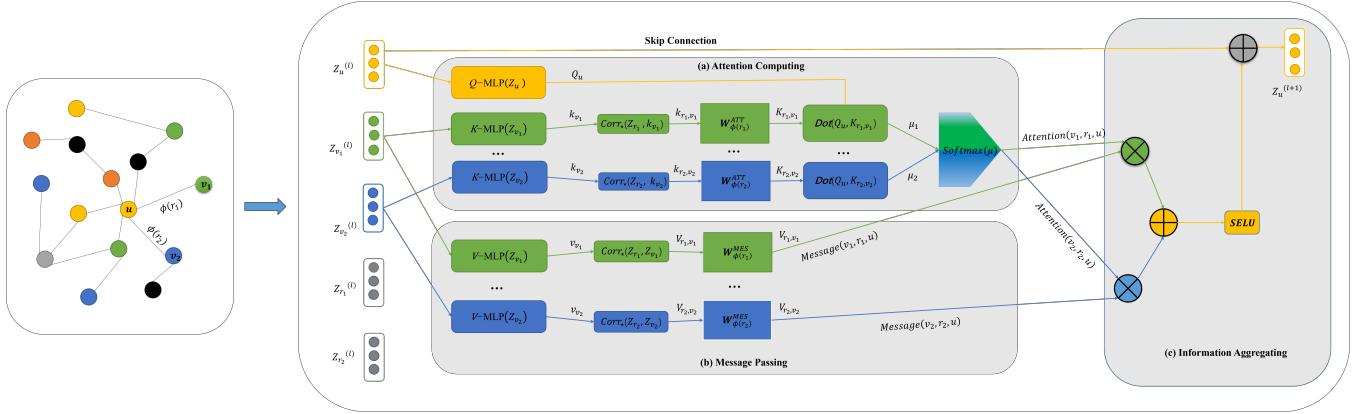


Figure 2: Relation-aware multi-head attention based KGE model.

$$v_u^{l,c} = \text{MLP}_{v,r,1}^{l,c}(z_u^l) \quad (7)$$

where  $q^{l,c}$  and  $k^{l,c}$  is the query vector and key vector for entities in multi-head attentions,  $\text{MLP}_{r,1}^{l,c}$  is a relation-type specific MLP without any bias terms,  $\mathcal{N}(u)$  represents the set of neighbors of entity  $e_u$ ,  $\mathcal{R}(u, w)$  denotes the set of relations between  $e_u$  and  $e_w$ ,  $\langle q, k \rangle = \exp\left(\frac{q^T k}{\sqrt{d}}\right)$ , and  $\star$  represents the circular-correlation mentioned in HolE [32].

The message passing information  $\text{Message } e_{u,r,v}^{l,c}$  is as follows:

$$\text{Message } e_{u,r,v}^{l+1,c} = \text{MLP}_{r,2}^{l,c}\left(v_t^{l,c} \star x_r^{l+1,c}\right) \quad (8)$$

where  $\text{MLP}_{r,2}^{l,c}$  is also an unbiased relation-type specific MLP as  $\text{MLP}_{r,1}^{l,c}$ .

### 4.3 Similarity scoring functions for relation predictions

Once embeddings of entities and relations are obtained, NC-KGE introduced a learnable similarity function to score the embeddings

of triplets in the knowledge graph as same as ConvE. It is a convolution over 2D shaped embeddings as formulated in Equation 9.

$$\psi(e_u, r, e_v) = f(\text{reshape}(f([\overline{e_u}, \bar{r}]) \otimes w)) * W \cdot e_v \quad (9)$$

Where  $\overline{e_u}$ ,  $\bar{r}$  denote a 2D reshaping of  $e_u$  and relation type  $r$ ,  $[\cdot]$  denotes a concatenation operation,  $\otimes w$  represents a 2D-convolutional layer with filters  $w$ ,  $f$  denotes a non-linear function,  $W$  is a linear transformation matrix, and  $\cdot$  represents the inner product operation. We need to note that NC-KGE can also integrate the classic statistic similarity measure functions [10, 38] as TransE, DisMult, ComplEx and SimplE as listed Table 1, but extensive experiments have shown that a learnable similarity function in Equation 9 outperforms these traditional similarity measures, as discussed in Section 5.4 Results.

### 4.4 Contrastive learning objective

NC-KGE aims to mining the hidden information between entities and relations by maximizing the similarity score between positive samples and minimizing it between negative ones. Thus, a classic contrastive training objective function [16] are introduced into

**Table 1: Examples of triple similarity measure functions proposed by TransE, DistMult, ComplEx, SimplE and ConvE. For triples  $(e_i, r_j, e_k)$ , we use  $e_i, r_j, e_k$  to represent the embedding of its components (in SimplE, they have two parts, which we use index to represent).  $\|\cdot\|_p$  represents the  $p$ -norm;  $\langle \cdot, \cdot \rangle$  is the generalized three-way dot product;  $\text{Re}(\cdot)$  is the real part of the complex number;  $\bar{e}_k$  is the complex conjugate of the complex-valued vector  $e_k$ .**

Model	Function
TransE [3]	$-\ e_i + r_j - e_k\ _p$
DistMult [53]	$\langle e_i, r_j, e_k \rangle$
ComplEx [42]	$\text{Re}(\langle e_i, r_j, \bar{e}_k \rangle)$
SimplE [20]	$\frac{1}{2} (\langle e_{i1}, r_{j1}, e_{k1} \rangle + \langle e_{i2}, r_{j2}, e_{k2} \rangle)$
ConvE [11]	$f(\text{vec}(f(\text{concat}(\bar{e}_i, \bar{r}_j) * \omega)) W) e_k$

NC-KGE for a given triplet  $(e_h, r, e_t)$  as shown in Equation 10.

$$\mathcal{L} = -\log \frac{\sum_{k=1}^{K^+} \exp(\tilde{S}(z_h, x_r, z_t^+)/\tau)}{\sum_{k=1}^{K^+} \exp(\tilde{S}(z_h, x_r, z_t^+)/\tau) + Q \sum_{k=1}^{K^-} \exp(\tilde{S}(z_h, x_r, z_t^-)/\tau)} \quad (10)$$

where the construction  $K^+$  and  $K^-$  are the number of positive triple pairs and negative triple pairs, respectively,  $Q$  is a scaling weight for negative pairs,  $\tau$  is a temperature factor controlling the discriminable capability of KGE model to negative pairs. To avoid extreme temperature coefficients affecting contrastive learning effects, we adopt a simulated annealing strategy to adjust the temperature factor dynamically in the range of [0.1, 1.5] according to the MRR metrics. Additionally, we need to note that the similarity scores  $S$  were layer-normalized to compute the contrastive loss, denoted as  $\tilde{S}$ , to avoid the numerical overflow in exponential operation and underflow in logarithmic operations.

## 5 EXPERIMENTS

Two commonly used benchmark datasets FB15k-237 and WN18RR [10] for KGE methods are utilized to evaluate the performance of NC-KGE on relation-predictions. We also perform benchmarks on three Bio-medical specific datasets including PharmKG8k-28, DRKG17k-21 and BioKG72k-14, derived from PharmKG, Drug Repositioning Knowledge Graph, and BioKG datasets. In these benchmarks, we could further evaluate the performance of NC-KGE on relation-predictions between different bio-meaningful entities. The detailed description of datasets is shown in Table 8 of Appendix A.

### 5.1 Baselines

The baselines are composed of three types: Translation Model, Semantic Matching Model and GNN-based Model. Translation Models include TransE [3], RotatE [39] and PaiRE [5]. Semantic Matching Models include DistMult [53], ComplEx [42], TuckER [1], ConvE [11], InteractE [43] and PROCRUSTES [35]. GNN-based Models include HyConvE [46], MEKER [6], RAGAT [27], HRGAT [61], RGAT [37], KBGAT [31], A2N [2], SACN [38], CompGCN [44] and SE-GNN [21].

### 5.2 Task and Evaluation

Relation Prediction, also termed Link Prediction, aims at inferring missing facts based on the facts in Knowledge Graph. Similar to question answering, we assess the quality of relation-predictions using the following ranking task: for all triplets  $(e_h, r, e_t)$  in both training sets and test sets, (1) we hidden the tail entity  $e_t$ . (2) we compute the similarity scores  $S(e_h, r, e_t^*)$  for all  $e_t^* \in \mathcal{E}$  as discussed in Section 4.3. (3) we sort values by decreasing order and (4) record the rank of the correct entity  $e_t$ . An identical process is repeated for prediction  $e_h$ .

Two kinds of metrics are introduced for evaluation, including the proportion of correct entities ranked in the top (such as top 1, 3, 10, denoted as Hits@1, Hits@3, Hits@10), and the mean reciprocal rank (MRR). Let  $r_{th}$  be the rank of the correct triplet  $t = (e_h, r, e_t)$  among all possible triples when hidden the head entity, while  $r_{tt}$  represent its rank when hidden the tail entity. MRR is the average of reciprocal rank of a set of correct fact triplets  $\mathcal{T}$  as shown in Equation 11.

$$MRR = \frac{1}{2|\mathcal{T}|} \sum_{t \in \mathcal{T}} \frac{1}{r_{th}} + \frac{1}{r_{tt}} \quad (11)$$

The Hits at 1 metric (H@1) is obtained by counting the number of times the correct triple appears at position 1. The H@3 and H@10 are computed similarly, considering the first 3 and 10 positions, respectively.

### 5.3 Experimental Setup

In our benchmark experiments, we employed a KGE encoder in NC-KGE comprising two layers of RAMHA. The RAMHA model utilized 10 heads, and the hidden layer dimension for all MLP was 200. For Equation 10, the positive pair number for each fact triplet  $K^+$  was set to 1 while all the negative pairs are used in node-based contrastive training. To stabilize the training phase and prevent overfitting, we introduced a dropout rate of 0.2 and applied batch layer normalization between the RAMHA layers. The AdamW [29] optimizer and cosine annealing learning rate scheduler [28] is used in training. The patience of simulated annealing strategy for temperature factor is set to 50 according to MRR metric.

### 5.4 Results

Here, we first evaluate the performance of NC-KGE on general knowledge graph FB15k-237 and WN18RR, the benchmark results are as shown in Table 2. Compared with 17 baselines, NC-KGE outperforms all the translation distance-based and semantic matching-based models, and is competitive with the SOTA method, SE-GNN. NC-KGE obtains obvious improvement compared to CompGCN, a typical GNN-based model, indicating node-based contrastive learning and RAMHA sufficient for better knowledge graph embedding.

For bio-medical specific knowledge graphs, NC-KGE outperforms all the baselines as shown in Table 3. In PharmKG8k-28, the knowledge entity can be clearly categorized into four bio-medical meaningful types: gene, disease and chemical compounds. Thus, the relation prediction can be divided into 6 types according to the types of head and tail entity of triplet. Then, we further evaluate

<sup>1</sup><https://github.com/DeepGraphLearning/>

**Table 2: Model reports on FB15k-237 and WN18RR test set. The best results are in bold.  $\dagger$  denotes that results are from the published paper. Other results are from SE-GNN [21].**

Dataset		FB15k-237				WN18RR			
Task		Link Prediction				Link Prediction			
Metric →	Model ↓	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
PaiRE		0.351	0.256	0.387	0.544	-	-	-	-
DistMult		0.241	0.155	0.263	0.419	0.430	0.390	0.440	0.490
ComplEx		0.247	0.158	0.275	0.428	0.440	0.410	0.460	0.510
Tucker		0.358	0.266	0.394	0.544	0.470	0.443	0.482	0.526
ConvE		0.325	0.237	0.356	0.501	0.430	0.400	0.440	0.520
InteractE		0.354	0.263	-	0.535	0.463	0.430	-	0.528
PROCRUSTES		0.345	0.249	0.379	0.541	0.474	0.421	0.502	0.569
HyConvE $\dagger$		0.339	0.212	-	0.458	0.461	0.432	-	0.534
MEKER $\dagger$		0.359	0.268	0.392	0.539	0.477	0.437	0.488	0.545
R-GCN		0.248	0.151	-	0.417	-	-	-	-
KBGAT		0.157	-	-	0.331	0.412	-	-	0.554
A2N		0.317	0.232	0.348	0.486	0.450	0.420	0.460	0.510
SACN		0.350	0.260	0.390	0.540	0.470	0.430	0.480	0.540
CompGCN		0.355	0.264	0.390	0.535	0.479	0.443	0.494	0.546
SE-GNN		0.365	0.271	<b>0.399</b>	<b>0.549</b>	0.484	0.446	<b>0.509</b>	<b>0.572</b>
NC-KGE (our)		<b>0.366</b>	<b>0.273</b>	0.392	0.542	<b>0.486</b>	<b>0.447</b>	0.499	0.556

**Table 3: The results of models on DRKG17k-21, BioKG72k-14 and PharmKG8k-28 Datasets.  $\dagger$  denotes that we reproduce the results using the codes <sup>1</sup>. For other results, we implemented the official codes.**

Dataset		DRKG17k-21				BioKG72k-14				PharmKG8k-28			
Task		Link Prediction				Link Prediction				Link Prediction			
Metric →	Model ↓	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE		0.321	0.035	0.558	0.744	0.116	0.026	0.149	0.276	0.116	0.038	0.127	0.269
DistMult		0.240	0.175	0.254	0.371	0.045	0.021	0.041	0.083	0.218	0.152	0.237	0.335
ComplEx		0.099	0.036	0.087	0.227	0.111	0.073	0.118	0.174	0.124	0.064	0.128	0.244
TruckER		0.460	0.411	0.535	0.557	0.226	0.174	0.237	0.327	0.182	0.103	0.202	0.336
HRGAT		0.540	0.483	0.582	0.720	0.103	0.061	0.105	0.185	0.134	0.063	0.144	0.271
SACN		0.487	0.393	0.534	0.665	0.179	0.118	0.192	0.299	0.156	0.085	0.170	0.296
CompGCN		0.562	0.466	0.619	0.739	0.221	0.170	0.230	0.321	0.193	0.110	0.216	0.352
SE-GNN		0.575	0.481	0.631	0.746	0.237	0.183	0.248	0.343	0.206	0.120	0.232	0.374
NC-KGE (our)		<b>0.590</b>	<b>0.505</b>	<b>0.637</b>	<b>0.747</b>	<b>0.240</b>	<b>0.185</b>	<b>0.256</b>	<b>0.344</b>	<b>0.228</b>	<b>0.145</b>	<b>0.252</b>	<b>0.390</b>

the relation prediction performance of NC-KGE on these sub-domains. As shown in Table 4, NC-KGE give obviously better results for Chemical-Chemical and Disease-Disease relation predictions, such as drug combination relationship and disease complications, which MRR is 0.483 and 0.471, respectively. Both of these sub-domains are belonging to “Interactions” category in PharmKG. Similarly, NC-KGE also achieve the best results, MRR is 0.640 on

Chemical-Chemical relation predictions, since no Chemical-Disease and Disease-Disease relations exists as shown in Table 5.

Additionally, NC-KGE proves to be more efficient for training, as illustrated in Figure 3, Figure 5, and Table 7. The Mean Reciprocal Rank (MRR) on test sets converges faster with NC-KGE than with any other methods. When we substitute the node-based contrastive objective with other functions like BCELoss (Binary CrossEntropy), MPLoss (Maximize Positive Sample Softmax Loss)

and MRLoss (Margin Ranking Loss), both accuracy and efficiency decrease. These findings highlight the efficiency and potential applications of NC-KGE in biomedical knowledge graphs.

**Table 4: Experiments on the PharmKG8k-28 subdataset**

Dataset		PharmKG8k-28		
Task		Link Prediction		
Metric →	SubDataset ↓	MRR	H@1	H@3
Gene-Chemical		0.160	0.093	0.169
Chemical-Disease		0.153	0.082	0.160
Disease-Disease		0.471	<b>0.377</b>	0.507
Gene-Gene		0.244	0.156	0.275
Gene-Disease		0.175	0.103	0.186
Chemical-Chemical		<b>0.483</b>	0.367	<b>0.537</b>
ALL		0.228	0.145	0.252
				0.390

**Table 5: Experiments on the DRKG17k-21 subdataset**

Dataset		DRKG17k-21		
Task		Link Prediction		
Metric →	SubDataset ↓	MRR	H@1	H@3
Gene-Chemical		0.333	0.258	0.363
Gene-Gene		0.243	0.160	0.272
Gene-Disease		0.222	0.138	0.223
Chemical-Chemical		<b>0.640</b>	<b>0.549</b>	<b>0.693</b>
ALL		0.590	0.505	0.637
				0.747

**Table 6: Results of baselines + NCL.**

Dataset		PharmKG8k-28		
Task		Link Prediction		
Metric →	Model ↓	MRR	H@1	H@3
CompGCN		0.193	0.110	0.216
SE-GNN		0.206	0.120	0.232
CompGCN + NCL		<b>0.194</b>	<b>0.112</b>	<b>0.217</b>
SE-GNN + NCL		<b>0.212</b>	<b>0.127</b>	<b>0.237</b>
				0.352
				0.374
				<b>0.354</b>
				0.377

## 5.5 Discussion

**The Universality of NC-KGE.** NC-KGE is a straightforward yet effective method that integrates contrastive learning with non-contrastive Knowledge Graph Embedding (KGE) techniques to enhance performance on biomedical knowledge graphs. In our study, we have integrated NC-KGE with existing models like CompGCN and SE-GNN, using PharmKG8k-28 as a representative example.

**Table 7: KGE with diffirent loss functions.**

Dataset		PharmKG8k-28			
Task		Link Prediction			
Metric →	Model ↓	MRR	H@1	H@3	H@10
KGE + BCELoss		0.222	0.141	0.244	0.379
KGE + MPLoss		0.205	0.124	0.226	0.363
KGE + MRLoss		0.163	0.082	0.176	0.326
<b>KGE+ NCloss</b>		<b>0.228</b>	<b>0.145</b>	<b>0.252</b>	<b>0.390</b>

The evaluation metrics for both CompGCN and SE-GNN demonstrated improvements, as outlined in Table 6. Furthermore, the evolution of Mean Reciprocal Rank (MRR) on the test set is depicted in Figure 4. NC-KGE not only enhances the accuracy of the models but also accelerates the convergence of CompGCN and SE-GNN. This suggests a general improvement resulting from node-based contrastive learning.

**The performance of NC-KGE with different similarity measure functions.** In this paper, we consider the effect of different similarity measure functions on NC-KGE. As shown in Table 9 of Appendix A, we consider 3 common classes of metric functions, including TransE based on translation, DistMult, SimplE and ComplEx based on tensor decomposition, ConvE with parameters based on convolutional networks. Among the five similarity measure functions, NC-KGE combined with ConvE has the best performance.

**The performance of NC-KGE under various temperature coefficients.** In this section, we discuss the temperature hyper-parameter for NC-KGE. From previous experience, smaller temperature coefficient usually makes the model tend to distinguish more difficult negative samples, and the learned embedding will be smoother. However, this may also lead to the model paying too much attention to difficult negative samples, resulting in insufficient attention to common negative samples, which is not conducive to learning good embeddings. As shown in Table 9 of Appendix A, we show the temperature coefficient is set to 0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 1.0, 1.1, 1.2, 1.3, 1.5 and dynamic adjustment. Among them, the temperature coefficient based on the dynamic adjustment strategy achieves better performance. However, when the temperature coefficient is set to a certain value, the performance of the model does not change significantly.

**The performance of NC-KGE with different number of negative samples.** The number of negative samples is always an important discussion topic for self-supervised learning or unsupervised learning. In general, more negative samples will allow the model to see more diverse data points, which will improve performance. However, more negative samples also mean higher sampling cost. So we need to make a trade-off between the two. Here, we also show the performance of NC-KGE with different number of negative samples. As shown in Table 9 of Appendix A, we find that NC-KGE achieves a decent performance when the number of negative samples for contrastive learning is 1000, and it performs best when the number is the whole.

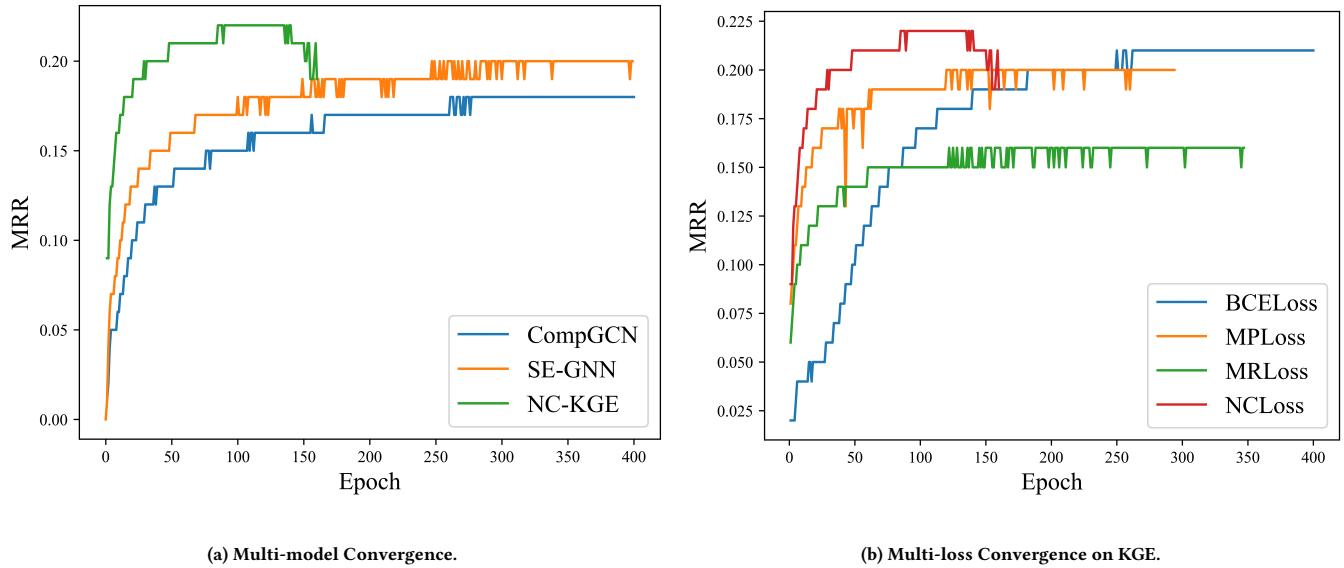


Figure 3: Performance of the models on PharmKG8k-28.

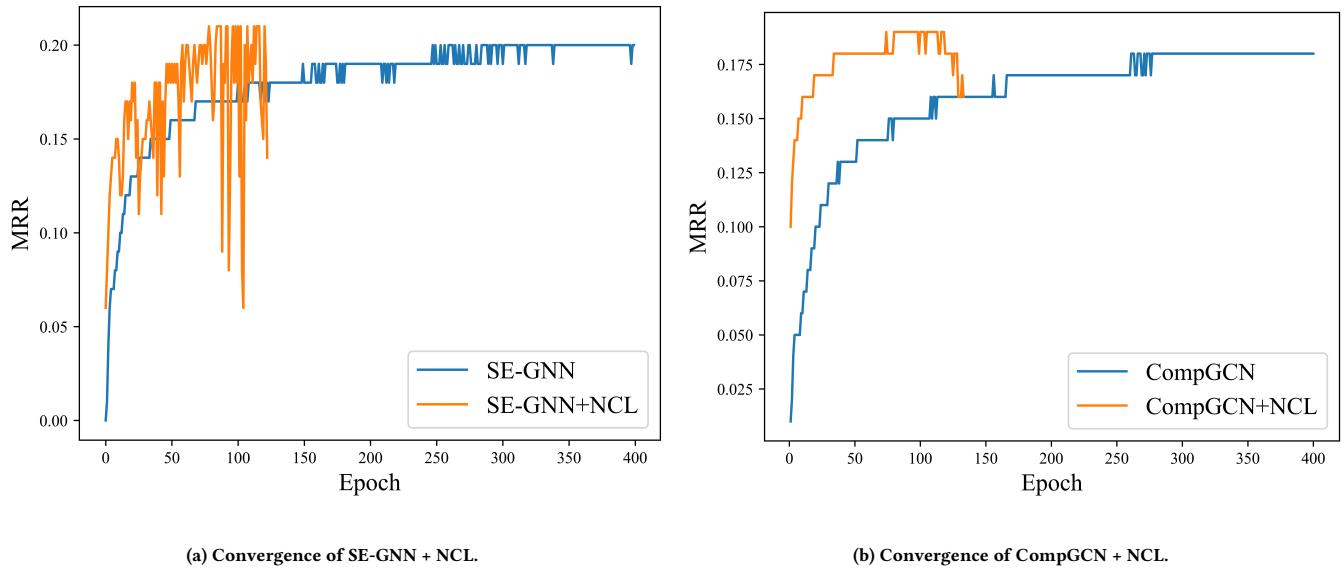


Figure 4: Performance of the baseline models on PharmKG8k-28.

## 6 CONCLUSION

In this study, we introduced a node-based contrastive knowledge graph embedding method called NC-KGE for relation prediction in biomedical knowledge graphs. NC-KGE is a simple and versatile approach that can be seamlessly integrated into existing KGE methods to enhance training convergence and improve overall performance. This distinctive feature allows NC-KGE to fully leverage advancements in state-of-the-art models. In the future work, we will expand the application of NC-KGE to diverse downstream tasks, including

but not limited to knowledge-guided molecular property prediction and the prediction of higher-order biomedical n-ary relationships. By leveraging semantically rich pre-trained node and relation embeddings, we aim to maximize the potential of NC-KGE in these domains.

## ACKNOWLEDGMENTS

Thanks to the readers.

## REFERENCES

- [1] Ivana Balažević, Carl Allen, and Timothy M Hospedales. 2019. Tucker: Tensor factorization for knowledge graph completion. *arXiv preprint arXiv:1901.09590* (2019).
- [2] Trudit Bansal, Da-Cheng Juan, Sujith Ravi, and Andrew McCallum. 2019. A2N: Attending to neighbors for knowledge graph inference. In *Proceedings of the 57th annual meeting of the association for computational linguistics*. 4387–4392.
- [3] 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 (2013).
- [4] 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*, Vol. 35. 6894–6902.
- [5] Linlin Chao, Jianshan He, Taifeng Wang, and Wei Chu. 2020. Pairre: Knowledge graph embeddings via paired relation vectors. *arXiv preprint arXiv:2011.03798* (2020).
- [6] Viktoriya Chekalina, Anton Razhigaev, Albert Sayapin, and Alexander Panchenko. 2022. MEKER: Memory Efficient Knowledge Embedding Representation for Link Prediction and Question Answering. *arXiv preprint arXiv:2204.10629* (2022).
- [7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PMLR, 1597–1607.
- [8] Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1. IEEE, 539–546.
- [9] Shivani Choudhary, Tarun Luthra, Ashima Mittal, and Rajat Singh. 2021. A survey of knowledge graph embedding and their applications. *arXiv preprint arXiv:2107.07842* (2021).
- [10] Daniel Daza, Michael Cochez, and Paul Groth. 2021. Inductive Entity Representations from Text via Link Prediction. In *Proceedings of the Web Conference 2021*. 798–808.
- [11] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Thirty-second AAAI conference on artificial intelligence*.
- [12] Alberto García-Durán, Antoine Bordes, and Nicolas Usunier. 2014. Effective blending of two and three-way interactions for modeling multi-relational data. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15–19, 2014. Proceedings, Part I*. Springer, 434–449.
- [13] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. 2016. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 conference on empirical methods in natural language processing*. 192–202.
- [14] Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 297–304.
- [15] Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2. IEEE, 1735–1742.
- [16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9729–9738.
- [17] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. 2018. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670* (2018).
- [18] Hongren Huang, Chen Li, Xutan Peng, Lifang He, Shu Guo, Hao Peng, Lihong Wang, and Jianxin Li. 2022. Cross-knowledge-graph entity alignment via relation prediction. *Knowledge-Based Systems* 240 (2022), 107813.
- [19] Vassilis N. Ioannidis, Xiang Song, Saurav Manchanda, Mufei Li, Xiaoqin Pan, Da Zheng, Xia Ning, Xiangxiang Zeng, and George Karypis. 2020. DRKG - Drug Repurposing Knowledge Graph for Covid-19. <https://github.com/gnn4dr/DRKG/>.
- [20] Seyed Mehran Kazemi and David Poole. 2018. Simple embedding for link prediction in knowledge graphs. *arXiv preprint arXiv:1802.04868* (2018).
- [21] Ren Li, Yanan Cao, Qiannan Zhu, Guanqun Bi, Fang Fang, Yi Liu, and Qian Li. 2022. How does knowledge graph embedding extrapolate to unseen data: a semantic evidence view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 5781–5791.
- [22] Rui Li, Jianan Zhao, Chaozhuo Li, Di He, Yiqi Wang, Yuming Liu, Hao Sun, Senzhang Wang, Weiwei Deng, Yanming Shen, et al. 2022. House: Knowledge graph embedding with householder parameterization. In *International Conference on Machine Learning*. PMLR, 13209–13224.
- [23] Huidong Liang, Xingjian Du, Bilei Zhu, Zejun Ma, Ke Chen, and Junbin Gao. 2023. Graph contrastive learning with implicit augmentations. *Neural Networks* 163 (2023), 156–164.
- [24] Ke Liang, Yue Liu, Sihang Zhou, Wenxuan Tu, Yi Wen, Xihong Yang, Xiangjun Dong, and Xinwang Liu. 2023. Knowledge Graph Contrastive Learning Based on Relation-Symmetrical Structure. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [25] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*.
- [26] Hanxiao Liu, Yuxin Wu, and Yiming Yang. 2017. Analogical inference for multi-relational embeddings. In *International conference on machine learning*. PMLR, 2168–2178.
- [27] Xiyang Liu, Huobin Tan, Qinghong Chen, and Guangyan Lin. 2021. RAGAT: Relation aware graph attention network for knowledge graph completion. *IEEE Access* 9 (2021), 20840–20849.
- [28] Ilya Loshchilov and Frank Hutter. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983* (2016).
- [29] Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
- [30] Ishan Misra and Laurens van der Maaten. 2020. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6707–6717.
- [31] Deepak Nathani, Jatin Chauhan, Charu Sharma, and Manohar Kaul. 2019. Learning attention-based embeddings for relation prediction in knowledge graphs. *arXiv preprint arXiv:1906.01195* (2019).
- [32] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016. Holographic embeddings of knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.
- [33] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Icml*.
- [34] Matteo Palmonari and Pasquale Minervini. 2020. Knowledge graph embeddings and explainable AI. *Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges* 47 (2020), 49.
- [35] Xutan Peng, Guanyi Chen, Chenghua Lin, and Mark Stevenson. 2021. Highly efficient knowledge graph embedding learning with orthogonal procrustes analysis. *arXiv preprint arXiv:2104.04676* (2021).
- [36] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. 2020. Contrastive learning with hard negative samples. *arXiv preprint arXiv:2010.04592* (2020).
- [37] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*. Springer, 593–607.
- [38] Chao Shang, Yun Tang, Jing Huang, Jinbo Bi, Xiaodong He, and Bowen Zhou. 2019. End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3060–3067.
- [39] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197* (2019).
- [40] Yonglong Tian, Dilip Krishnan, and Phillip Isola. 2020. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*. Springer, 776–794.
- [41] 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*. 57–66.
- [42] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*. PMLR, 2071–2080.
- [43] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, Nilesh Agrawal, and Partha Talukdar. 2020. Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3009–3016.
- [44] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. 2019. Composition-based multi-relational graph convolutional networks. *arXiv preprint arXiv:1911.03082* (2019).
- [45] Brian Walsh, Sameh K. Mohamed, and Vit Novák. 2020. BioKG: A Knowledge Graph for Relational Learning On Biological Data. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19–23, 2020*, Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux (Eds.). ACM, 3173–3180. <https://doi.org/10.1145/3340531.3412776>
- [46] Chenxu Wang, Xin Wang, Zhao Li, Zirui Chen, and Jianxin Li. 2023. HyConvE: A Novel Embedding Model for Knowledge Hypergraph Link Prediction with Convolutional Neural Networks. In *Proceedings of the ACM Web Conference 2023*. 188–198.
- [47] Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. 2022. SimKGC: Simple contrastive knowledge graph completion with pre-trained language models. *arXiv preprint arXiv:2203.02167* (2022).

- [48] Xiaolong Wang and Abhinav Gupta. 2015. Unsupervised learning of visual representations using videos. In *Proceedings of the IEEE international conference on computer vision*. 2794–2802.
- [49] Zhigang Wang, Juanzi Li, Zhiyuan Liu, and Jie Tang. 2016. Text-enhanced representation learning for knowledge graph. In *Proceedings of International joint conference on artificial intelligent (IJCAI)*. 4–17.
- [50] 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*, Vol. 28.
- [51] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. 2018. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3733–3742.
- [52] Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. 2016. Representation learning of knowledge graphs with entity descriptions. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 30.
- [53] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575* (2014).
- [54] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-BERT: BERT for knowledge graph completion. *arXiv preprint arXiv:1909.03193* (2019).
- [55] Yihang Yin, Qingzhong Wang, Siyu Huang, Haoyi Xiong, and Xiang Zhang. 2022. Autogcl: Automated graph contrastive learning via learnable view generators. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 36. 8892–8900.
- [56] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph contrastive learning with augmentations. *Advances in neural information processing systems* 33 (2020), 5812–5823.
- [57] Donghan Yu, Yiming Yang, Ruohong Zhang, and Yuxin Wu. 2021. Knowledge embedding based graph convolutional network. In *Proceedings of the Web Conference 2021*. 1619–1628.
- [58] Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. 2019. Quaternion knowledge graph embeddings. *Advances in neural information processing systems* 32 (2019).
- [59] Zhanqui Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020. Learning hierarchy-aware knowledge graph embeddings for link prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3065–3072.
- [60] Zhao Zhang, Fuzhen Zhuang, Hengshu Zhu, Zhi-Ping Shi, Hui Xiong, and Qing He. 2020. Relational Graph Neural Network with Hierarchical Attention for Knowledge Graph Completion. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020*. AAAI Press, 9612–9619. <https://doi.org/10.1609/aaai.v34i05.6508>
- [61] Shuangjia Zheng, Jiahua Rao, Ying Song, Jixian Zhang, Xianglu Xiao, Evandro Fei Fang, Yuedong Yang, and Zhangming Niu. 2020. PharmKG: a dedicated knowledge graph benchmark for biomedical data mining. *Briefings in Bioinformatics* 22, 4 (12 2020), bbaa344. <https://doi.org/10.1093/bib/bbaa344> arXiv:<https://academic.oup.com/bib/article-pdf/22/4/bbaa344/39139048/bbaa344.pdf>
- [62] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. 2020. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131* (2020).

## A APPENDIX

The detailed description of datasets is as follows:

- **FB15k-237** [41] dataset contains knowledge base relation triples and textual mentions of Freebase node pairs, as used in the work published in [41]. The knowledge base triples are a subset of the FB15K [3], originally derived from Freebase. The inverse relations are removed in FB15k-237. To obtain node descriptions and types, we employ the datasets made available by [10].
- **WN18RR** [11] is created from WN18 [3], which is a subset of WordNet. WN18 consists of 18 relations and 40,943 nodes. However, many text triples obtained by inverting triples from the training set. Thus WN18RR dataset [11] is created to ensure that the evaluation dataset does not have inverse relation test leakage. To obtain node descriptions and types, we employ the datasets made available by [10].
- **PharmKG8k-28** is built based on PharmKG dataset [61]. PharmKG is a multi-relational, attributed biomedical KG,

composed of more than 500 000 individual interconnections between genes, drugs and diseases, with 29 relation types over a vocabulary of ~ 8000 disambiguated nodes. Each node in PharmKG is attached with heterogeneous, domain-specific information obtained from multi-omics data, i.e. gene expression, chemical structure and disease word embedding, while preserving the semantic and biomedical features. We obtained PharmKG8k-28 dataset after simple deduplication cleaning of PharmKG.

- **DRKG17k-21** is built based on Drug Repositioning Knowledge Graph (DRKG) [19]. DRKG is a comprehensive biological knowledge graph relating genes, compounds, diseases, biological processes, side effects and symptoms. DRKG includes information from six existing databases including DrugBank, Hetionet, GNBR, String, IntAct and DGIdb, and data collected from recent publications particularly related to Covid19.
- **BioKG72k-14** is built based on BioKG dataset [45]. BioKG is a new more standardised and reproducible biological knowledge graph which provides a compilation of curated relational data from open biological databases in a unified format with common, interlinked identifiers. BioKG can be used to train and assess the relational learning models in various tasks related to pathway and drug discovery.

**Table 8: Statistics of the datasets.**

	<b>FB15k-237</b>	<b>WN18RR</b>	<b>PharmKG8k-28</b>	<b>DRKG17k-21</b>	<b>BioKG72k-14</b>
Relations	237	11	28	21	14
Nodes	14,541	40,943	7,599	17,508	72,569
Triples	310,116	93,003	480,500	587,425	1,157,739
Training					
Nodes	14,505	40,559	7,520	16,252	70,108
Triples	272,115	86,835	384,880	470,526	927,348
Validation					
Nodes	9,809	5,173	6,637	8,186	43,806
Triples	17,535	3,034	47,570	58,156	114,617
Test					
Nodes	10,348	5,323	6,628	8,242	43,989
Triples	20,466	3,134	48,050	58,743	115,774

**Table 9: Ablation experiments on the PharmKG8k-28 and FB15k-237 datasets. 'Neg100' means that the number of negative samples is 100. 'Tem0.1' means that the temperature coefficient is 0.1. 'DynamicTem' means to adopt the dynamic temperature coefficient strategy.**

<b>Dataset</b>	<b>PharmKG8k-28</b>				<b>Dataset</b>	<b>FB15k-237</b>			
<b>Task</b>	<b>Link Prediction</b>				<b>Task</b>	<b>Link Prediction</b>			
<b>Metric → Model ↓</b>	MRR	H@1	H@3	H@10	<b>Metric → Model ↓</b>	MRR	H@1	H@3	H@10
NC-KGE + TransE	0.140	0.077	0.147	0.263	NC-KGE + TransE	0.321	0.235	0.352	0.494
NC-KGE + DistMult	0.169	0.097	0.177	0.317	NC-KGE + DistMult	0.322	0.235	0.353	0.499
NC-KGE + SimplE	0.200	0.126	0.217	0.344	NC-KGE + SimplE	0.318	0.232	0.347	0.494
NC-KGE + ComplEx	0.205	0.134	0.220	0.345	NC-KGE + ComplEx	0.323	0.236	0.354	0.500
<b>NC-KGE +ConvE</b>	<b>0.228</b>	<b>0.145</b>	<b>0.252</b>	<b>0.390</b>	<b>NC-KGE +ConvE</b>	<b>0.362</b>	<b>0.270</b>	<b>0.392</b>	<b>0.542</b>
NC-KGE + Neg100	0.214	0.126	0.241	0.385	NC-KGE + Neg100	0.342	0.248	0.377	0.531
NC-KGE + Neg500	0.214	0.126	0.241	0.384	NC-KGE + Neg500	0.352	0.259	0.387	0.538
NC-KGE + Neg1000	0.220	0.133	0.246	0.386	NC-KGE + Neg1000	0.354	0.258	0.388	0.539
<b>NC-KGE +NegAll</b>	<b>0.228</b>	<b>0.145</b>	<b>0.252</b>	<b>0.390</b>	<b>NC-KGE +NegAll</b>	<b>0.362</b>	<b>0.270</b>	<b>0.392</b>	<b>0.542</b>
NC-KGE + Tem0.1	0.218	0.137	0.239	0.376	NC-KGE + Tem0.1	0.347	0.255	0.382	0.531
NC-KGE + Tem0.2	0.220	0.138	0.242	0.379	NC-KGE + Tem0.2	0.348	0.255	0.384	0.530
NC-KGE + Tem0.3	0.217	0.135	0.240	0.377	NC-KGE + Tem0.3	0.348	0.258	0.381	0.528
NC-KGE + Tem0.5	0.220	0.138	0.242	0.379	NC-KGE + Tem0.5	0.348	0.257	0.381	0.529
NC-KGE + Tem0.7	0.225	0.142	0.247	0.386	NC-KGE + Tem0.7	0.348	0.254	0.383	0.534
NC-KGE + Tem0.8	0.223	0.140	0.247	0.383	NC-KGE + Tem0.8	0.347	0.256	0.381	0.531
NC-KGE + Tem1.0	0.225	0.142	0.248	0.386	NC-KGE + Tem1.0	0.349	0.257	0.382	0.535
NC-KGE + Tem1.1	0.225	0.143	0.248	0.386	NC-KGE + Tem1.1	0.349	0.256	0.383	0.536
NC-KGE + Tem1.2	0.227	0.145	0.250	0.384	NC-KGE + Tem1.2	0.348	0.258	0.384	0.537
NC-KGE + Tem1.3	0.225	0.142	0.250	0.386	NC-KGE + Tem1.3	0.348	0.259	0.385	0.536
NC-KGE + Tem1.5	0.225	0.142	0.249	0.386	NC-KGE + Tem1.5	0.349	0.258	0.386	0.536
<b>NC-KGE +DynamicTem</b>	<b>0.228</b>	<b>0.145</b>	<b>0.252</b>	<b>0.390</b>	<b>NC-KGE +DynamicTem</b>	<b>0.362</b>	<b>0.270</b>	<b>0.392</b>	<b>0.542</b>

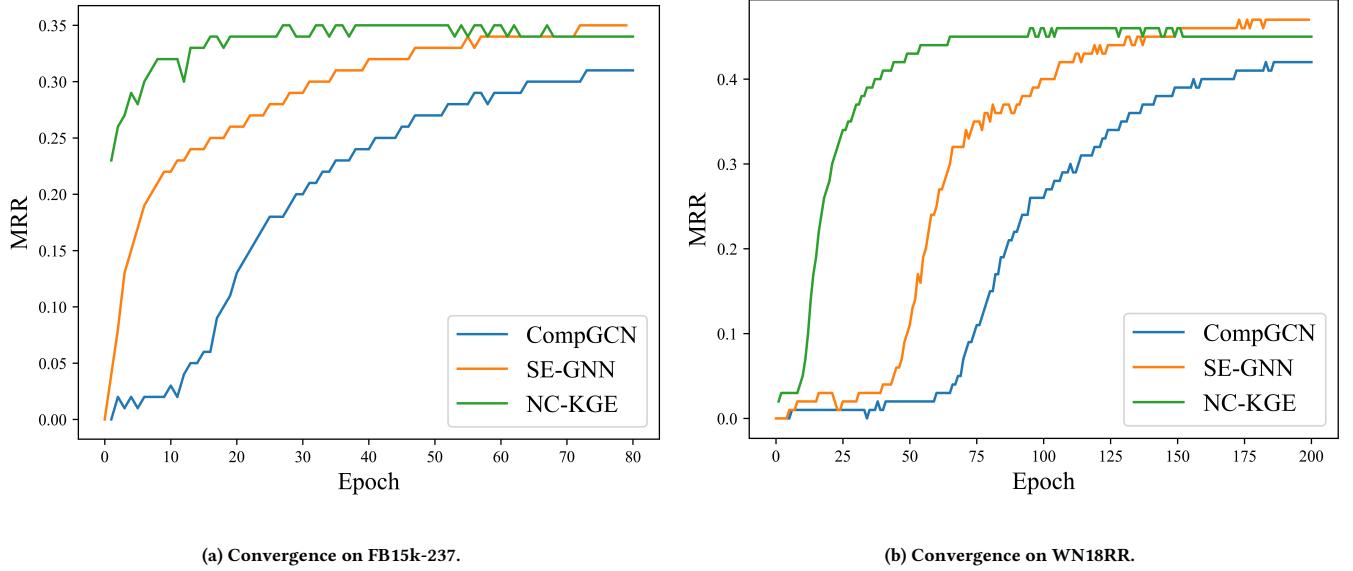


Figure 5: Performance of the models on FB15k-237 and WN18RR.

Table 10: Test results of the models on subsets of PharmKG8k-28

Model	SE-GNN				CompGCN				NC-KGE			
Task	Link Prediction				Link Prediction				Link Prediction			
Metric SubDataset	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
Gene-Gene	0.192	0.096	0.226	0.373	0.186	0.095	0.218	0.356	0.244	0.156	0.275	0.413
Gene-Disease	0.182	0.111	0.192	0.324	0.172	0.102	0.183	0.313	0.175	0.103	0.186	0.320
Chemical-Disease	0.160	0.091	0.162	0.296	0.151	0.084	0.155	0.283	0.153	0.082	0.160	0.291
Gene-Chemical	0.151	0.085	0.158	0.286	0.143	0.080	0.148	0.268	0.160	0.093	0.169	0.296
Chemical-Chemical	0.462	0.347	0.511	0.691	0.437	0.327	0.484	0.659	0.483	0.367	0.537	0.713
Disease-Disease	0.418	0.317	0.463	0.619	0.400	0.295	0.440	0.623	0.471	0.377	0.507	0.653
ALL	0.206	0.120	0.232	0.374	0.193	0.110	0.216	0.352	0.228	0.145	0.252	0.390

Table 11: Test results of the models on subsets of DRKG17k-21

Model	SE-GNN				CompGCN				NC-KGE			
Task	Link Prediction				Link Prediction				Link Prediction			
Metric SubDataset	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
Chemical-Chemical	0.637	0.540	0.709	0.810	0.632	0.532	0.696	0.815	0.640	0.549	0.693	0.811
Gene-Chemical	0.326	0.259	0.360	0.474	0.321	0.245	0.354	0.478	0.333	0.258	0.363	0.481
Gene-Gene	0.207	0.118	0.234	0.373	0.199	0.118	0.219	0.360	0.243	0.160	0.272	0.400
Gene-Disease	0.241	0.149	0.237	0.447	0.246	0.130	0.237	0.404	0.222	0.138	0.234	0.447
ALL	0.575	0.481	0.631	0.746	0.562	0.466	0.619	0.739	0.590	0.505	0.637	0.747

**Table 12: Z-test of the models on subsets of PharmKG9k-28**

SubData	Dataset		PharmKG8k-28
	Task	Link Prediction	
		Metric	Model
Gene-Gene	NC-KGE, SE-GNN	0.000	
	NC-KGE, CompCCN	0.000	
	CompCCN, SE-GNN	0.015	
Gene-Disease	NC-KGE, SE-GNN	0.048	
	NC-KGE, CompCCN	0.329	
	CompCCN, SE-GNN	0.003	
Chemical-Disease	NC-KGE, SE-GNN	0.130	
	NC-KGE, CompCCN	0.621	
	CompCCN, SE-GNN	0.043	
Gene-Chemical	NC-KGE, SE-GNN	0.117	
	NC-KGE, CompCCN	0.002	
	CompCCN, SE-GNN	0.121	
Chemical-Chemical	NC-KGE, SE-GNN	0.017	
	NC-KGE, CompCCN	0.000	
	CompCCN, SE-GNN	0.006	
Disease-Disease	NC-KGE, SE-GNN	0.234	
	NC-KGE, CompCCN	0.102	
	CompCCN, SE-GNN	0.658	

**Table 13: Z-test of the models on subsets of DRKG17k-21**

SubData	Dataset		DRKG17-21
	Task	Link Prediction	
		Metric	Model
Chemical-Chemical	NC-KGE, SE-GNN	0.000	
	NC-KGE, CompCCN	0.001	
	CompCCN, SE-GNN	0.000	
Gene-Chemical	NC-KGE, SE-GNN	0.063	
	NC-KGE, CompCCN	0.233	
	CompCCN, SE-GNN	0.509	
Gene-Gene	NC-KGE, SE-GNN	0.000	
	NC-KGE, CompCCN	0.000	
	CompCCN, SE-GNN	0.126	
Gene-Disease	NC-KGE, SE-GNN	0.638	
	NC-KGE, CompCCN	0.544	
	CompCCN, SE-GNN	0.884	