

## Research Article

# Attention-Based Learning for Predicting Drug-Drug Interactions in Knowledge Graph Embedding Based on Multisource Fusion Information

Yu Li <sup>1</sup>, Zhu-Hong You <sup>1</sup>, Shu-Min Wang <sup>2</sup>, Cheng-Gang Mi,<sup>3</sup> Mei-Neng Wang,<sup>4,5</sup> Yu-An Huang,<sup>1</sup> and Hai-Cheng Yi<sup>1</sup>

<sup>1</sup>School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China

<sup>2</sup>Department of Obstetrics and Gynecology, The Affiliated Hospital of Inner Mongolia Medical University, Huhhot 010050, China

<sup>3</sup>Foreign Language and Literature Institute, Xi'an International Studies University, Xi'an 710129, China

<sup>4</sup>School of Mathematics and Computer Science, Yichun University, Yichun, Jiangxi 336000, China

<sup>5</sup>School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

Correspondence should be addressed to Shu-Min Wang; [shumin\\_wang0915@163.com](mailto:shumin_wang0915@163.com)

Received 23 November 2023; Revised 16 January 2024; Accepted 17 February 2024; Published 2 March 2024

Academic Editor: Gennaro Vessio

Copyright © 2024 Yu Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drug combinations can reduce drug resistance and side effects and enable the improvement of disease treatment efficacy. Therefore, how to effectively identify drug-drug interactions (DDIs) is a challenging problem. Currently, there exist several approaches that leverage advanced representation learning and graph-based techniques for DDIs prediction. While these methods have demonstrated promising results, a limited number of approaches effectively utilize the potential of knowledge graphs (KGs), which provide information on drug attributes and multirelation among entities. In this work, we introduce a novel attention-based KGs representation learning framework. To encode drug SMILES sequence, a pretrained model is used, while molecular structure information is mapped as the initialization of nodes within the KG using a message-passing neural network. Additionally, the knowledge-aware graph attention network is employed to capture the drug and its topological neighbor representation in the KG representation module. To prevent the oversmoothing problem, the residual layer is used in the DDI prediction module. Comprehensive experiments on several datasets have demonstrated that the proposed method outperforms the state-of-the-art algorithms on the DDI prediction task across a range of evaluation metrics. It achieves an accuracy of 0.924 and an AUC of 0.9705 on the KEGG dataset and attains an ACC of 0.9777 and an AUC of 0.9959 on the OGB-biokg dataset. These experimental findings affirm that our approach is a dependable model for predicting the association of drugs.

## 1. Introduction

Drug combinations are a very promising therapeutic strategy [1]. However, it may increase the risk of unexpected adverse drug reactions (ADRs), such as reduced drug efficacy or increased drug toxicity, which can result in injuries and deaths [2, 3]. Hence, it becomes necessary to effectively detect the potential drug-drug interactions (DDIs) to mitigate the influence of unexpected pharmacological effects.

There are some machine learning-based approaches that define the potential DDIs by utilizing a range of drug-related

similarity features, such as drug structure [4–6], adverse or side effects [7, 8], and phenotypic similarity [9, 10]. However, these studies heavily rely on manual characteristics and domain expertise. Other works attempt to use the embedding approach to automatically discover drugs representations. Various methods have been employed to model DDIs, such as matrix factorization [11–13], random walk [14], and graph neural networks [15]. Despite the remarkable performance achieved by the above methods, a notable limitation is that they treat DDIs as independent data samples, without due consideration of their interrelation correlations, such as drug-

target pairs. Several researchers have explored the application of knowledge graphs (KGs) for DDIs, as seen in the works of Celebi et al. [16] and Karim et al. [17]. The drug embedding is obtained using conventional machine learning methods, such as ComplEx [18]. Lin et al. introduced an end-to-end neural method based on the knowledge graph to predict potential DDIs [19]. Although these studies effectively used a knowledge graph with ample biomedical information, resulting in improved performance, they regrettably omitted the consideration of the attribute information of drugs. It has been proved that node attributes are essential for accurately analysing complex networks [20–22]. Su et al. introduce an innovative framework for KG representation learning that incorporates attention mechanisms, employing the simplified molecular input line entry system (SMILES) sequence as an attribute for drug entities within the knowledge graph [23]. It is common knowledge that drugs contain chemical molecules with a distinct spatial structure [24, 25]. However, the sequence alone cannot fully represent the spatial structure of the drug. Therefore, it is crucial to include an atomic graph of the drug as an additional attribute in the knowledge graph.

This research presents a novel attention-based KG representation learning framework, which considers two drug attributes (molecular graph and SMILES sequence) and triple facts in KG. There are three blocks in our framework. The first block is the drug representation initialization module, where we find the SMILES sequence and molecular graph as its drug attributes. Based on these attributes, we leverage message-passing neural network (MPNNs) [26] and the SELFormer pretrained model [27] to initialize the drug embedding. The second block, known as the knowledge graph representation learning module, is tasked with acquiring the drug and its topological neighbourhood representation through knowledge-aware graph attention network. Both high-order structures information and semantic relationship features can be found in this block. The potential DDIs are predicted as binary classification tasks in the final block. We summarize our research's main contributions as follows:

- (i) For the DDIs prediction, we have devised a novel KG representation learning framework that exploits attention mechanisms. This framework effectively utilizes the information of biomedical KGs, as well as drug molecular structure and sequence information.
- (ii) We establish a drug representation initialization module designed to acquire the initial drug embeddings, which are based on their attributes within the KGs. At the same time, we use the knowledge-aware graph attention network to calculate attention weights by considering the drug node, its surrounding neighbourhoods, and triple facts, which are adapted to better learn the drug and its topological neighbourhood representation.
- (iii) Through a series of experiments conducted on two biomedical datasets, our model consistently demonstrates superior performance compared to the current state-of-the-art methods.

## 2. Related Work

Recently, numerous works have emerged to address the problem of predicting DDIs. Based on the hypothesis that similar drugs exhibit a higher propensity for interacting, some previous studies have attempted to predict drug interactions using analogous feature derived from molecular structure [28] and various properties (e.g., phenotypic [29], functionality [30], and side effects [31]). In the study by Ruy et al, DDI types were estimated by using a deep neural network (DNN) model trained on chemical structure similarity [32]. With the development of neural network methods in the graph domain, several investigators have employed graph neural network methodologies to extract features related to molecular properties and structure for molecular interaction tasks. Xu et al. introduced an approach that uses multiple graph convolution layers to extract features of nodes from their neighbours in a structured entity graph [33]. Additionally, Deac et al. introduce a method that the type of side effect and the molecular structures are combined by a co-attentional mechanism to generate the drug-level representation [24]. GNN-DDI constructs a five-layer graph attention network (GAT) encoder to capture the drug representation [34]. These above works demonstrate the crucial role of drug molecular structure in DDIs prediction.

With the advancement of deep learning techniques and the increasing popularity of extensive biomedical networks, network-based methods employ various advanced techniques, which are roughly categorized into three groups: graph embedding, link prediction, and knowledge graph-based methods. Some works adopt diverse graph embedding algorithms to capture potentially influential network-based features, such as GraRep [12], HOPE [35], DeepWalk [14], and node2vec [36]. Other works treat DDI prediction as a link prediction task within the drug-drug interaction graph or network. Based on a graph auto-encoder, Decagon et al. proposed an approach to predict multirelational links on multimodal graphs comprising multiple types of interaction on the different entities (e.g., drug-protein target interactions) [8]. Liu et al. constructed several drug feature networks and utilized the graph embedding technique to obtain drug representations from these networks [37]. Feng et al. introduce a method that uses the relational graph convolution network (RGCN)-based encoder and similarity regularisation of multi-drug features to learn the topological features of DDI networks [38]. In addition, some works show that using information from multiple sources (e.g., knowledge graph) improves the prediction performance. KG-DDI is a specialised task for identifying drug-drug interactions (DDIs), which uses different embedding approaches node representation within the KGs [17]. The KGNN method can use GCNs and neighbourhood sampling to effectively identify and analyse relationships between neighbourhoods [19]. DDKG takes drug attributes in the KG to learn drug embeddings and uses an attention mechanism to jointly consider neighbouring node embeddings and triple facts simultaneously [23]. DDKG is the most relevant

to our work. Our method differs from the DDKG method in that we include the molecular graph as an additional drug attribute in KGs. Additionally, we utilize knowledge-aware graph network that considers the features of both entities (nodes) and relations (edges) in a multihop neighbourhood of a designated entity/node.

### 3. Materials and Methods

**3.1. Datasets.** In this research, we have utilized two datasets to evaluate the effectiveness of our proposed model. They are KEGG-drug [39] and OGB-biokg [40]. First, to obtain the attribute information of drugs, we processed the two datasets separately to select proven drugs according to the latest version of the DrugBank [41]. Second, we also use these datasets to conduct the KGs. It is important to ensure that knowledge graphs do not contain any information about drug-drug interactions (DDIs). Therefore, we meticulously removed the relationships labelled as *drug-drug* from OGB-biokg and the information presented as *URL: Drug-Drug-Interaction* from KEGG-drug datasets, respectively. Table 1 shows the statistics of the remaining knowledge graph datasets and drug data.

**3.2. Problem Formulation.** In our study, given the DDI matrix  $Y$  and the biomedical knowledge graph  $G_{\text{kg}}$  for the DDI prediction problem, we aim to learn the function  $f$  for estimating the probability of interaction about drug  $d_i$  and drug  $d_j$ , as follows:

$$\widehat{y}_{ij} = f(d_i, d_j | Y, G_{\text{kg}}, \Theta), \quad (1)$$

where  $\Theta$  represents the set of trainable parameters associated with the function  $f$ . The DDI matrix  $Y$  and the knowledge graph (KGs) are specifically described below.

**3.2.1. Knowledge Graph.** The knowledge graph is a collection of knowledge bases about the real world. It is formally presented as a set of triples, denoted as  $G_{\text{kg}} = \{(h, r, t) | h, t \in E, r \in R\}$ , where  $E$  and  $R$  represent the set of entities and the set of relationships, respectively. Each triple  $(h_i, r_i, t_i)$  indicates the relationship  $r_i$  between  $h_i$  and  $t_i$ , where  $i \in \{1, 2, \dots, N_{\text{kg}}\}$ , and  $N_{\text{kg}}$  is the number of triples within KGs.

**3.2.2. DDI Matrix.** Given the drug set  $D = \{d_1, d_2, d_3, \dots, d_{N_d}\}$  and the corresponding set of SMILES sequence as  $\{s_1, s_2, s_3, \dots, s_{N_d}\}$ , where  $N_d$  refers the number of drugs in the DDI matrix. For the DDIs prediction task, we construct the DDI matrix  $Y$ .  $Y$  is the set of  $y_{ij} \in \{0, 1\}$ , where  $y_{ij}$  indicates that there is a reaction between drug  $d_i$  and drug  $d_j$ . It is important to note that if  $y_{ij} = 0$ , it does not necessarily mean that there is no interaction between the two drugs in KG, as this could be a potential interaction that has not yet been identified.

TABLE 1: The statistics of the KEGG-drug dataset and OGB-biokg dataset.

	KEGG-drug	OGB-biokg
Drug	1925	2221
Drug interactions	56983	127196
Entity	129910	84323
Relation type	167	13
KG triplets	362870	3892464

**3.3. Methodology.** Figure 1 illustrates the overall framework of our approach, which includes three modules for predicting DDIs: (a) drug representation initialization module, (b) knowledge graph representation learning module, and (c) DDI prediction module. In the drug representation initialization module, we first handle different attribute features such as SMILES sequences and drug molecular graphs. Then, we concatenate the different embeddings to initialize the drug nodes of the knowledge graph. In the knowledge graph representation learning module, to capture high-order neighbourhood topologies of drugs in KG, we employ similar convolutions that aggregate and integrate topological neighbourhood information. Simultaneously, we use knowledge-aware attention mechanisms to capture entity and relational features in a multihop neighbourhood of a given drug. In the final classifier module, the final latent representations of the given drug pair  $(d_i, d_j)$  are employed to calculate the probability of interaction.

**3.4. Drug Representation Initialization Module.** Within the drug initialization representation module, we have developed both the drug molecular graph representation module and the drug sequence representation module to acquire diverse drug representations from different perspectives.

**3.4.1. Drug Molecular Graph Representation Module.** Figure 2 shows the structure of the drug molecular graph representation module. For each drug  $d_i \in D$ , its 2D molecular graph  $g_i \in G_{\text{drug}}$  can be generated according to its SMILE sequence by the RDKit tool [42]. In particular,  $g_i = \{v, \varepsilon\}$  where  $v$  denotes atoms, and  $\varepsilon$  represents chemical bonds. For generating the structure representation, we employed MPNNs. This process can be divided into two main phases as follows:

- (1) The message-passing phase. For node  $p \in v$ , we first aggregate relevant information from its neighboring nodes. Then, we conducted  $k$  iterations to update the representation of it. Formally, equation (1) describes this phase.

$$e_p^k = U_{(k-1)} \left( e_p^{(k-1)}, \sum_{q \in \mathcal{N}(p)} M_{k-1}(e_p^{(k-1)}, e_q^{(k-1)}, o_{pq}) \right), \quad (2)$$

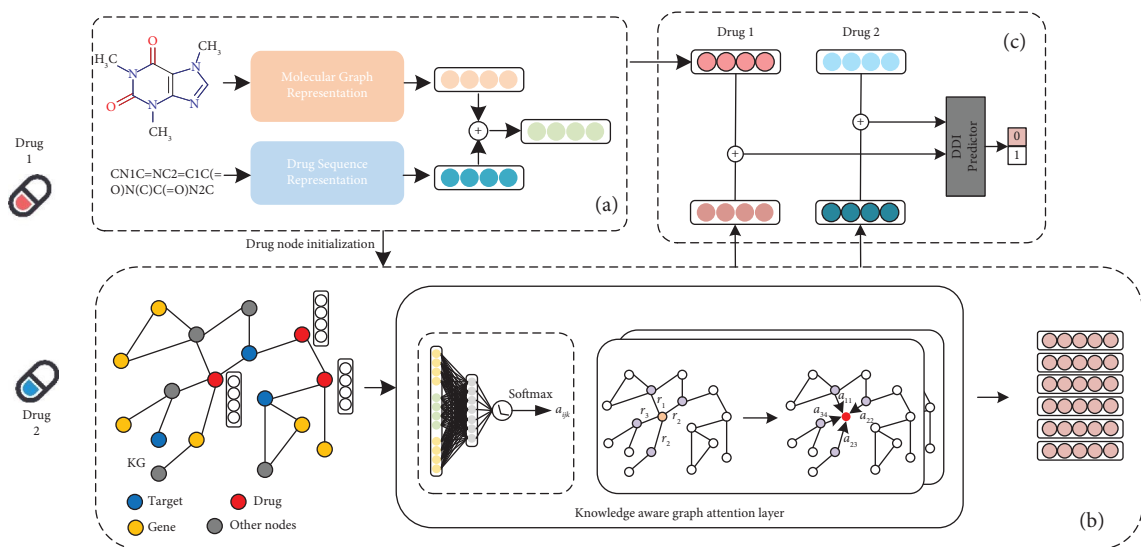


FIGURE 1: The overview of our approach. (a) Drug representation initialization module. (b) Knowledge graph representation learning module. (c) DDI prediction module.

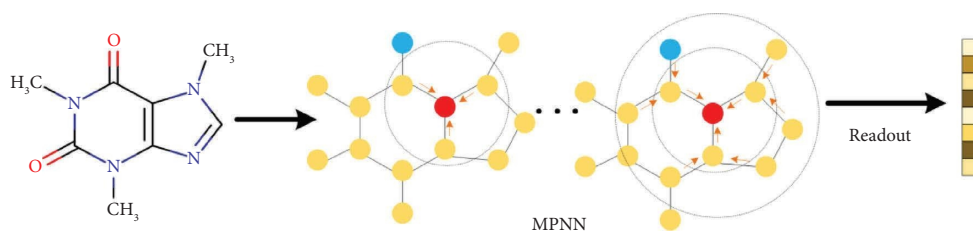


FIGURE 2: The structure of the drug molecular graph representation module.

where  $U_{(k-1)}$  denotes node update functions,  $M_{k-1}$  denotes message functions,  $\mathcal{N}(p)$  implies a set of neighbors of node  $p$  in graph  $g_i$ ,  $o_{pq}$  means the representation of edge between of node  $p$  and node  $q$ , and  $e_p^{(k-1)}$  denotes the representation of node  $p$  after  $k-1$  iterations.

- (2) Readout phase. The global drug graph representation, denoted as  $e_{\text{mole}} \in \mathbb{R}^d$ , can be obtained by relying on the representation of node  $p$  generated after  $k$  iterations. This phase can be represented by the following formula:

$$e_{\text{mole}} = \text{Readout}(\{e_p^k \mid p \in V\}), \quad (3)$$

Here, Readout denotes the readout function. We use the average Readout function [43] which computes the average of all node representations to generate the graph-level representation.

**3.4.2. Drug Sequence Representation Module.** Figure 3 displays the structure of the drug sequence representation module. The SMILES sequence is the most common way to represent drugs composed of molecular characters. It contains more information in comparison to molecular graphs and provides essential functional information about the atoms. Additionally, it enables the representation of

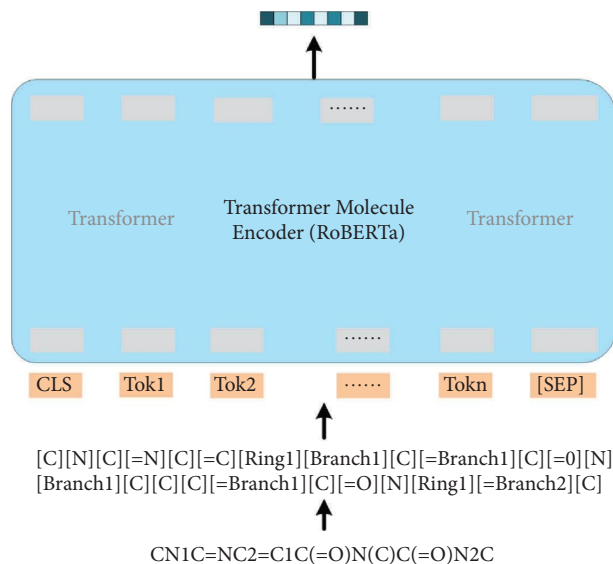


FIGURE 3: The structure of the drug sequence representation module.

long-term dependencies, whereas the molecular graphs of drugs illustrate the interconnectivity between atoms. While some researchers find the SMILES sequence has some disadvantages. For example, a valid SMILES string may exhibit

invalid chemical properties, such as surpassing the natural valency of an atom. Krenn et al. introduced SELFIES (SELF-referencing Embedded String), which is a string-based representation of molecular graphs known for its complete robustness [42]. Therefore, we convert SMILES notations into SELFIES representations utilizing the SELFIES API and use the Byte-level byte-pair encoding (BPE) to tokenize the SELFIES sequence within all drug datasets. Then we use the SELFormer, a pre-trained model proposed by Yüksel et al. [27], to generate the drug embedding  $e_{\text{seq}}$ . It can be described as follows:

$$e_{\text{seq}} = \text{SELFormer}(s_i), \quad (4)$$

when we get the molecular embedding and sequence embedding, we use the simple method to combine those embeddings:

$$e_{\text{init}} = e_{\text{mole}} + e_{\text{seq}}. \quad (5)$$

**3.5. Knowledge Graph Representation Learning Module.** When initializing the knowledge graph with the initial embedding  $e_{\text{init}}$ , it becomes essential to discern the associations between drug nodes and other entities within the knowledge graph. Within the KGs, each node is connected to multiple neighbouring nodes, which carrying distinct levels of significance. Furthermore, entities assume different roles based on the relations they are linked with. So, we employ the knowledge-aware graph attention network based on entity and relationship. Figure 4 illustrates the knowledge-aware attention network.

As shown in Figure 4, given the triple  $(h_i, r_k, t_j)$ , we concatenate the embeddings of entity and relation feature and generate the representation of triple  $\vec{c}_{i,j,k}$  by a linear transformation. This step is shown in equation(6):

$$\vec{c}_{i,j,k} = W_1 \left[ \vec{e}_{h_i} \parallel \vec{e}_{h_j} \parallel \vec{e}_{r_k} \right], \quad (6)$$

where vector  $\vec{e}_{h_i}$ ,  $\vec{e}_{h_j}$ , and  $\vec{e}_{r_k}$  denote the embedding of entity and relationship.  $W_1$  represents the linear transformation matrix. Subsequently, the significance of each triple can be ascertained by employing a linear transformation and applying the LeakyRelu nonlinearity:

$$b_{i,j,k} = \text{LeakyReLU}(W_2 c_{i,j,k}), \quad (7)$$

where  $W_2$  represents the linear transformation matrix.  $b_{i,j,k}$  represents the importance of each triple. To get the relative attention values  $a_{i,j,k}$ , the softmax is employed on  $b_{i,j,k}$ :

$$\begin{aligned} a_{i,j,k} &= \text{softmax}(b_{ijk}) \\ &= \frac{\exp(b_{ijk})}{\sum_{n \in \mathcal{N}_i} \sum_{r \in \mathcal{R}_{in}} \exp(b_{inr})}, \end{aligned} \quad (8)$$

where  $\mathcal{N}_i$  denotes the neighborhood of entity  $i$  and  $\mathcal{R}_{in}$  represents a set of relationships between entity  $i$  and entity  $n$ .

By summing the triple representations with their associated attention weights, we derive the revised embedding  $e_{\mathcal{N}_i}$  for the entity  $h_i$ .

$$e_{\mathcal{N}_i} = \sigma \left( \sum_{j \in \mathcal{N}_i} \sum_{r \in \mathcal{R}_{ij}} a_{ijr} c_{ijr} \right), \quad (9)$$

where  $\sigma$  is activation function.  $\mathcal{N}_i$  denotes the neighborhood of entity  $i$ , and  $\mathcal{R}_{ij}$  represents a set of relationships between entity  $i$  and entity  $j$ .

After obtaining the propagated neighboring information with equation (9), we combine the initial embedding  $\vec{e}_{h_i}$  and  $e_{\mathcal{N}_i}^1$  to update the representation of the drug. It can be described as follows:

$$e_i^1 = \xi(e_{\mathcal{N}_i}, \vec{e}_{h_i}), \quad (10)$$

where  $\xi$  denotes the aggregation function.

To better learn the global representation, we adopt the stack layer to broaden the entity's receptive field. To elaborate further, given a total of  $L$  propagation layers, the representation of entity in the  $l$ -th layer is denoted as follows:

$$e_i^l = \xi(e_{\mathcal{N}_i}^{l-1}, e_i^{l-1}), \quad (11)$$

where  $e_i^0$  is equal to the initial embedding  $e_{\text{init}}$  generated by the drug representation initialization module.

**3.6. DDI Prediction Module.** Due to the challenge of node smoothing that emerges with the growing depth of graph neural networks, we introduce a residual layer to mitigate this problem. In this layer, we combine the initialized drug representation  $e_{\text{init}}$  with the global drug representation  $e_i^l$  to obtain the final drug embedding vector.

$$e_{\text{final}} = e_{\text{init}} + e_i^l. \quad (12)$$

Finally, we generate the final representations of the drug pair, employing it as an input to function  $f$  for getting the prediction scores. Our objective is to minimize the discrepancy between the predictions and the actual labels:

$$\text{Loss} = \sum_{(i,j) \in Y} -y_{i,j} \log \hat{y}_{i,j} - (1 - y_{i,j}) \log (1 - \hat{y}_{i,j}), \quad (13)$$

where  $y_{i,j}$  is the real label for the drug pair.

## 4. Experiment

**4.1. Baselines.** To illustrate the superior performance of our method, we use some state-of-art models as the baselines:

- (i) Laplacian [11]: It is an exemplary matrix factorization method, which generates the network embedding by factorizing the input data matrix into lower dimensional matrices

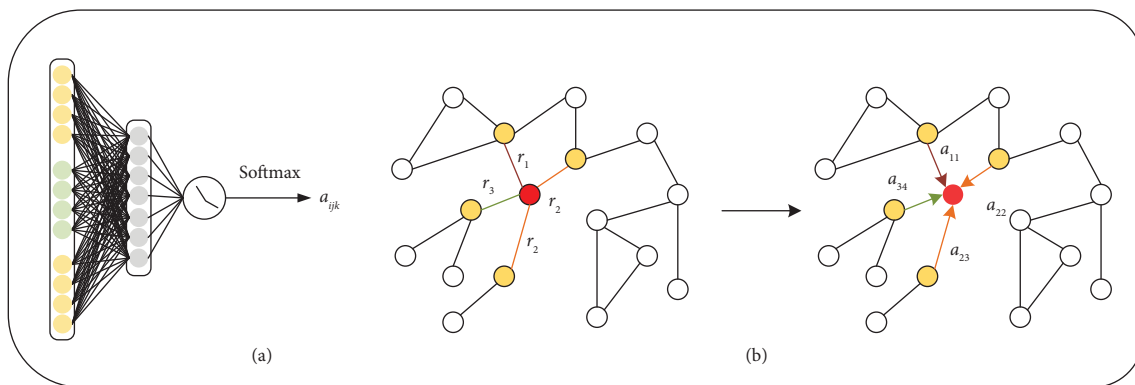


FIGURE 4: The illustration of knowledge-aware attention graph network. (a) The knowledge-aware attention mechanism, where orange circles represent the entity embedding vector and green circles represent the relation embedding vectors. (b) The neighbor sampling and information propagation of red nodes.

- (ii) DeepWalk [14]: This method employs the random walk to generate node sequences, which are subsequently fed to the skip-gram algorithm [44] for node embedding learning
- (iii) LINE [45]: This method uses the neural network and incorporates local and global graph information for node embedding learning
- (iv) TransE [46]: It assesses the credibility of a fact by considering the proximity between two entities, typically following a transformation performed by the relation
- (v) ComplEx [18]: It measures the plausibility of facts by aligning latent semantics of entities and relations, which are embedding into the vector space representations
- (vi) KGNN [19]: This method employs a graph neural network on the knowledge graph for capture both structural information and rich semantic features
- (vii) DDKG [23]: It acquires the drug embedding from their attributes within the knowledge graph. Subsequently, it leverages an attention mechanism to consider neighboring node embeddings and triple facts
- (viii) LaGAT [47]: It is a link-aware graph attention method, which considers various links between drug pairs in the knowledge graph to create multiple attention pathways for the drug entity

**4.2. Experimental Settings.** In the experiment, we fixed the embedding size to 32 for all methods. For the Laplacian method, the DeepWalk method, and the LINE method, we re-implement the BioNEV toolkit [48] to generate the node embedding. We also train the TransE and the ComplEx based on the Dgl-ke toolkits [49] and generate node embedding of the knowledge graph. For KGNN and DDKG, we configure the parameters to match their original settings as described in their original works, respectively. As the parameters of our work, we set the number of depths to 2 and the neighborhood size to 4, the batch-size is 4096. We adopt

the Adam algorithm to optimize all trainable parameters. For the LaGAT method, we set some parameters to compare with our method, e.g., the number of neighbor samples and the number of depths. For both datasets, we utilize all drug pairs that have been approved to interact as positive samples and randomly allocate them into training, validation, and test sets in an 8 : 1 : 1 ratio. Furthermore, we randomly extract an equivalent quantity of data from the complementary set of positive samples as the negative samples.

## 5. Results and Discussion

In this section, we show and analyse the performance of our method and all baselines with 5-fold cross-validation. Several metrics are employed to evaluate the prediction performance, including area under precision-recall (AUPR), accuracy (Acc.), F1 scores, and area under curve (AUC). Table 2 indicates the results of those metrics on the KEGG-drug dataset and the OGB-biokg dataset. The AUC and AUPR curves of each algorithm are depicted in Figures 5 and 6.

Table 2 shows that our approach demonstrates a significant performance advantage over the baseline methods across two datasets, as evidenced by superior results in four metrics. On the KEGG-drug dataset, our method improved by 0.6% on AUC, 0.93% on Acc., 0.89% on F1 scores and 0.93% on AUPR compared to the LaGAT, which achieved the best result of all baselines. On the OGB-biokg dataset, the improvement in AUC and AUPR is not obvious, only 0.07 and 0.08. On the F1 score and ACC indicators, our method improved by 0.41%. This can be attributed to our model's integration of drug attribute information, including molecular graph and drug sequence data, into the representation learning process. In addition, LAP, DeepWalk and LINE show a relatively poor performance compared to other baselines, the reason is that these approaches only focus on the topological properties of biomedical networks and lack the wealth of attribute information presented in biomedical entities and their relationships. ComplEx and TransE perform better performance than LAP, DeepWalk and LINE, but not as well as the other baselines. This is because

TABLE 2: Temperature and wildlife count in the three areas covered by the study.

Methods	KEGG				OGB-biokg			
	AUC	Acc	F1	AUPR	AUC	Acc	F1	AUPR
LAP	0.8677	0.7981	0.8047	0.8354	0.9428	0.9053	0.8992	0.9313
DeepWalk	0.9121	0.8396	0.8451	0.8980	0.9900	0.9572	0.9581	0.9887
LINE	0.9040	0.8273	0.8338	0.8896	0.9910	0.9599	0.9607	0.9898
TransE	0.9319	0.8594	0.8651	0.9208	0.9930	0.9703	0.9708	0.9908
Complex	0.9454	0.8753	0.8812	0.9364	0.9942	0.9710	0.9715	0.9924
KGNN	0.9414	0.8845	0.8899	0.9187	0.9949	0.9715	0.972	0.9943
DDKG	0.9534	0.9012	0.9041	0.9359	0.9951	0.973	0.9733	0.9945
LaGAT	0.9643	0.9154	0.9175	0.9518	0.9952	0.9737	0.9739	0.9946
Ours	0.9705	0.9240	0.9257	0.9611	0.9959	0.9777	0.9779	0.9954

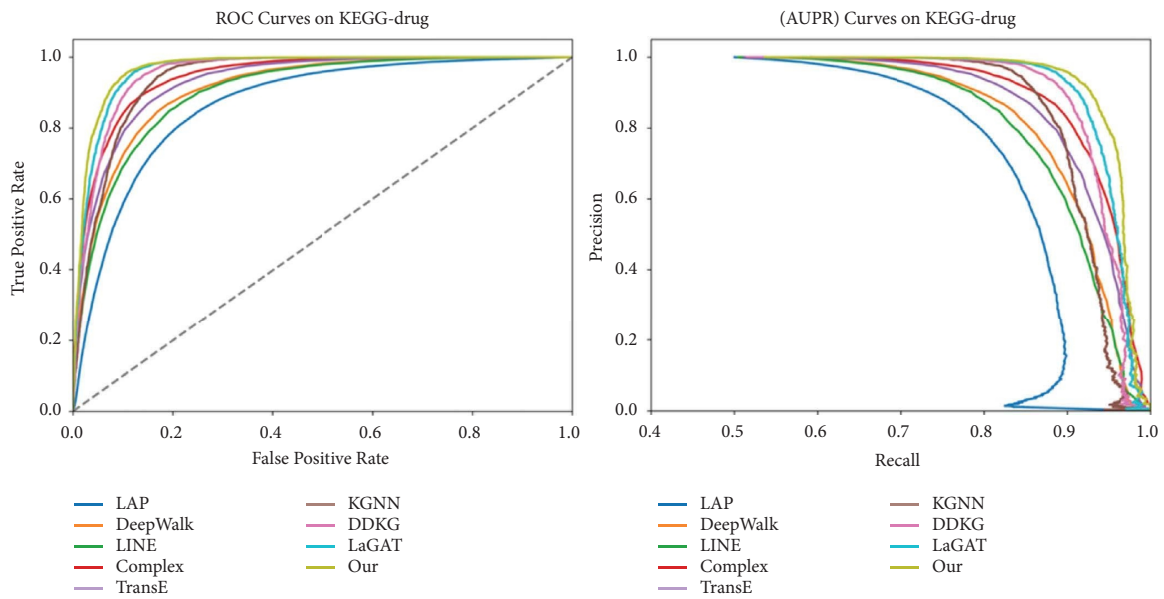


FIGURE 5: The AUC and AUPR curves of each algorithm on KEGG-drug dataset.

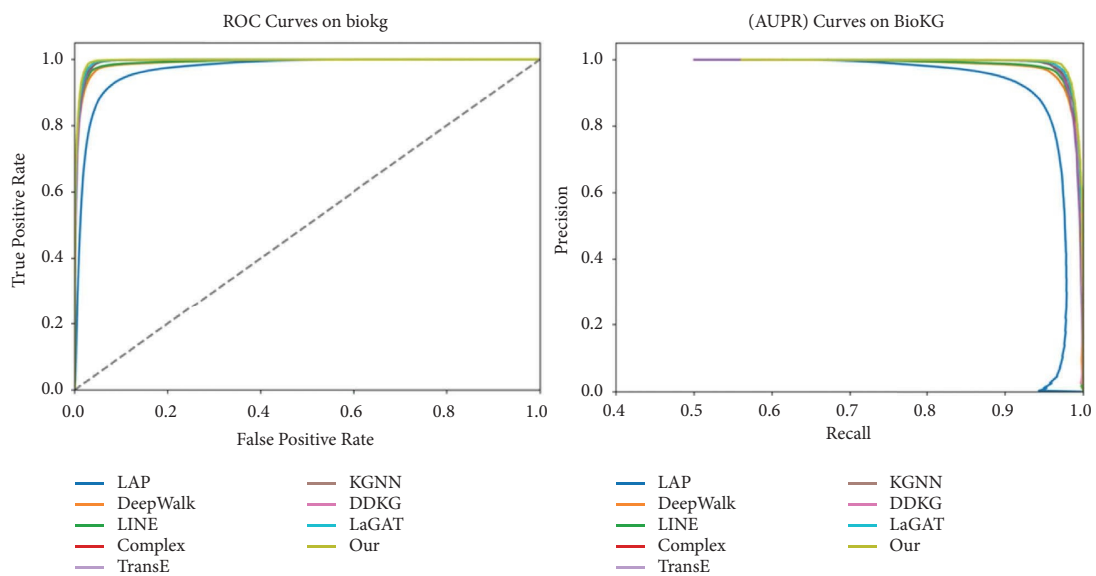
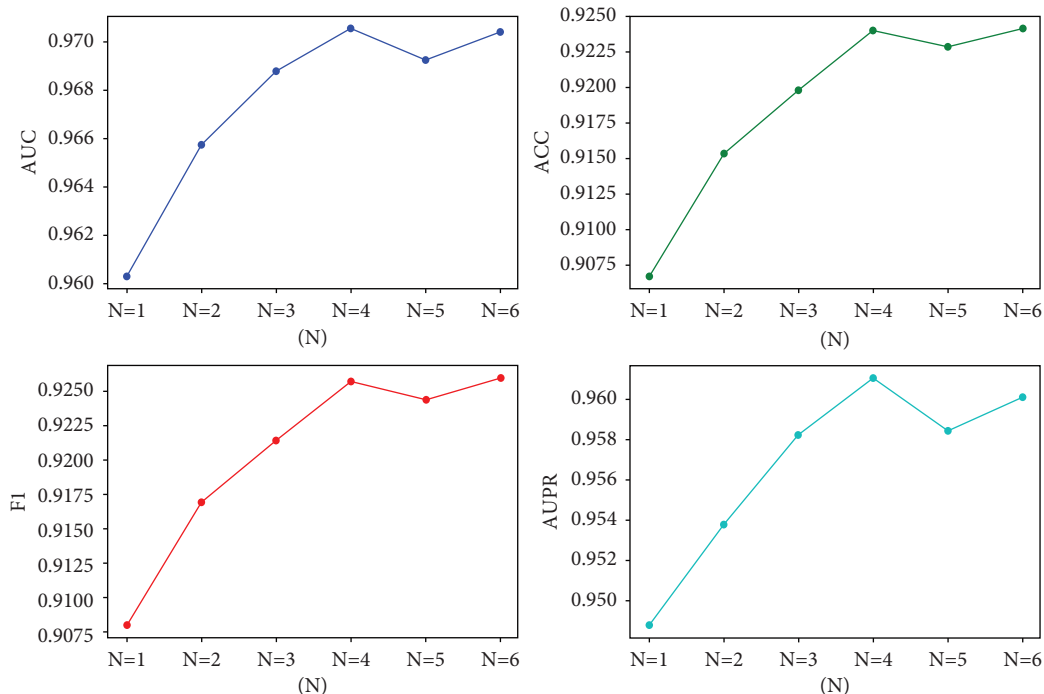


FIGURE 6: The AUC and AUPR curves of each algorithm on OGB-biokg dataset.

TABLE 3: Ablation experimental results.

Method	KEGG-drug				OGB-biokg			
	AUC	Acc	F1	AUPR	AUC	Acc	F1	AUPR
Ours	0.9705	0.9240	0.9257	0.9611	0.9959	0.9778	0.978	0.9953
W/o residual	0.9516	0.9005	0.9035	0.9292	0.9949	0.9729	0.9732	0.9943
W/o attention	0.9663	0.9161	0.9175	0.9555	0.9957	0.9765	0.9768	0.9950
W/o seq	0.9413	0.8894	0.8948	0.9146	0.9947	0.9717	0.9720	0.9939
W/o mole	0.9639	0.9125	0.9141	0.9514	0.9953	0.9749	0.9752	0.9947
W/o both	0.9442	0.8915	0.8921	0.9190	0.9952	0.9735	0.9722	0.9963

FIGURE 7: The performance of the different  $N$  values.

Complex and TransE take both entities and relationships into account and consider the 1-hop information of KGs to perform the representation learning. In contrast, other baseline methods (e.g. KGNN, DDKG) employ spatial-based graph convolutional networks (GCN) to capture multihop information. Furthermore, DDKG performs better than KGNN because it not only considers  $t$  the topological neighbourhood representations of drugs and relationships but also incorporates drug attribute information. DDKG also does not outperform our method. Based on DDKG, we excavate the information of the molecular structure information and take it as an additional attribute of drugs in KGs. Besides, we adopt the residual mechanism to avoid node smoothing caused by graph neural network.

**5.1. Ablation Study.** As mentioned before, the core of our work is to incorporate the molecular graph as another drug attribute in KGs and to utilize an attention mechanism that considers the features of both entities (nodes) and relations (edges) in a multihop neighbourhood of a designated entity/node. Additionally, the residual mechanism is adopted to

reduce node smoothing caused by graph neural networks. To investigate the effectiveness of the central idea, additional variant experiments have been implemented. Their brief descriptions are given as below:

- (i) w/o residual mechanism: It removes the residual mechanism compared with the origin model
- (ii) w/o attention: This exam uses the same attention mechanism as DDKG compared with the origin model
- (iii) w/o seq: It only uses the information of the drug molecular graph as the initial feature
- (iv) w/o mole: It only uses the information of drug sequence as the initial feature
- (v) w/o both: It does not use the information of drug sequence or information of the drug molecular graph as the initial feature

Table 3 exhibits the experimental results. From Table 3, all variant experiments show slightly lower results than the original model. When the residual mechanism is removed,



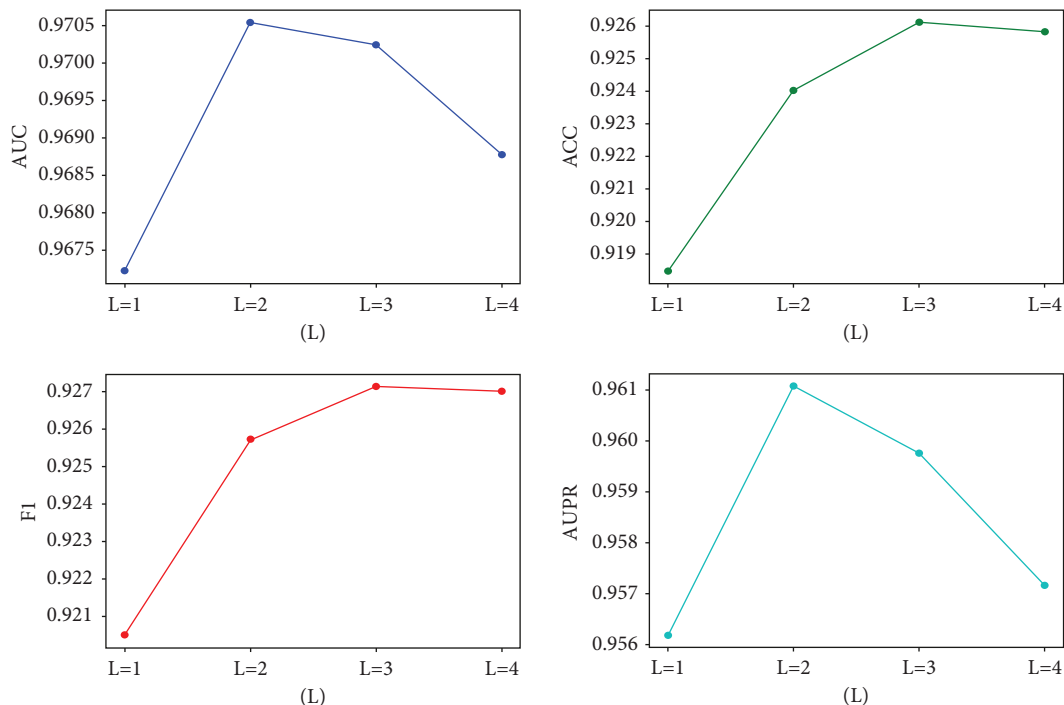


FIGURE 8: The performance of the different  $L$  values.

the performance decreases uniformly on different datasets. This indicates that the residual mechanism can reduce node smoothing by combining the attribute information of the drug. Moreover, the use of the attention mechanism in DDKG has a lower score than our method. The attention mechanism used in our method tries to consider both entity and relation features in a multihop neighbourhood and capture the diverse roles that an entity plays in different relations. To some extent, this method enhances the expressive capacity of drug representations. In the absence of drug sequence information or molecular graph information, the performance is an unavoidable loss in two datasets, confirming the importance of integrating drug attribute information when initializing drug embeddings.

**5.2. Parameter Analysis.** Here, we will investigate how several key parameters affect the performance of our approach. When examining a particular parameter, the remaining parameters are fixed. Figure 4 displays the value of AUC, Acc, F1 scores, and AUPR for KEGG-drug datasets.

**5.2.1. Impact of Neighbourhood Size.** First, we examined the influence of different values of  $N$ , which represents the number of neighboring nodes selected, ranging from 1 to 5. The results corresponding to the different values are depicted in Figure 7. We observe that (1) our model performs the best when  $N$  is set to 4. (2) The unsatisfactory performance observed with small values of  $N$  can be attributed to their limited capacity to capture enough neighboring nodes during the information aggregation process. (3) When  $N$  is set to 5, the performance begins to deteriorate due to the presence of unexpected noise.

**5.2.2. Impact of the Depth of Receptive Field.** Second, the effect of the depth of the receptive field, denoted as  $L$ , is investigated through a series of experiments where  $L$  is varied from 1 to 4. Figure 8 demonstrates the result. It is noted that the performance of the proposed model is better when  $L$  is set to 2. As the value of  $L$  increases, the perceptual path for each drug node becomes longer, causing an influx of noisy data. This will adversely affect the performance. Moreover, a higher value of  $L$  will require more CPU time and a greater number of epochs to achieve convergence. According to the experimental results, we imply  $L$  to be 2.

## 6. Conclusions

In this paper, we develop a new attention-based KG representation learning framework for drug-drug interaction. First, by taking the sequence and molecular graph information of drugs as the attributes of drug nodes in KGs, we use the drug representation initialization module to acquire the initial embeddings of drugs. Then, we adopt a knowledge-aware graph attention network that considers the features of both entities (nodes) and relations (edges) in a multi-hop neighbourhood of a given entity/node. Based on this, the proposed model can capture the topological structure information and the semantic relationship within KGs. Furthermore, the residual mechanism is added to the proposed model to avoid the node smoothing caused by GNN. Finally, we implemented the proposed model and conducted extensive experiments using two benchmark biomedical KG datasets. The results of these experiments consistently demonstrate the superior performance of our model compared to several state-of-the-art drug-drug interaction prediction models across a spectrum of evaluation metrics.

## Data Availability

The dataset and source code can be accessed at <https://github.com/Nokeli/MFIKGDDI>. All datasets in this study are from public resource. KEGG is available at <https://www.genome.jp/kegg/>. DrugBank is available at <https://go.drugbank.com/releases/5-1-10/downloads/all-full-database>; OGB-biokg is available at <https://ogb.stanford.edu/docs/linkprop/>.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported in part by the Nature Science Fund for Distinguished Young Scholars of China (62325308).

## References

- [1] G. Lee, C. Park, and J. Ahn, "Novel deep learning model for more accurate prediction of drug-drug interaction effects," *BMC Bioinformatics*, vol. 20, pp. 1–8, 2019.
- [2] S. Vilar, E. Uriarte, L. Santana et al., "Similarity-based modeling in large-scale prediction of drug-drug interactions," *Nature Protocols*, vol. 9, no. 9, pp. 2147–2163, 2014.
- [3] A. Kumar, P. P. Bansod, and S. Kansal, "Prediction of instantaneous heart rate using adaptive algorithms," *International Journal of Adaptive and Innovative Systems*, vol. 2, no. 4, pp. 267–281, 2019.
- [4] M. Asada, M. Miwa, and Y. Sasaki, "Enhancing drug-drug interaction extraction from texts by molecular structure information," 2018, <https://arxiv.org/abs/1805.05593>.
- [5] A. K. Nyamabo, H. Yu, Z. Liu, and J.-Y. Shi, "Drug–drug interaction prediction with learnable size-adaptive molecular substructures," *Briefings in Bioinformatics*, vol. 23, no. 1, Article ID bbab441, 2022.
- [6] T. Takeda, M. Hao, T. Cheng, S. H. Bryant, and Y. Wang, "Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge," *Journal of Cheminformatics*, vol. 9, no. 1, pp. 16–19, 2017.
- [7] X. Lin, Z. Quan, Z.-J. Wang, H. Huang, and X. Zeng, "A novel molecular representation with BiGRU neural networks for learning atom," *Briefings in Bioinformatics*, vol. 21, no. 6, pp. 2099–2111, 2020.
- [8] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling poly-pharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.
- [9] C. Yan, G. Duan, Y. Zhang, F.-X. Wu, Y. Pan, and J. Wang, "Predicting drug-drug interactions based on integrated similarity and semi-supervised learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 168–179, 2022.
- [10] P. Li, C. Huang, Y. Fu et al., "Large-scale exploration and analysis of drug combinations," *Bioinformatics*, vol. 31, no. 12, pp. 2007–2016, 2015.
- [11] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [12] S. Cao, W. Lu, and Q. Xu, "Grarep: learning graph representations with global structural information," in *Proceedings of the 24th ACM international on conference on information and knowledge management*, pp. 891–900, Melbourne Australia, October 2015.
- [13] W. Zhang, Y. Chen, D. Li, and X. Yue, "Manifold regularized matrix factorization for drug-drug interaction prediction," *Journal of Biomedical Informatics*, vol. 88, pp. 90–97, 2018.
- [14] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710, New York, NY, USA, August 2014.
- [15] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [16] R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, and M. Dumontier, "Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings," *BMC Bioinformatics*, vol. 20, no. 1, pp. 726–814, 2019.
- [17] M. R. Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, and S. Decker, "Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network," in *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*, pp. 113–123, Niagara Falls, NY, USA, September 2019.
- [18] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proceedings of the International conference on machine learning*, pp. 2071–2080, New York, NY, USA, June 2016.
- [19] X. Lin, Z. Quan, Z.-J. Wang, T. Ma, and X. Zeng, "KGNN: knowledge graph neural network for drug-drug interaction prediction," in *Proceedings of the IJCAI*, pp. 2739–2745, Yokohama, Japan, January 2020.
- [20] L. Hu, X. Pan, Z. Tang, and X. Luo, "A fast fuzzy clustering algorithm for complex networks via a generalized momentum method," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 9, pp. 3473–3485, 2022.
- [21] L. Hu, X. Pan, H. Yan, P. Hu, and T. He, "Exploiting higher-order patterns for community detection in attributed graphs," *Integrated Computer-Aided Engineering*, vol. 28, no. 2, pp. 207–218, 2021.
- [22] L. Hu, K. C. Chan, X. Yuan, and S. Xiong, "A variational Bayesian framework for cluster analysis in a complex network," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 11, pp. 2115–2128, 2020.
- [23] X. Su, L. Hu, Z. You, P. Hu, and B. Zhao, "Attention-based knowledge graph representation learning for predicting drug-drug interactions," *Briefings in Bioinformatics*, vol. 23, no. 3, Article ID bbac140, 2022.
- [24] A. Deac, Y.-H. Huang, P. Veličković, P. Liò, and J. Tang, "Drug-drug adverse effect prediction with graph co-attention," 2019, <https://arxiv.org/abs/1905.00534>.
- [25] S. Pang, Y. Zhang, T. Song, X. Zhang, X. Wang, and A. Rodriguez-Patón, "AMDE: a novel attention-mechanism-based multidimensional feature encoder for drug–drug interaction prediction," *Briefings in Bioinformatics*, vol. 23, no. 1, Article ID bbab545, 2022.
- [26] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proceedings of the international conference on machine learning*, pp. 1263–1272, Sydney, Australia, August 2017.

- [27] A. Yüksel, E. Ulusoy, A. Ünlü, and T. Doğan, “Selfformer: molecular representation learning via selfies language models,” *Machine Learning: Science and Technology*, vol. 4, no. 2, Article ID 025035, 2023.
- [28] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, and C. Friedman, “Drug–drug interaction through molecular structure similarity analysis,” *Journal of the American Medical Informatics Association*, vol. 19, no. 6, pp. 1066–1074, 2012.
- [29] F. Cheng and Z. Zhao, “Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties,” *Journal of the American Medical Informatics Association*, vol. 21, no. e2, pp. e278–e286, 2014.
- [30] R. Ferdousi, R. Safdari, and Y. Omid, “Computational prediction of drug–drug interactions based on drugs functional similarities,” *Journal of Biomedical Informatics*, vol. 70, pp. 54–64, 2017.
- [31] H. Huang, P. Zhang, X. A. Qu, P. Sanseau, and L. Yang, “Systematic prediction of drug combinations based on clinical side-effects,” *Scientific Reports*, vol. 4, no. 1, p. 7160, 2014.
- [32] J. Y. Ryu, H. U. Kim, and S. Y. Lee, “Deep learning improves prediction of drug–drug and drug–food interactions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 18, pp. E4304–E4311, 2018.
- [33] N. Xu, P. Wang, L. Chen, J. Tao, and J. Zhao, “Mr-gnn: multi-resolution and dual graph neural network for predicting structured entity interactions,” 2019, <https://arxiv.org/abs/1905.09558>.
- [34] Y.-H. Feng and S.-W. Zhang, “Prediction of drug–drug interaction using an attention-based graph neural network on drug molecular graphs,” *Molecules*, vol. 27, no. 9, p. 3004, 2022.
- [35] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu, “Asymmetric transitivity preserving graph embedding,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1105–1114, San Francisco, CA, USA, August 2016.
- [36] A. Grover and J. Leskovec, “node2vec: scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, San Francisco, CA, USA, August 2016.
- [37] S. Liu, Y. Zhang, Y. Cui et al., “Enhancing drug–drug interaction prediction using deep attention neural networks,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 2, pp. 976–985, 2023.
- [38] Y.-H. Feng, S.-W. Zhang, Q.-Q. Zhang, C.-H. Zhang, and J.-Y. Shi, “deepMDDI: a deep graph convolutional network framework for multi-label prediction of drug–drug interactions,” *Analytical Biochemistry*, vol. 646, Article ID 114631, 2022.
- [39] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, “KEGG: kyoto encyclopedia of genes and genomes,” *Nucleic Acids Research*, vol. 27, no. 1, pp. 29–34, 1999.
- [40] W. Hu, M. Fey, M. Zitnik et al., “Open graph benchmark: datasets for machine learning on graphs,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 22118–22133, 2020.
- [41] D. S. Wishart, Y. D. Feunang, A. C. Guo et al., “DrugBank 5.0: a major update to the DrugBank database for 2018,” *Nucleic Acids Research*, vol. 46, no. D1, pp. D1074–D1082, 2018.
- [42] M. Krenn, F. Häse, A. Nigam, P. Friederich, and A. Aspuru-Guzik, “Self-referencing embedded strings (SELFIES): a 100% robust molecular string representation,” *Machine Learning: Science and Technology*, vol. 1, no. 4, Article ID 045024, 2020.
- [43] Y. Chen, T. Ma, X. Yang, J. Wang, B. Song, and X. Zeng, “MUFFIN: multi-scale feature fusion for drug–drug interaction prediction,” *Bioinformatics*, vol. 37, no. 17, pp. 2651–2658, 2021.
- [44] K. W. Church, “Word2Vec,” *Natural Language Engineering*, vol. 23, no. 1, pp. 155–162, 2017.
- [45] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “Line: large-scale information network embedding,” in *Proceedings of the 24th international conference on world wide web*, pp. 1067–1077, Florence, Italy, May 2015.
- [46] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi-relational data,” *Advances in Neural Information Processing Systems*, vol. 26, 2013.
- [47] Y. Hong, P. Luo, S. Jin, and X. Liu, “LaGAT: link-aware graph attention network for drug–drug interaction prediction,” *Bioinformatics*, vol. 38, no. 24, pp. 5406–5412, 2022.
- [48] X. Yue, Z. Wang, J. Huang et al., “Graph embedding on biomedical networks: methods, applications and evaluations,” *Bioinformatics*, vol. 36, no. 4, pp. 1241–1251, 2020.
- [49] D. Zheng, X. Song, C. Ma et al., “Dgl-ke: training knowledge graph embeddings at scale,” in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 739–748, China, July 2020.