MuCoS: Efficient Drug–Target Discovery via Multi-Context-Aware Sampling in Knowledge Graphs

Haji Gul[1a], Abdul Ghani Naim[1b], Ajaz Ahmad Bhat*[1],

[1]School of Digital Science, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, Brunei Darussalam,
(23h1710[a], ghani.naim[b], ajaz.bhat*)@ubd.edu.bn

Abstract

Accurate prediction of drug–target interactions is critical for accelerating drug discovery. In this work, we frame drug–target prediction as a link prediction task on heterogeneous biomedical knowledge graphs (KG) that integrate drugs, proteins, diseases, pathways, and other relevant entities. Conventional KG embedding methods such as TransE and ComplEx-SE are hindered by their reliance on computationally intensive negative sampling and their limited generalization to unseen drug–target pairs. To address these challenges, we propose Multi-Context-Aware Sampling (MuCoS), a novel framework that prioritizes high-density neighbours to capture salient structural patterns and integrates these with contextual embeddings derived from BERT. By unifying structural and textual modalities and selectively sampling highly informative patterns, MuCoS circumvents the need for negative sampling, significantly reducing computational overhead while enhancing predictive accuracy for novel drug–target associations and drug targets. Extensive experiments on the KEGG50k and PharmKG-8k datasets demonstrate that MuCoS outperforms baselines, achieving up to a 13% improvement in MRR for general relation prediction on KEGG50k, a 22% improvement on PharmKG-8k, and a 6% gain in dedicated drug–target relation prediction on KEGG50k.

1 Introduction

Drug target discovery lies at the core of modern therapeutic development, enabling the identification of new biological targets, the prediction of non-target effects, and opportunities for drug repurposing — while significantly reducing experimental costs and accelerating translational timelines (Sachdev and Gupta, 2019). Recent computational advances leverage knowledge graphs (KGs) to integrate heterogeneous biomedical data (e.g., drugs, proteins, diseases, side effects, pathways) into unified networks where nodes represent entities and edges capture relationships, essentially framing discovery as a link prediction problem (Himmelstein et al., 2017). For example, KG's such as KEGG50k (Mohamed et al., 2019) PharmKG-8k (Zheng et al., 2021) and Hetionet (Himmelstein et al., 2017) provide comprehensive, structured representations of biological components and their intricate associations.

Biomedical KGC methods, however, face a critical trade-off: structural embedding methods such as ComplEx-SE (Mohamed et al., 2019) capture explicit drug-target relationships but fail to generalize to unseen entities like novel drugs due to rigid geometric constraints. Conversely, graph neural approaches like NeoDTI Progeni (Liu et al., 2024) integrate probabilistic reasoning with GNNs for state-of-the-art drug-target prediction but remain unevaluated on relation-centric benchmarks like KEGG50k. Furthermore, none of these methods exploit the rich textual semantics embedded in biomedical triples (e.g., "DRUG X $\rightarrow$ DRUG-TARGET-GENE $\rightarrow$ GENE Z"), which could provide inductive signals for unseen entities by contextualizing relationships beyond structural adjacency.

We posit that PharmKG-8k's and KEGG50k's relational triples are inherently compatible with textual encoding strategies and therefore believe that we can leverage a language transformer model like BERT's bidirectional attention to jointly model the explicit relationships through syntactic patterns in entity-relation-entity chains. Moreover, we propose to exploit, (a) the rich contextual information inherent in the graph's structure such as neighbouring entities and relations associated with a given head entity and query relation, like GNNs do, and (b) associated features such as node degrees and connectivity that affect the performance of KG techniques(Cattaneo et al., 2024).

We therefore propose **MuCoS** (*Multi-Context-Aware Sampling*), a KG completion framework that overcomes these limitations by aggregating filtered

Proceedings of the 24th Workshop on Biomedical Language Processing (BioNLP 2025), pages 319–327

contextual information from adjacent entities and their relationships, and then integrating this semantically enriched context into a BERT model for better prediction of relationships and entities. In doing so, MuCoS advances drug target discovery in the following key ways:

- **Drug–Target Relation Prediction:** By leveraging optimized neighbouring contextual information around nodes and relations, MuCoS outperforms traditional models in predicting general and drug–target relationships.
- **Target-tail Prediction:** The method accurately predicts potential target tails (such as genes etc.) by incorporating contextualized structural information derived from the head entity and relationship.
- **Efficient Multi-Context Sampling:** By prioritizing informative structural patterns through density-based sampling, MuCoS reduces computational overhead while preserving high predictive accuracy.
- **Elimination of Auxiliary Data Requirements:** Operating effectively without reliance on extensive entity descriptions or negative sampling, MuCoS is particularly well-suited for sparse biomedical datasets.

2 Related Work

Drug target discovery has been approached from multiple computational perspectives. Similarity-based methods quantify relationships by computing pairwise distances—often using Euclidean or other metric functions—between drugs and their target proteins (Shi and Li, 2018). These methods typically rely on handcrafted similarity measures to distinguish interacting pairs. Feature-based techniques, predominantly employing support vector machines (Zhang et al., 2017), formulate the problem as a binary classification or two-class clustering task to differentiate between positive and negative drug–target associations based on engineered features.

Recent graph-based methods leverage heterogeneous networks that integrate multiple similarity metrics—such as drug–drug, target–target, and cross-modal associations—to exploit the homophily principle in biological systems (Ban et al., 2019). These approaches infer missing links by modelling complex interdependencies among drugs, proteins, diseases, and pathways. In parallel, the application of embedding-based techniques has evolved considerably (Bordes et al., 2013; Yang et al., 2014; Trouillon et al., 2016). For instance, Mohamed et al. (Mohamed et al., 2019) introduced ComplEx-SE, a variant of the ComplEx KGE model that adopts a squared error-based loss for enhanced accuracy. Recent works like NeoDTI (Liu et al., 2024) combine graph neural networks with probabilistic reasoning to achieve state-of-the-art performance in drug–target prediction.

Despite these advances, current KGC methods still face challenges in drug target discovery. Traditional embedding models depend on static, pre-trained embeddings, which hinder their ability to generalize to novel entities and interactions in rapidly evolving biomedical data (Gul et al., 2024). Text-based and large language model approaches require rich and consistent annotations—a resource often sparse in biomedical domains (Gul et al., 2025). Additionally, the reliance on extensive negative sampling during training imposes significant computational burdens, particularly for large-scale datasets. These limitations motivate us to develop MuCoS as a flexible, context-aware and computationally efficient model that integrates both structural and textual cues to drive the discovery of new drug targets.

3 Methodology

MuCoS addresses two knowledge graph completion tasks: **(1) Link Prediction** (inferring missing relations in triples like $(h, ?, t)$) and **(2) Tail Prediction** (identifying missing tail entities in $(h, r, ?)$). Both tasks are divided into general and drug-target-specific subtasks to balance broad applicability with a biomedical focus. Using the PharmKG-8k and the KEGG50k dataset, general subtasks predict relations/tails across all entities and relations, while drug-target subtasks use a filtered subset to predict specific relations.

MuCoS builds on the MuCo-KGC model (Gul et al., 2025) to boost computational efficiency by strategically *sampling* high-density contextual information *(i.e., entities or relations that appear most frequently)* from both entity and relation-neighbouring contexts before integrating it with BERT for precise predictions. For the transformer part of MuCoS, DistilBERT (base, uncased) has been employed, which is a smaller model that helps MuCoS run efficiently while still capturing context

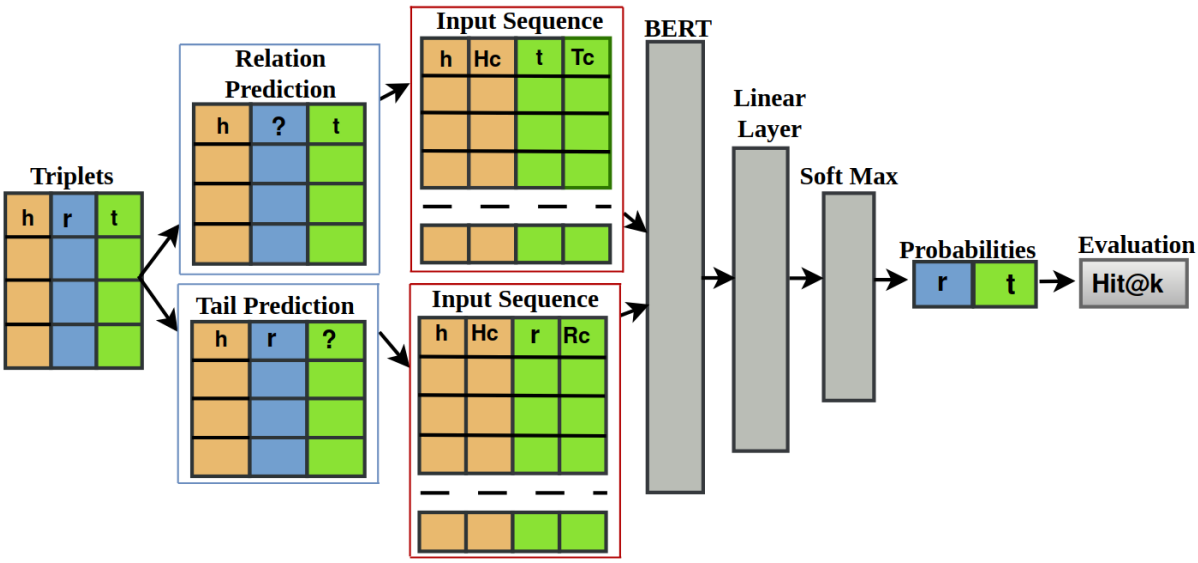

Figure 1: A concise overview of the MuCoS model pipeline, which is designed to predict general and drug-target relations and tail entities. The boxes on the left show the input sequence to the BERT model, where (h) head, $(\mathcal{H}_c)$ head context, (t) tail, $(\mathcal{T}_c)$ tail context, (r) relation, and $(\mathcal{R}_c)$ relation context. This integrated context is passed through the BERT model with a linear classifier and softmax function to generate probabilities for relations and tail.

well. We selected DistilBERT for its efficiency, retaining 95% of BERT's performance while being 40% smaller, making it suitable for large-scale knowledge graph tasks (Sanh et al., 2019). Figure 1 provides an overview of the MuCoS pipeline. The subsequent sections detail the computations of the contextual information and the sampling process in the MuCoS pipeline.

Given a head (h), tail (t), a relation (r) between them, MuCoS first figures out the corresponding neighbouring contexts, i.e., the head context $(\mathcal{H}_c)$, the tail-context $(\mathcal{T}_c)$ or the relationship context $(\mathcal{R}_c)$ and then selects out the high-density contexts. Based on the task at hand, relevant contexts are then concatenated and passed on to a BERT model with a linear classifier and softmax function to generate probabilities for relations or tails.

Head Context $\mathcal{H}_c$: To extract the contextual information for the head, i.e., $\mathcal{H}_c$, we first identify the relations associated with the head entity h, i.e., the relation neighbourhood $\mathcal{R}(h)$. If l relations are associated with h from the set $\mathcal{R}$ of all relations r_i in the graph,G, then:

$$\mathcal{R}(h) = A_{i=1}^{l}\left(\{r_i \mid (h, r_i, e_j) \in \mathcal{T}, e_j \in \mathcal{E}\}\right) \tag{1}$$

where $A(\cdot)$ is the concatenation operation $\|$, $\mathcal{T}$ is the set of training triples, $\mathcal{E}_t$ is the set of all tail entities, and r_i represents each relation associated with h. Next, we find the tail entities e that are neighbours (i.e., have a direct connection) with the head entity h, i.e., tail neighbourhood $\mathcal{E}(h)$, using the identified relations in $\mathcal{R}(h)$. Assuming m neighbour tails, $\mathcal{E}(h)$ is expressed as:

$$\mathcal{E}(h) = A_{i=1}^{m}\left(\{t_i \mid (h, r_j, e_i) \in \mathcal{T}, r_j \in \mathcal{R}\}\right) \tag{2}$$

where $\mathcal{E}(h)$ is the set of all tail entities t_i directly associated with the h through some relation r_j.

Sampling: While MuCo-KGC (Gul et al., 2025) integrates $\mathcal{R}(h)$ and $\mathcal{E}(h)$ calculates the head context, this study introduce a density-based sampling for context calculation $\mathcal{H}_c$, where the density $\rho(e)$ of an entity $e \in \mathcal{E}(h)$ is defined as its frequency of appearance in $\mathcal{T}$.

$$\rho(t) = |\{(h, r, t) \in \mathcal{T}\}|, for\ any\ h, r \tag{3}$$

$\rho(t)$ denotes the *density* of the tail entity t, defined as the number of times t appears as the tail in triples (h, r, t). Using these density values, we select n entities of highest density values and the relationships between head h and these top-n selected entities:

$$\text{top}_n(\mathcal{E}(h)) = \text{sort}(\mathcal{E}(h), \text{by } \rho(e))[:n] \tag{4}$$

$$\mathcal{R}^*(h) = A_{i=1}^{n}\big(\{r_i \mid (h, r_i, e_j) \in \mathcal{T}, \\ e_j \in \text{top}_n(\mathcal{E}(h))\}\big) \tag{5}$$

$\text{top}_n(\mathcal{E}(h))$ selects the top n tail entities from $\mathcal{E}(h)$ sorted by their density $\rho(e)$. $\mathcal{R}^*(h)$ concatenates the relations r_i connected h and the selected top-n

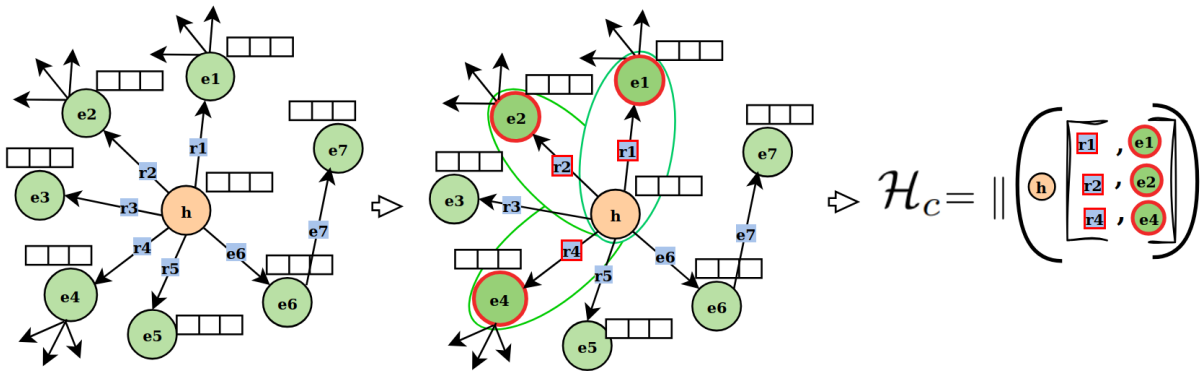

Figure 2: MuCoS $\mathcal{H}_c$ construction. The left graphical view illustrates one hop head h context, which consists of the set of relations $\mathcal{R}(h)$ $(r_1, r_2, r_3, r_4, r_5, r_6)$ and the set of neighbouring tail entities $\mathcal{E}(h)$ $(e_1, e_2, e_3, e_4, e_5, e_6)$ associated with the head entity h. The middle view shows the sampling process, where only the top-n (suppose $n = 3$) tail entities e are selected and concatenated ($\|$) based on their density $\rho(e)$, to calculate the optimized head context $\mathcal{H}_c$.

tail entities. The optimized head context $\mathcal{H}_c$ is then defined as:

$$\mathcal{H}_c = \mathcal{R}^*(h) \cup \text{top}_n(\mathcal{E}(h)) \tag{6}$$

Figure 2 illustrates this sampling process, highlighting only a select subset of high-density neighbours (shown in red border) used to compute the aggregated context $\mathcal{H}_c$. We follow the same procedure to compute the tail context $\mathcal{T}_c$ (for a given tail) required along with head context $\mathcal{H}_c$ in the relation prediction task.

Relation Context $\mathcal{R}_c$: To acquire the relation context $\mathcal{R}_c$, we identify all entities (heads and tails) associated with the operational relation r in the knowledge graph $\mathcal{G}$. This includes the set of heads (e.g., drugs) e_i and tails (e.g., genes) e_j connected by r:

$$\mathcal{E}(r) = A^{o}_{i,j=1}\left(\{e_i, e_j\} \mid (e_i, r, e_j) \in \mathcal{T}\right) \tag{7}$$

$\mathcal{E}(r)$ is the concatenation of all head-tail entity pairs (e_i, e_j) connected by the relation r in the knowledge graph.

Sampling: From the set of entities in $\mathcal{E}_c$, the top-k elements with the highest density values $\rho(e)$ are selected to generate the optimized relationship context $\mathcal{R}_c$.

$$\mathcal{R}_c = \text{top}_k(\mathcal{E}(r)) = \text{sort}(\mathcal{E}(r),\ \text{by}\ (\rho(e_i) + \rho(e_j)))[:k] \tag{8}$$

$\mathcal{R}_c$ therefore provides a focused global perspective on r's patterns, enhancing generalization without excessively raising the time complexity. Figure 3 depicts the sampling process involved in computing $\mathcal{R}_c$, highlighting the selection of k high-density entity pairs (shown in red border) involved with the relation r to form the optimized relationship context. Following the extraction of contextual information via density-based sampling, MuCoS integrates these contexts into a BERT-based framework for prediction. The process for each subtask, leveraging the KEGG50k dataset and its filtered drug-target subset, is detailed below:

- For task **(1), link prediction**, which includes two subtasks: *General link prediction* $(h, ?, t)$: The concatenated representations $\mathcal{H}_c$ (head context) and $\mathcal{T}_c$ (tail context) are combined with the head entity h and tail entity t to form the input sequence $[h, \mathcal{H}_c, t, \mathcal{T}_c]$. This sequence passes through BERT's transformer layers, generating a contextualized representation for each token. A classification layer then predicts the relation r, with a softmax function calculating the probability distribution over all relations:

$$P(r \mid h, t) = \text{softmax}(W \cdot \text{BERT}(h, \mathcal{H}_c, t, \mathcal{T}_c)) \tag{9}$$

  Drug-target link prediction $(h, ?, t)$: Following Mohamed et al (Mohamed et al., 2019) in this case, we filter the dataset to consider drug-target relations only. Other than that, we follow the same methodology as above, where the input sequence $[h, \mathcal{H}_c, t, \mathcal{T}_c]$ is processed by BERT to predict the drug-target-specific relations r.

- For task **(2), tail prediction**, which includes two subtasks: *General tail prediction* $(h, r, ?)$:

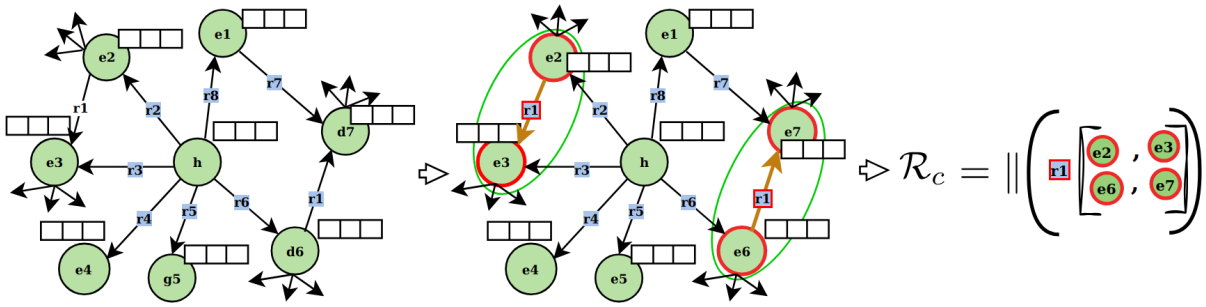

Figure 3: $\mathcal{R}_c$ construction. The left view illustrates the relationship r_1 and entities (head, tail) connected by r_1. The graph in the middle depicts optimization, selecting the top k (suppose $k = 2$) entities based on density ρ, retaining pairs such as (e_2, e_3) and (e_6, e_7) The optimized context $\mathcal{R}_c$ is aggregated using concatenation ($\|$), as shown in the right section.

The concatenated representations $\mathcal{H}_c$ (head context) and $\mathcal{R}_c$ (relation context) are combined with the head entity h and relation r to form the input sequence $[h, \mathcal{H}_c, \text{r}, \mathcal{R}_c]$, using the full KEGG50k dataset. BERT processes this sequence, and a classification layer predicts the tail entity t:

$$P(t \mid h, r) = \text{softmax}(W \cdot \text{BERT}(h, \mathcal{H}_c, r, \mathcal{R}_c)) \quad (10)$$

Drug-target tail prediction $(h, r, ?)$: Following above, we use a filtered drug-target subset of the KEGG50k dataset, to predict the tail entity t.

We train the model using cross-entropy loss. For link prediction, Equation 11 defines the loss with y_i as the one-hot true label for relation r_i and $P(r_i \mid h, t)$ as the predicted probability. For tail prediction, Equation 12 defines the loss with y_i as the true label for tail entity t_i and $P(t_i \mid h, r)$ as its predicted probability.

$$\textbf{(a)}\ \mathcal{L} = -\sum_{i=1}^{N} y_i \log P(r_i \mid h, t), \quad (11)$$

$$\textbf{(b)}\ \mathcal{L} = -\sum_{i=1}^{N} y_i \log P(g_i \mid h, r) \quad (12)$$

where y_i is the true label for the relation r_i, and $P(r_i \mid h, t)$ is the predicted probability of the relation given h and t. On the other hand, $P(t_i \mid h, r)$ is the predicted probability of the tail given h and r.

3.1 Computational Advantage of MuCoS over MuCo-KGC

Compared to MuCo-KGC (Gul et al., 2025), MuCoS reduces computational complexity by sampling only the most significant neighbours (based on density) from the full entity and relation contexts. MuCoS employs two sampling thresholds: n for the head entity context $\mathcal{H}_c$ and k for the relation context $\mathcal{R}_c$. To compute the complexities, we first define two terms: (i) the average density $(avg\_density)$ as the average number of neighbours per entity in the knowledge graph, and (ii) average appearance $(avg\_appearance)$ of a relation r in the dataset.

$$avg\_density = \frac{|T|}{|E|}, \qquad avg\_appearance = \frac{|T|}{|R|} \quad (13)$$

where $|T|$ is the total number of triples, $|E|$ entities, and $|R|$ unique relations.

For **MuCo-KGC**, the complexity of computing the head context $\mathcal{H}_c$ and the relation context $\mathcal{R}_c$ is based on full neighbourhoods without sampling. The complexity of $\mathcal{H}_c$ depends on the number of relations involving the head entity h, denoted as $|\mathcal{R}(h)|$, and the number of neighbouring entities $|\mathcal{E}(h)|$, both approximated by $avg\_density$ (see Equation 15). The complexity of $\mathcal{R}_c$ is determined by the number of entity pairs connected by relation r, $|\mathcal{E}(r)|$, estimated using $avg\_appearance$ (see Equation 16). Therefore, the overall complexity for context computation in MuCo-KGC is defined equals:

$$O(2 \cdot avg\_density + avg\_appearance) \quad (14)$$

$$\text{where,} \quad O(|\mathcal{H}_c|) = O(|\mathcal{R}(h)| + |\mathcal{E}(h)|) = O(2 \cdot avg\_density) \quad (15)$$

$$\text{and,} \quad O(|\mathcal{R}_c|) = O(|\mathcal{E}(r)|) = O(avg\_appearance) \quad (16)$$

For **MuCoS**, the head context $\mathcal{H}_c$ is computed by selecting the top-n high-density neighbouring entities and their corresponding relations, and the

relation context $\mathcal{R}_c$ are computed by selecting the top-k high-density entity pairs. The complexity of $\mathcal{H}_c$ is $O(n)$ for the sampled entities and $O(n)$ for the corresponding relations, and $\mathcal{R}_c$ is $O(k)$ for the sampled entity pairs. Thus, the overall complexity for context computation in MuCoS is:

$$O(2 \cdot n + k) \tag{17}$$

Since sampling threshold values n and k are much smaller than $avg\_density$ and $avg\_appearance$ in large datasets like KEGG50k, MuCoS achieves a significant reduction in computational cost compared to MuCo-KGC.

For example, in case of the KEGG50k dataset (with triplets $|T| = 63,080$, entities $|\mathcal{E}| = 16,201$, and relations $|\mathcal{R}| = 9$), $avg\_density \approx 3.895$, and $avg\_appearance \approx 7,008.89$. Therefore, the complexity of MuCo-KGC on the KEGG50k dataset is: $O(2{\cdot}3.895{+}7,008.89) = O(7,016.68)$. For MuCoS (with $n = 15$, $k = 10$: the complexity is $O(2 \cdot 15 + 10) = O(40)$. This is a speed up by a factor of ≈ 175.42 in context computation, i.e., the process of extracting and aggregating relevant neighbourhood information associated with a given head entity and relation. Sampling the context reduces the input token length, which further contributes to the efficiency slightly. The primary computational gains however arise from our selective sampling strategy, which significantly limits the amount of nodes/relations processed for context extraction.

Sampling size values of n at 15 and k at 10, although empirical, are motivated from the ablation studies on MuCo-KGC, suggesting that the head context plays a greater role than the relationship context in model performance (see Table 1 and Table 3 for details).

3.2 Experimental Setup

We evaluate MuCoS on two prediction tasks using KEGG50k and PharmKG-8k datasets: link and tail prediction. Each task is evaluated in two settings: the full KEGG50k dataset and a drug-target subset. In **link prediction**, we infer the missing relation in $(h, ?, t)$, with general and drug-target variants. Similarly, in **tail prediction**, we predict the missing entity in $(h, r, ?)$ for both settings. Below we provide the details of the dataset used in our experiments, the hyperparameter settings, and the evaluation criteria.

Datasets: *KEGG50k* [1] medical domain dataset, comprises 63,080 triples split into 57,080 training, 3,000 validation, and 3,000 testing instances (i.e. a 90:5:5 ratio split). Drug-target only triplet counts are 10769, 585, and 650 for the train, valid, and test sets. The dataset comprises 16,201 unique entities $\mathcal{E}$ where $(\mathcal{E}_d \cup \mathcal{E}_g) \subset \mathcal{E}$ and 9 distinct types of drug-target relationships, enabling a comprehensive mapping of pharmacological interactions. *PharmKG-8k* [2] comprises 400,788 training triplets, 49,536 testing triplets, and 50,036 validation triplets, covering 7,601 entities. These are categorized into Chemical, Disease, and Gene types, integrating data from DrugBank, TTD, OMIM, PharmGKB, and GNBR.

Hyperparameters: The input sequence is tokenized with a maximum length of 128 tokens. Training is conducted over 50 epochs using the AdamW optimizer with a learning rate of 5×10^{-5} and a batch size of 16. Experiments were performed on an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory.

Evaluation: Model performance is assessed using standard metrics, Mean Reciprocal Rank (MRR) and Hits@k, as defined in Equations 18 and 19, to evaluate the accuracy of general and drug-target relations and tail predictions:

$$\text{MRR} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\text{rank}_i}, \tag{18}$$

$$\text{Hits@k} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\text{rank}_i \leq k), \tag{19}$$

MRR measures the average of the reciprocal ranks of the correct item across all queries. A higher MRR indicates better ranking performance. $H@k$ measures the proportion of queries where the correct item appears in the top k ranks. It provides a metric for evaluating ranking quality at different points.

3.3 Results and Discussion

Link Prediction: Table 1 demonstrates that MuCoS outperforms state-of-the-art baselines on the **KEGG50k** dataset. It achieves an MRR of 0.65 for general link prediction across all relations, a 13% improvement over ComplEx-SE's 0.52, and its Hits@1 score of 0.52 exceeds ComplEx-SE's 0.45 by 7%. Moreover, Hits@3 and Hits@10 scores of 0.60 and 0.86 further underscore the robust ranking performance of MuCoS. Although MuCo-KGC

[1] KEGG50k: https://shorturl.at/pWSJO

[2] PharmKG-8k: https://zenodo.org/records/4525237

Table 1: Relationship prediction results over the KEGG50k dataset on both general links and drug target links only.

Model	General link prediction				Drug-target link prediction			
	MRR	**Hits@1**	**Hits@3**	**Hits@10**	**MRR**	**Hits@1**	**Hits@3**	**Hits@10**
TransE (Bordes et al., 2013)	0.46	0.38	0.50	0.63	0.75	0.69	0.79	0.86
DistMult (Yang et al., 2014)	0.37	0.27	0.42	0.57	0.61	0.50	0.69	0.81
ComplEx (Trouillon et al., 2016)	0.39	0.31	0.43	0.57	0.68	0.61	0.71	0.82
ComplEx-SE (Mohamed et al., 2019)	0.52	0.45	0.56	0.68	0.78	0.73	0.81	0.88
MuCoS- (H_c Only)	0.52	0.44	0.55	0.69	0.75	0.65	0.77	0.1
MuCoS (T_c Only)	0.45	0.37	0.51	0.61	0.70	0.59	0.70	0.1
MuCo-KGC (Gul et al., 2025)	**0.79**	**0.58**	**0.73**	**0.92**	**0.94**	**0.91**	**0.96**	**1**
MuCoS	0.65	0.52	0.60	0.86	0.84	0.74	0.84	1

(Gul et al., 2025) achieves state-of-the-art performance, MuCoS offers a significant computational advantage with a small reduction in accuracy.

In drug–target prediction,which focuses on identifying relationships between drugs and their targets (e.g., genes), MuCoS achieves an MRR of 0.84—a 6% improvement over ComplEx-SE's 0.78—demonstrating the benefit of contextual head/tail information. It also records Hits@1 of 0.74 (vs. 0.73), Hits@3 of 0.84 (a 3% gain), and a perfect Hits@10 of 1.00 (12% improvement), outperforming TransE, DistMult, and ComplEx. Although MuCo-KGC attains higher accuracy (e.g., an MRR of 0.94), its prohibitive computational cost limits scalability. MuCoS, by offering competitive performance with substantial efficiency gains, provides a scalable solution for real-world, large-scale drug discovery.

Table 2 shows that MuCoS achieves state-of-the-art performance on **PharmKG-8k**. It attains an MRR of 0.452, compared to NC-KGE's 0.228, and a Hits@1 of 0.258 versus 0.145. Additionally, MuCoS records Hits@3 and Hits@10 scores of 0.602 and 0.676, respectively.

Table 2: PharmKG8k-28 Results for Link Prediction Task.The symbol □ denotes that the results are taken from Paper (Zheng et al., 2021), while the symbol △ results are taken from Paper (Fan et al., 2023). R2N results are reported from (Diligenti et al., 2023).

Model	**MRR**	**H@1**	**H@3**	**H@10**
TransR □	0.075	0.030	0.071	0.155
RESCAL □	0.064	0.023	0.057	0.122
ConvE □	0.086	0.038	0.087	0.169
ConvKB □	0.106	0.052	0.107	0.209
RGCN □	0.067	0.027	0.062	0.139
HRGAT □	0.154	0.075	0.172	0.315
TransE △	0.116	0.038	0.127	0.269
DistMult △	0.218	0.152	0.237	0.335
ComplEx △	0.124	0.064	0.128	0.244
TruckER △	0.182	0.103	0.202	0.336
HRGAT △	0.134	0.063	0.144	0.271
SACN △	0.156	0.085	0.170	0.296
CompGCN △	0.193	0.110	0.216	0.352
SE-GNN △	0.206	0.120	0.232	0.374
R2N	0.215	0.145	0.234	0.342
NC-KGE △	0.228	0.145	0.252	0.390
MuCoS	**0.452**	**0.258**	**0.602**	**0.676**

Tail Prediction: Table 3 compares tail prediction performance between MuCoS and MuCo-KGC under both general and drug-target settings. While MuCo-KGC (without sampling) achieves higher MRR, Hits@1, and Hits@3 in the general scenario, MuCoS (sampling-based) excels in drug-target cases, particularly in Hits@10. Thus, sampling enhances prediction accuracy for drug targets at a slight cost in the general scenario, and MuCoS offers a significant computational advantage while outperforming other models on KEGG50k.

4 Ablation Study

We analyze the contributions of the Head Context ($\mathcal{H}_c$) and Tail Context ($\mathcal{T}_c$) components for relation (link) prediction, and Head Context ($\mathcal{H}_c$) and Relation Context ($\mathcal{R}_c$) for the prediction of the tail. The results are presented in Tables 1 and 3.

Relationship Prediction: Table 1 reports the results for both general link prediction and drug-target link prediction scenarios. MuCo-KGC (Gul et al., 2025), the earlier method, demonstrates strong performance across all metrics, achieving an MRR of 0.79 for general link prediction and 0.94 for drug-target link prediction. These results highlight its ability to leverage both $\mathcal{H}_c$ (Head Context) and $\mathcal{R}_c$ (Relation Context) effectively, excelling particularly in Hits@1 (0.58 and 0.91) and Hits@10 (0.92 and 0.1).

Tail Prediction: Table 3 presents the results for both general tail prediction and drug-target-specific

Table 3: Tail prediction results on the KEGG50k dataset were evaluated for both general and drug target scenarios using methods with and without sampling.

Model	General tail prediction				Drug-target tail prediction			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
MuCoS (H_c Only)	0.26	0.20	0.34	0.55	0.38	0.30	0.41	0.78
MuCoS (R_c Only)	0.21	0.15	0.28	0.39	0.31	0.19	0.36	0.69
MuCo-KGC	0.39	0.34	0.521	0.718	0.567	0.457	0.628	0.917
MuCoS	0.31	0.215	0.40	0.57	0.442	0.259	0.46	0.868

scenarios. MuCo-KGC (Gul et al., 2025) delivers robust performance, achieving an MRR of 0.39 for general tail prediction and 0.567 for drug-target tail prediction. Its superior Hits@1 scores (0.34 and 0.457) and Hits@10 scores (0.71 and 0.917) confirm its effectiveness in capturing complex relational patterns in the graph.

Across both prediction tasks, the $\mathcal{H}_c$-Only configuration consistently outperforms achieving an MRR of 0.52 (general links) and 0.75 (drug-target links) for relationships, and 0.26 (general tails) and 0.38 (drug-target tails) for tail predictions. This highlights the critical role of localized contextual information over global relational patterns, which tend to underperform when used in isolation $\mathcal{T}_c$-Only MRR: 0.45 and 0.70 for links; $\mathcal{R}_c$-Only MRR: 0.21 and 0.31 for tails).

5 Conclusion

The study introduces MuCoS, a multi-context-aware sampling method that uses DistilBERT to improve drug-target relation predictions and tail entity predictions in biomedical knowledge graphs. MuCoS employs a dual strategy combining transformer-based textual modeling with context-aware sampling to overcome limitations of existing models, such as poor generalization, negative sampling, and the need for descriptive entity information. It extracts and optimizes contextualized information from the head, tail, and relation entities using density-based sampling and its lexical semantics, capturing richer structural patterns and reducing computational complexity. Experimental results show superior performance over state-of-the-art models, with improvements in MRR and Hits@1 for general and drug-target relationship prediction on both KEGG50k and PharmKG-8k datasets. Future work could focus on adaptive sampling to dynamically adjust n and k for sparse KGs, and integrate multimodal data like protein sequences or chemical structures to enhance drug-target prediction.

References

Tomohiro Ban, Masahito Ohue, and Yutaka Akiyama. 2019. Nrlmfβ: Beta-distribution-rescored neighborhood regularized logistic matrix factorization for improving the performance of drug–target interaction prediction. *Biochemistry and biophysics reports*, 18:100615.

A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in NIPS*, volume 26. Curran Associates, Inc.

Alberto Cattaneo, Thomas Martynec, Stephen Bonner, Carlo Luschi, and Daniel Justus. 2024. Towards linking graph topology to model performance for biomedical knowledge graph completion. In *ICML'24 Workshop ML for Life and Material Science: From Theory to Industry Applications*.

Michelangelo Diligenti, Francesco Giannini, Stefano Fioravanti, Caterina Graziani, Moreno Falaschi, and Giuseppe Marra. 2023. Enhancing embedding representations of biomedical data using logic knowledge. In *2023 IJCNN*, pages 1–8. IEEE.

Zhiguang Fan, Yuedong Yang, Mingyuan Xu, and Hongming Chen. 2023. Node-based knowledge graph contrastive learning for medical relationship prediction. In *https://arxiv.org/abs/2310.10138*.

Haji Gul, Abdul Ghani Naim, and Ajaz A Bhat. 2024. A contextualized bert model for knowledge graph completion. In *MusIML, Advances in Neural Information Processing Systems (NeurIPS)*, pages https://neurips.cc/virtual/2024/affinity–event/105018.

Haji Gul, Abdul Ghani Naim, and Ajaz A Bhat. 2025. Muco-kgc: Multi-context-aware knowledge graph completion. In *The 29th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), https://arxiv.org/abs/2503.03091*.

Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, and Sergio E Baranzini. 2017. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *elife*, 6:e26726.

Chang Liu, Kaimin Xiao, Cuinan Yu, Yipin Lei, Kangbo Lyu, Tingzhong Tian, Dan Zhao, Fengfeng Zhou, Haidong Tang, and Jianyang Zeng. 2024. A probabilistic knowledge graph for target identification. *PLOS Comp Bio*, 20(4):e1011945.

Sameh K Mohamed, Aayah Nounu, and Vít Nováček. 2019. Drug target discovery using knowledge graph embeddings. In *Proceedings of the 34th ACM/SIGAPP symposium on applied computing*, pages 11–18.

Kanica Sachdev and Manoj Kumar Gupta. 2019. A comprehensive review of feature based methods for drug target interaction prediction. *Journal of biomedical informatics*, 93:103159.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Zezhi Shi and Jianhua Li. 2018. Drug-target interaction prediction with weighted bayesian ranking. In *Proceedings of the 2nd ICBB*, pages 19–24.

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

Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*.

Jun Zhang, Muchun Zhu, Peng Chen, and Bing Wang. 2017. Drugrpe: random projection ensemble approach to drug-target interaction prediction. *Neurocomputing*, 228:256–262.

Shuangjia Zheng, Jiahua Rao, Ying Song, Jixian Zhang, Xianglu Xiao, Evandro Fei Fang, Yuedong Yang, and Zhangming Niu. 2021. Pharmkg: a dedicated knowledge graph benchmark for bomedical data mining. *Briefings in bioinformatics*, 22(4):bbaa344.