JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

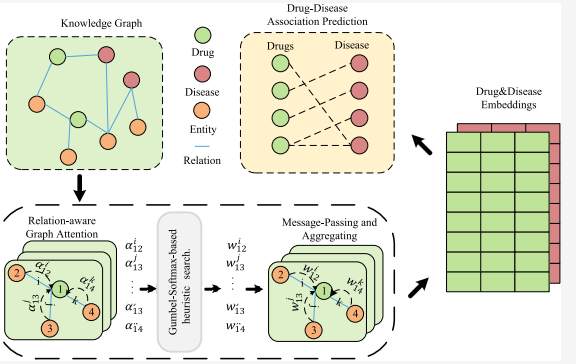
Knowledge Graph Convolutional Network with Heuristic Search for Drug Repositioning

Xiang Du, Xinliang Sun, and Min Li*

Cite This: J. Chem. Inf. Model. 2024, 64, 4928−4937 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Drug repositioning is a strategy of repurposing approved drugs for treating new indications, which can accelerate the drug discovery process, reduce development costs, and lower the safety risk. The advancement of biotechnology has significantly accelerated the speed and scale of biological data generation, offering significant potential for drug repositioning through biomedical knowledge graphs that integrate diverse entities and relations from various biomedical sources. To fully learn the semantic information and topological structure information from the biological knowledge graph, we propose a knowledge graph convolutional network with a heuristic search, named KGCNH, which can effectively utilize the diversity of entities and relationships in biological knowledge graphs, as well as topological structure information, to predict the associations between drugs and diseases. Specifically, we design a relation-aware attention mechanism to compute the attention scores for each neighboring entity of a given entity under different relations. To address the challenge of randomness of the initial attention scores potentially impacting model performance and to expand the search scope of the model, we designed a heuristic search module based on Gumbel-Softmax, which uses attention scores as heuristic information and introduces randomness to assist the model in exploring more optimal embeddings of drugs and diseases. Following this module, we derive the relation weights, obtain the embeddings of drugs and diseases through neighborhood aggregation, and then predict drug−disease associations. Additionally, we employ feature-based augmented views to enhance model robustness and mitigate overfitting issues. We have implemented our method and conducted experiments on two data sets. The results demonstrate that KGCNH outperforms competing methods. In particular, case studies on lithium and quetiapine confirm that KGCNH can retrieve more actual drug−disease associations in the top prediction results.



INTRODUCTION

Drug repositioning (also called drug repurposing, reprofiling, or retasking) is a strategy for identifying new uses for approved or investigational drugs that are outside the scope of the original medical indication.[1] Drug repositioning provides a safe and efficient way to facilitate the market entry of potential drugs, as repositioned drugs have already completed early trials and have proven to be sufficiently safe in preclinical models and humans. Therefore, drug repositioning requires less time and investment than developing new drugs for specific indications. For example, some repositioned drugs, such as remdesivir, ritonavir, and tocilizumab, have provided rapid response to the global coronavirus disease (COVID-19) pandemic,[2] indicating that drug repositioning is a promising strategy to fight against diseases.

With the continuous advancement of biomedical research and the gradual accumulation of drug and pathological information,[3] the scale of biological data is constantly increasing, including various omics data such as genomics, transcriptomics, proteomics, and metabolomics and related information such as diseases, drugs, nutrition, etc. For instance, in genomics, next-generation sequencing (NGS), the release of the first truly high-throughput sequencing platform in the mid-2000s, heralded a 50,000-fold drop in the cost of human genome sequencing. The amount of sequencing data is increasing exponentially.[4] In transcriptomics, the Human Protein Atlas[5] and the Genotype-Tissue Expression (GTEx)[6] project have generated atlases of gene expression of tissues, the brain, diseases, blood, and cells and quantified expression over common tissues, respectively. Given the diverse and complex biological data sources, a knowledge graph (KG), a semantic network comprising entities and their relations in the real world,[7] serves as an efficient means to integrate and store this information.[8] Formally, a KG is a type of labeled multigraph[9] that represents entities (also known as nodes or vertices) and their relations (also known as edges, facts, or links). Two

Received: April 28, 2024
Accepted: May 21, 2024
Published: June 5, 2024

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937

entities connected by a relation are also known as the head and tail or source node and target node.

Knowledge graph embedding (KGE) techniques aim to discover missing links between entities in a knowledge graph. KGE methods can be roughly classified into two main categories: distance-based methods and semantic-matching-based methods. Distance-based methods use translational operations to measure the distance between the head and the tail after applying the relation vector. The most representative method in this category is TransE,[10] which assumes that relations are translations from head entities to tail entities. However, TransE cannot handle complex relations such as one-to-many, many-to-one, or many-to-many. To overcome this limitation, several variants of TransE have been proposed, such as TransH,[11] TransR,[12] and TransD,[13] which use different strategies to map entities and relations to different spaces. Another distance-based method is RotatE,[14] which is inspired by Euler's identity and defines relations as rotations from heads to tails in the complex plane. Semantic matching-based methods use scoring functions to measure the semantic similarity between the head, the relation, and the tail.[15] RESCAL[16] is a seminal method in this category, which uses a bilinear scoring function to solve a three-way matrix factorization problem. However, RESCAL has a high computational cost and a large number of parameters, as it uses a dense matrix to represent each relation. DistMult[17] reduces the number of parameters by constraining the relation matrices to be diagonal. ComplEx[18] extends DistMult to the complex domain, which allows the same node to have different representations depending on its position. Specifically, if a node is represented as h when it is a head, then it is represented as $\overline{h}$ when it is a tail, where $\overline{h}$ is the complex conjugate of h. This enables ComplEx to model antisymmetric relations more effectively. QuatE[19] further extends ComplEx to the quaternion domain and defines relations as rotations of head entities. Then it computes the quaternion inner product between the rotated head and the tail as the score. The goal of semantic-matching-based methods is to maximize the scores of positive triples and minimize the scores of negative triples. The details and comparisons of these methods are provided in the overview in ref 20. Despite the performance of the KGE methods discussed above being promising, they often encounter challenges in extracting high-order nonlinear features, resulting in insufficient utilization of topological structure information.

Graph neural networks (GNNs) can effectively extract the structure and feature information on graph data. GNN[21] is proposed as the first graph neural network, which is based on information diffusion and relaxation mechanisms. Using the spectrum of the graph Laplacian, Bruna et al.[22] first proposed convolutional networks for graphs. In follow-up works, GCN[23] constructed a simple graph convolution via a localized first-order approximation. LightGCN[24] simplified the design of GCN by including only the most essential component in GCN—neighborhood aggregation. Inspired by the attention mechanisms, GAT[25] defined the graph attention networks, which learn the weight values between a node and its neighbors. BridgeDPI[26] first constructs virtual nodes to bridge the gap between drugs and proteins and then uses a Graph Neural Network to capture the network-level information among diverse drugs and proteins for predicting drug−protein interactions. CGraphDTA[27] designed a fusion protocol based on multiscale convolutional neural networks and graph neural networks, which can leverage both target sequence and structure for predicting drug−target binding affinity. GraphscoreDTA,[28] taking a combination of graph neural networks, bitransport information mechanisms, and physics-based distance terms, effectively captures mutual information between protein−ligand pairs for predicting protein−ligand binding affinity. Guo et al.[29] designed a variational autoencoder based on a gated mechanism and graph convolutions to extract multilevel dependency information for predicting disease−miRNA associations. PSGCN[30] is a GCN-based method that leverages the graph structure to capture the contextual information on drug−disease pairs for drug repositioning. DRGCL[31] treats known drug−disease associations as a topological graph and improves drug repositioning performance by constructing semantic graphs and applying graph contrastive learning. AdaDR[32] is an adaptive GCN method for drug repositioning that integrates both node features and topological structures and models interactive information between them with an adaptive graph convolution operation. However, most of the GNN-based methods discussed above are primarily designed for bipartite graphs, which do not require consideration of diverse types of entities and relations, thus limiting their performance on biomedical knowledge graphs.

In recent years, with the development of GNNs, many KG-based methods have adopted GNNs as a basic component. For example, in a recommendation system, KGCN[33] focuses on the user preference for the relations in the KG and learns item representations by aggregating and integrating neighborhood information with bias. KGAT[34] designs a relational attention mechanism that calculates attention scores according to the distance between head and tail entities in the relation space. KACL[35] employs an attention mechanism and contrastive learning to reduce recommendation-irrelevant information in knowledge graphs and alleviate interaction domination issues. In drug−drug interaction prediction, KGNN[36] introduced a framework that predicts drug−drug interactions by capturing drugs and their potential neighborhoods via mining their associated relations in a KG. In drug repositioning, DREAM-walk,[37] a method for drug repositioning leveraging biomedical knowledge graphs, employs semantic information-guided random walks to generate sequences of drug and disease nodes, addressing challenges such as inadequate representation due to gene dominance and the limited number of drug and disease entities. DRKF[38] first extracts relations related to Parkinson's disease from medical literature to construct a medical literature knowledge graph and then employs TransE and DistMult methods to predict drug repositioning candidates. Zhang et al.[39] proposed a neural network-based literature-based discovery approach, which uses SemRep, filtering rules, and an accuracy classifier developed on a BERT variant to construct triples from PubMed and other COVID-19-focused research literature, and then uses TransE, RotatE, DistMult, and ComplEx methods to predict drug repositioning candidates. However, in drug repositioning, more research focuses on how to extract effective information from the literature to construct knowledge graphs, and the models used for prediction have limitations mentioned previously.

To overcome the mentioned limitations, we proposed a knowledge graph convolutional network with a heuristic search (KGCNH) for drug repositioning. We summarize the contributions of this work as follows:

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937

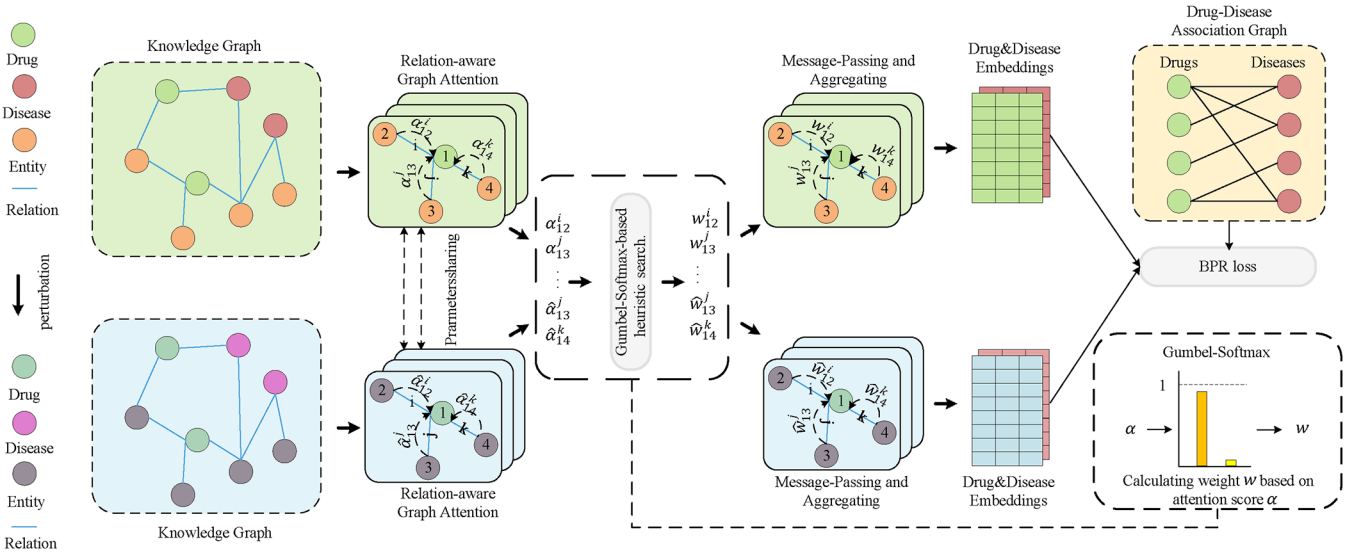


Figure 1. Overall architecture of KGCNH. KGCNH mainly consists of three modules: relation-aware attention mechanism, Gumbel-Softmax-based heuristic search, and feature-based graph augmentation.

- We introduce KGCNH, an novel end-to-end framework for drug repositioning that utilizes biomedical knowledge graphs. It employs a relation-aware attention mechanism to learn the importance of different neighbors of entities under various relations. KGCNH selectively aggregates neighborhood information based on learned attention scores, enhancing the influence of relevant neighbors while mitigating interference from irrelevant ones. Furthermore, to augment KGCNH's performance and robustness, we introduce feature-based enhanced views.
- We introduce randomness into our model by using the Gumbel-Softmax-based heuristic search module, which enables the model to explore the optimal embeddings of drugs and diseases.
- We conduct extensive experiments to evaluate the performance of our proposed method. The results show that KGCNH outperforms competitive methods. Moreover, the case studies present the potential of KGCNH for real-world application.

Table 1. Notations and Explanations

notations	explanations
G_d	drug-disease association graph
$U, V, \mathcal{E}$	set of drugs, diseases, and edges
A	adjacency matrix of drug−disease association graph
$\|U\|, \|V\|$	number of drugs and diseases
G_k	knowledge graph
E, R	set of entities and relations
$\|E\|$	number of entities
$\|R\|$	number of relation types
$y(u, v)$	score of drug u being associated with disease v
$\mathbf{e}, \mathbf{r}$	embeddings of entities and relations
$e_h, r(<h, t>), e_t$	entity h, relation type between h and t, entity t
$\mathbf{e}_h, \mathbf{r}_{r(<h,t>)}, \mathbf{e}_t$	embedding of entity h, relation between h and t and entity t
$\mathbf{e}^0$	initial entity embeddings
$\tilde{\mathbf{e}}$	embeddings of drugs and diseases for prediction
$N_{(h)}$	set of neighbor entities of h
α_{ht}	attention score of entity t to h
w_{ht}	relation weight of entity t to h

MATERIALS AND METHODS

In this section, we describe the problem formulation, the data sets, and the KGCNH framework. As Figure 1 depicts, our framework comprises three components: (i) a relation-aware attention mechanism, (ii) a Gumbel-Softmax-based heuristic search, and (iii) a feature-based graph augmentation. The main notations utilized throughout this article are presented in Table 1.

Problem Formulation. We begin by presenting the drug−disease association graph and knowledge graph and then describe the KG-based drug−disease association prediction task.

Drug−Disease Association Graph. We construct a bipartite drug−disease association graph, $G_d = \{U, V, \mathcal{E}\}$, where U represents drug nodes set in which $|U|$ is the number of drugs, V represents disease nodes set in which $|V|$ is the number of diseases, and $\mathcal{E} = \{(u, v)|u \in U, v \in V\}$ is the edge set representing observed associations between drugs and diseases. The adjacency matrix of the graph is $A \in \{0,1\}^{|U|\times|V|}$. In the matrix, for each entry A_{uv} in A, if $A_{uv} = 1$, it means that drug u is associated with disease v. Note that if $A_{uv} = 0$, it does not necessarily mean there is no association between u and v, as it may be an unobserved potential association.

Knowledge Graph. We also have a knowledge graph $G_k = \{E, R\}$, which is comprised of entity−relation−entity triples, where E represents set of entities in which $|E|$ is the number of entities and R represents the set of relations in which $|R|$ is the number of relation types. For any knowledge graph triple (h, r, t), it indicates that there is a relation from head entity h to tail entity t, where $h, t \in E$ and $r \in R$. For example, the triple (methotrexate, treats, muscle cancer) states the fact that methotrexate can be used to treat muscle cancer.

KG-Based Drug Repositioning. Given a drug−disease association graph G_d and knowledge graph G_k, we defined a score function $y(u, v) = f(u, v|\theta, G_d, G_k)$, where y_{uv} is the score of drug u being associated with disease v. The aim of the optimization is generally to score an observed association of

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937

drug−disease higher than an unobserved one. A higher score implies a higher likelihood of the drug treating the disease.

Data Sets. To evaluate the performance of our proposed model, we conducted experiments using the Hetionet data set[40] and the DrugBank data set from the DRKG.[41] Table 2 displays the overall statistical results of the data sets, and more detailed statistics are available in the Supporting Information S1.

Table 2. Statistics of Data Sets

	Hetionet	DrugBank
no. drugs	1538	9708
no. diseases	136	1182
no. associations	1145	4968
no. entities	45158	19911
no. relations	15	6
no. triples	2249052	1419822

Hetionet contains 11 types of entities and 24 relation types from 29 publicly available data sources. In order to reduce the computational complexity, we performed the following preprocessing on the Hetionet data set. We merge relations of the same name. For instance, the relation "upregulates" between compounds and genes, diseases and genes, and anatomy and genes is treated as the same relation by us. As a result, the number of relations decreased from 24 to 16. To balance the data, we changed the "palliates" relation of 390 triples to "treats". Thus, the preprocessed data contain 15 relations and 1145 associations between drugs and diseases. Ioannidis et al. constructed a biomedical knowledge graph based on the DrugBank database (version 5.1.5), which contains 6 relation types, 1419822 triples, 4968 drug−disease associations, and 19911 entities.

Relation-Aware Attention Mechanism. Inspired by the interpretability and effectiveness of GAT, many previous methods adopt GAT as the base architecture. However, the original graph attention mechanism ignores the influence of different types of relations on the attention scores. Therefore, we designed a relation-aware knowledge attention mechanism to calculate the attention score by incorporating the relation embedding. Formally, the attention score $\alpha_{ht} \in \mathbb{R}$ of the tail entity t to the head entity h under the relation type $r(\langle h, t\rangle)$ is computed as follows:

$$\bar{\alpha}_{ht} = \text{Leaky ReLU}(a^T W[(\mathbf{e}_h \odot \mathbf{r}_{r(\langle h,t\rangle)}) \| \mathbf{e}_t]) \tag{1}$$

$$\alpha_{ht} = \frac{\exp(\bar{\alpha}_{ht})}{\sum_{j \in N_{(h)}} \exp(\bar{\alpha}_{hj})} \tag{2}$$

where $\mathbf{e} \in \mathbb{R}^{|E| \times d}$ and $\mathbf{r} \in \mathbb{R}^{|R| \times d}$ are embeddings of entities and relations, respectively, which are randomly initialized and trainable, d is the dimension of embeddings, $\mathbf{e}_h$ and $\mathbf{e}_t$ denote the embeddings of head entity h and tail entity t, $\mathbf{r}_{r(\langle h,t\rangle)}$ represents the embedding of relation $r(\langle h, t\rangle)$, $\odot$ denotes the Hadamard product, $\|$ denotes the concatenation operation, $N_{(h)}$ denotes the set of neighbor entities of h, and a and W are trainable parameters.

Gumbel-Softmax-Based Heuristic Search. As previously mentioned, the embeddings of entities and relations are randomly initialized, resulting in random initial attention scores. Directly utilizing these attention scores may negatively impact the model's performance. To mitigate the effects of embedding initialization on model training and performance, we drew inspiration from the annealing algorithm, utilized the attention scores as heuristic information and introduced randomness through Gumbel-Softmax.[42] Specifically, the weight w_{ht} is calculated as follows:

$$w_{ht} = \frac{\exp((\log(\alpha_{ht}) + g_{ht})/\tau)}{\exp((\log(\alpha_{ht}) + g_{ht})/\tau) + \exp((\log(1 - \alpha_{ht}) + g'_{ht})/\tau)} \tag{3}$$

where g_{ht} and g'_{ht} are independent and identically distributed samples drawn from standard Gumbel distribution and τ is the temperature. The standard Gumbel distribution can be sampled using inverse transform sampling by drawing a $\epsilon \in \text{Uniform}(0, 1)$ and computing $g = -\log(-\log(\epsilon))$.

Gumbel-Softmax has two properties. One is that for low temperature ($\tau \leq 1$) the expected value of Gumbel-Softmax random variables is close to the expected value of categorical variables, i.e., the expected value of w_{ht}, $\mathbb{E}(w_{ht}) \approx \alpha_{ht}$. As the temperature increases, the expected value of Gumbel-Softmax random variables converges to a uniform distribution, i.e., $\mathbb{E}(w_{ht}) \approx 0.5$. The other property is that when temperature τ goes to 0, samples from Gumbel-Softmax distributions move toward binary, i.e., w_{ht} is either close to 0 or close to 1; as temperature τ increases without bound, samples converge toward a uniform distribution, i.e., w_{ht} is close to 0.5.

Considering the properties of the Gumbel-Softmax function, we employed a higher temperature coefficient during the initial stages of model training. As training progresses, the temperature gradually decreases, which is akin to the annealing algorithm. At higher temperatures, the sampled value w_{ht} is more uniformly distributed, resulting in a smaller impact of the attention score α_{ht} on w_{ht}. As the temperature decreases, the expected value of w_{ht} approaches the attention score α_{ht}. By incorporating randomness with the Gumbel-Softmax distribution, the model can explore more optimal embeddings of drugs and diseases.

For each entity h, the representation of its neighbors through the linear combination is computed as follows:

$$\mathbf{e}_{N_{(h)}} = \sum_{t \in N_{(h)}} w_{ht} \mathbf{e}_t \tag{4}$$

Next, we used Aggregator inspired by LightGCN, which discards the feature transformation and nonlinear activation to update the representation of entity h. More formally, the representation of entity h at the lth layer is updated as

$$\mathbf{e}_h^l = \mathbf{e}_h^{l-1} + \mathbf{e}_{N_{(h)}}^{l-1} \tag{5}$$

where $\mathbf{e}_h^{l-1}$ and $\mathbf{e}_{N_{(h)}}^{l-1}$ are the representations of h and its neighbors at the $(l-1)$th graph propagation layer. For $l = 1$, the representations of entities $\mathbf{e}^0$ are randomly initialized embeddings. At the lth graph propagation layer, which is denoted as the final layer of graph propagation, the representations of entities are given in e^l. Subsequently, we extract the representations of drugs and diseases from e^l, denoted as $\hat{\mathbf{e}}$.

Feature-Based Graph Augmentation. To enhance model performance and mitigate overfitting, we employed feature-based graph augmentation to enrich the representation of drugs and diseases by creating new features from existing ones. Specifically, we randomly initialized the embeddings of entities $\mathbf{e}^0$, and then applied aggregation to obtain the

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937

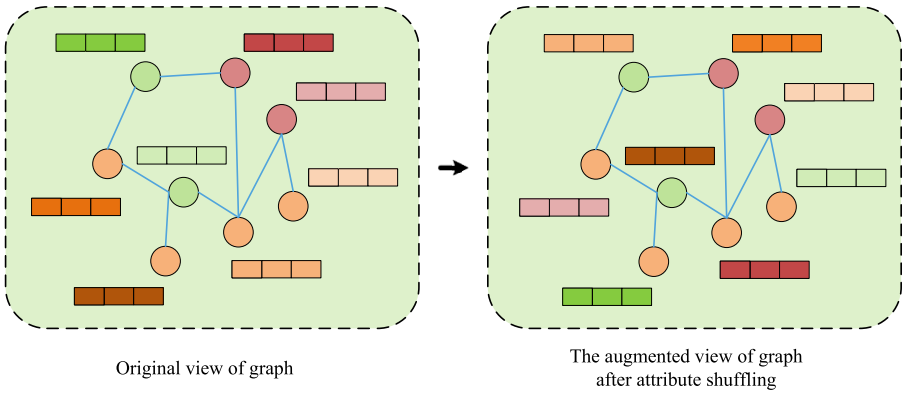


Figure 2. A toy example of generating an augmented view of a graph by using attribute shuffling.

embeddings of drugs and diseases $\hat{\mathbf{e}}_{v1}$. We subsequently introduced perturbation to $\mathbf{e}^0$ using attribute shuffling to construct the feature-based augmentation view. This augmented view, with the same topological structure as the original graph but with permuted order of entity embeddings, was then fed into the parameter-shared graph neural networks, resulting in the augmented embeddings of drugs and diseases $\hat{\mathbf{e}}_{v2}$. As shown in Figure 2, this is a toy example demonstrating the generation of an augmented view of a graph through attribute shuffling.

Model Training. We simply concatenated embeddings $\hat{\mathbf{e}}_{v1}$ and $\hat{\mathbf{e}}_{v2}$ to obtain the final representation $\tilde{\mathbf{e}}$ of drugs and diseases for predicting drug−disease associations. We employed the inner product as a score function, and drug−disease association score $y(u, v)$ is computed as follows:

$$y(u, v) = \tilde{\mathbf{e}}_u^T \tilde{\mathbf{e}}_v \tag{6}$$

The BPR loss[43] is a commonly employed method for optimizing models to capture pairwise associations. It is defined as follows:

$$L_{\text{bpr}}(u, v^+, v^-) = -\log \sigma(y(u, v^+) - y(u, v^-)) \tag{7}$$

where (u, v^+) represents the observed association in the drug−disease association graph, (u, v^-) denotes an unobserved association with $A_{uv^-} = 0$, and σ denotes the sigmoid function. We utilize the Adam algorithm to optimize the trainable parameters.

RESULTS AND DISCUSSIONS

Evaluation Metrics. We conducted 10-fold cross-validation to evaluate our approach. We treat all of the known drug−disease associations as positive samples and randomly divided them into ten equally sized subsets. In each fold, we used nine subsets as the positive training set and the remaining subset as the positive testing set. Then, we constructed negative samples for both the positive training and the test sets. Using the training set as an example, in each training iteration, for every positive sample, n negative samples are generated by randomly pairing drugs and diseases that are not associated. Importantly, these negative samples share the same drug as the positive sample but differ from each other. Through the merging of positive and negative samples, a training set is generated, maintaining a positive to negative sample ratio of 1:n, where n is a hyperparameter. The ratio of positive to negative samples in the training set is specified in the hyperparameters' setting section. The test set had a balanced ratio of 1:1. We employed the Area Under the Receiver Operating Characteristic curve (AUROC) and the Area Under the Precision-Recall curve (AUPRC) as evaluation metrics to assess the model performance, given their common utilization in drug repositioning prediction tasks. To ensure reliable performance estimation, we iteratively conducted the 10-fold cross-validation procedure ten times and subsequently reported the average results.

Baseline Methods. To evaluate the performance of our proposed model, we compared KGCNH with the seven methods listed below.

- RotatE[14] treats each relation as a rotation from the head to the tail entity in the complex vector space.
- ComplEx[18] extends DistMult to the complex domain, which allows the same node to have different representations depending on its position. This enables ComplEx to model antisymmetric relations more effectively.
- QuatE[19] further extends ComplEx to the quaternion domain and defines each relation as a quaternion rotation from the head to the tail in the quaternion space.
- GCN[23] is a model that applies neural networks to graph data. GCN constructed a simple graph convolution via a localized first-order approximation.
- LightGCN[24] simplifies the design of GCN by including only the most essential component in GCN−neighborhood aggregation.
- GAT[25] is inspired by the attention mechanisms and defines the graph attention networks, which learn the weight values between a node and its neighbors.
- KGNN[36] introduces a framework that predicts DDIs by capturing drugs and their potential neighborhoods via mining their associated relations.

Hyperparameter Setting. In the KGCNH model, the embedding dimension is set to 64 with a dropout rate of 0.2. For the KGCNH model, the learning rates on the Hetionet and DrugBank data sets are 0.002 and 0.0045, respectively, with optimal positive-to-negative sample ratios of 1:10 and 1:50. The results of KGCNH performance under various positive-to-negative sample ratios are provided in Supporting Information S2. Detailed hyperparameters for the baseline models are provided in Supporting Information S3.

Performance of KGCNH in the Cross-Validation. For a fair comparison, we reported the average results of 10 times of 10-fold cross-validation on two data sets to demonstrate result stability. As shown in Table 3, we emphasize the best results in

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937

Table 3. Performance Comparison of 10 Times 10-Fold Cross-Validation between Our Method and Baselines over Hetionet and DrugBank Data Sets

	Hetionet		DrugBank	
model	AUROC	AUPRC	AUROC	AUPRC
ComplEX	0.8073	0.7954	0.8320	0.8800
RotatE	0.7627	0.7279	0.7867	0.8311
QutatE	0.7334	0.7353	0.8106	0.8592
GCN	0.8770	0.8671	0.8371	0.8350
LightGCN	0.9094	0.9046	0.8158	0.8526
GAT	0.9058	0.9035	0.8537	0.8799
KGNN	0.7680	0.7487	0.8035	0.8105
KGCNH	**0.9367**	**0.9383**	**0.8834**	**0.9091**

bold. Notably, KGCNH consistently attained the best AUROC and AUPRC values over both data sets. Specifically, KGCNH achieves AUROC values of 0.9367 and 0.8834 on Hetionet and DrugBank, respectively, which are improvements of 3.00% and 3.47% compared to the second best methods LightGCN and GAT. For the AUPRC values, KGCNH outperforms all other methods, resulting in an average improvement of 3.73% over LightGCN on HetioNet and about 3.30% over ComplEx on DrugBank. The results of each 10-fold cross-validation are highly consistent, proving that our model has convincing performance and robustness(Figure 3). Additionally, to validate KGCNH's performance across diverse data sets, we assessed its efficacy on the KEGG data set curated by Bang et al. KGCNH achieved AUROC and AUPRC values of 0.93 and 0.945, respectively. Further details are provided in Supporting Information S4.

Test of KGCNH on Sparse Data. In this section, we investigate the model's performance under sparse data conditions. We randomly removed a fraction of data from Hetionet at a ratio of 5%, 10%, 15%, and 20% to simulate the scenario of incomplete data and evaluated the model by 10-fold cross-validation. As shown in Figure 4, the observed results show that more known triples are positively correlated with better prediction results, and further demonstrates that the number of triples is an important factor for the drug–disease association prediction. Additionally, our proposed model consistently achieves high performance, outperforming baseline models even when varying proportions of triplets are removed.

Ablation Study. To assess the impact of key components on the model's performance, we devised two model variants: KGCNH-w/o-GS, which removes the Gumbel-Softmax-based heuristic search module, and KGCNH-w/o-Aug, which removes the augmentation view module. We then employed 10-fold cross-validation to evaluate their performance. As shown in Table 4, we highlight the best results in bold. Our findings are as follows: (i) The complete version of KGCNH consistently outperforms other variants, underscoring the contribution of each component. (ii) The Gumbel-Softmax-based heuristic search module is essential, facilitating the model in finding superior embeddings and resulting in a notable performance enhancement.

Case Study. To evaluate the practical use of KGCNH, we conducted detailed case studies on the computationally predicted candidate diseases for the two drugs, namely, lithium and quetiapine. Specifically, we used all the known drug–disease associations in Hetionet as the training set and considered the missing drug–disease associations as candidate pairs. We subsequently ranked all the candidate diseases by the computed drug–disease association scores and verified the top 5 potential drug–disease associations for each drug in the CTD,[44] DrugCentral[45] databases, and PubMed.

Lithium. Lithium, an element in the alkali metals family, has the atomic symbol Li, atomic number 3, and atomic weight [6.938; 6.997]. Table 5 shows that KGCNH predicted five candidate diseases for lithium, all of which are confirmed by authoritative public databases (100% hit rate). For example, bipolar disorder is a top predicted candidate disease, and lithium has been proven as a first line treatment for it.[46] Figure 5 in DrugMechDB illustrates the path that represents the mechanism of action from lithium to bipolar disorder.

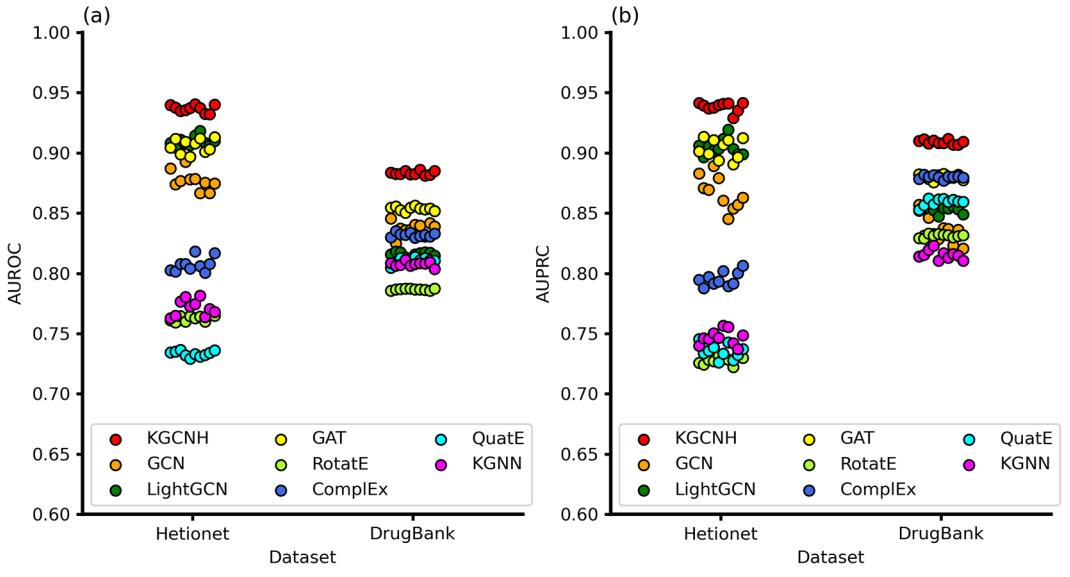


Figure 3. Performance comparison of 10 times 10-fold cross-validation between our method and baselines over Hetionet and DrugBank data sets. (a) Area under the receiver operating characteristic curves (AUROC) of prediction results. (b) Area under the precision-recall curves (AUPRC) of prediction results.

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928–4937

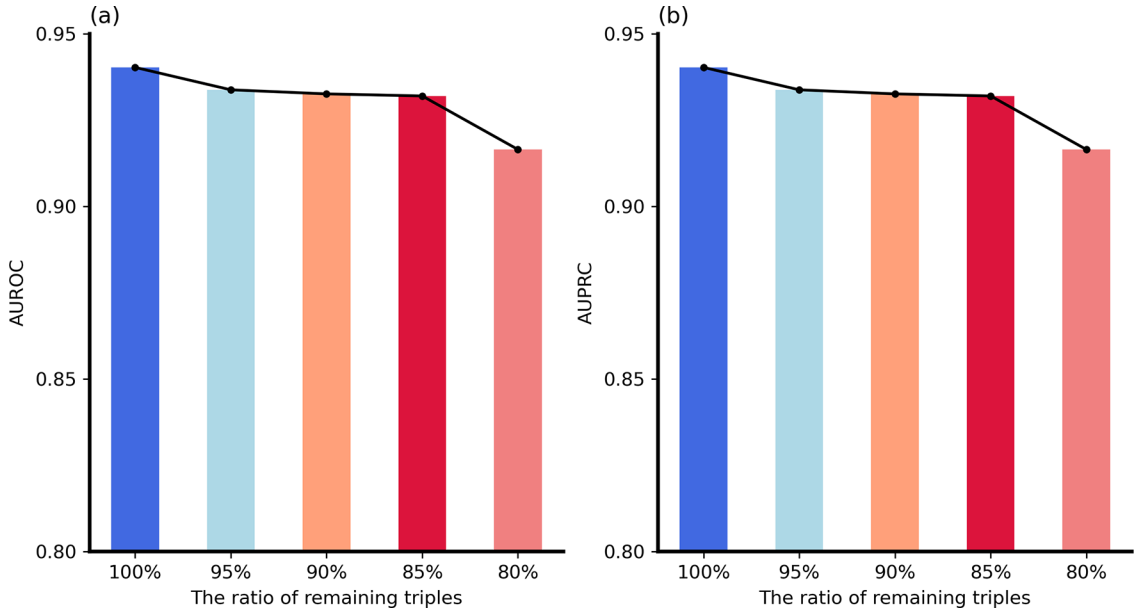


Figure 4. KGCNH performance for different sparsity ratios on Hetionet.

Table 4. Performance of KGCNH and Various Variants on Hetionet and DrugBank Data Sets

Model	Hetionet		DrugBank	
	AUROC	AUPRC	AUROC	AUPRC
KGCNH-w/o-GS	0.9226	0.9289	0.8699	0.9002
KGCNH-w/o-Aug	0.9339	0.9339	0.8816	0.9052
KGCNH	**0.9367**	**0.9383**	**0.8833**	**0.909**

Table 5. Top 5 Predicted Diseases Potentially Treatable by Lithium

rank	disease	MESH	evidence
1	Bipolar Disorder	D001714	46
2	Tourette Syndrome	D005879	47
3	Depressive Disorder	D003866	48
4	Autistic Disorder	D001321	49
5	Alzheimer's Disease	D000544	50

Tourette syndrome is another predicted candidate disease. Erickson et al.[47] found that when the Li^+ blood levels had stabilized at 0.8 to 0.9 mEq/L the major tics and involuntary sounds cleared dramatically and the patients experienced no side effects and were followed for several months without recurrence of the original symptoms. For Depressive Disorder also known as unipolar depression, Coppen[48] provided substantial evidence that lithium treatment decreases morbidity, suggesting that systematic, long-term lithium treatment of unipolar depression significantly reduced suicide rates. Autistic disorder is a developmental disability that affects how people interact with others and behave. Wang et al.[49] concluded that environment-related lithium exposure protected against neurobehavioral deficits in the rat valproic acid model of autism, implying that it may be a potential drug for the treatment of autism. Alzheimer's disease is a type of dementia that affects memory, thinking, and behavior. Toledo and Inestrosa[50] concluded that lithium and rosiglitazone, possibly by the activation of the Wnt signaling pathway, reduced various Alzheimer's disease neuropathological markers and may be considered as potential therapeutic agents against the disease.

Quetiapine. Similar to lithium, we also focused on analyzing the top 5 disease candidates for quetiapine predicted by KGCNH. Table 6 shows that these five potential diseases are verified by reliable evidence with a 100% hit rate. Depressive Disorder is a top predicted candidate disease, and Baune et al.[51] found an independent influence of quetiapine on improved depression, motor activity, and sleep. Matur and Üçok[52] described a case of a young man who had Tourette disorder and obsessive-compulsive disorder for 13 years and developed mania on clomipramine. The mania resolved and the symptoms of both disorders improved after he received 600 mg of quetiapine daily. Conduct disorder is a mental disorder characterized by a repetitive and persistent pattern of behavior that violates the basic rights of others or major age-appropriate societal norms or rules. Barzman et al.[53] found that quetiapine can be used to treat impulsiveness and aggression in adolescents with bipolar and disruptive behavior disorder. Parkinson's disease is a progressive, degenerative neurologic disease. Quetiapine is an atypical antipsychotic with sedative properties frequently used to treat hallucinations and psychosis in Parkinson's disease as reported by Juri et al.[54] Manic Disorder is a type of anxiety disorder characterized by unexpected panic attacks that last minutes or, rarely, hours. Takahashi et al.[55] reported three patients who suffered from panic attacks, and their symptoms improved significantly after quetiapine.

In summary, these case studies demonstrate the promising ability of KGCNH for discovering potential diseases of specific drugs. We expect that the candidate diseases predicted by KGCNH can provide a meaningful reference for clinicians in practical applications.

Discussion. Compared with GCN, LightGCN, and GAT, the KGCNH model variant without the Gumbel-Softmax-based heuristic search module achieved superior performance, mainly due to its relation-aware mechanism, which effectively handles diverse entities and relationships, enhancing the influence of relevant neighbors while mitigating interference from irrelevant ones. Compared with ComplEx, RotatE, and QuatE, KGCNH demonstrates superior performance, indicat-

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937

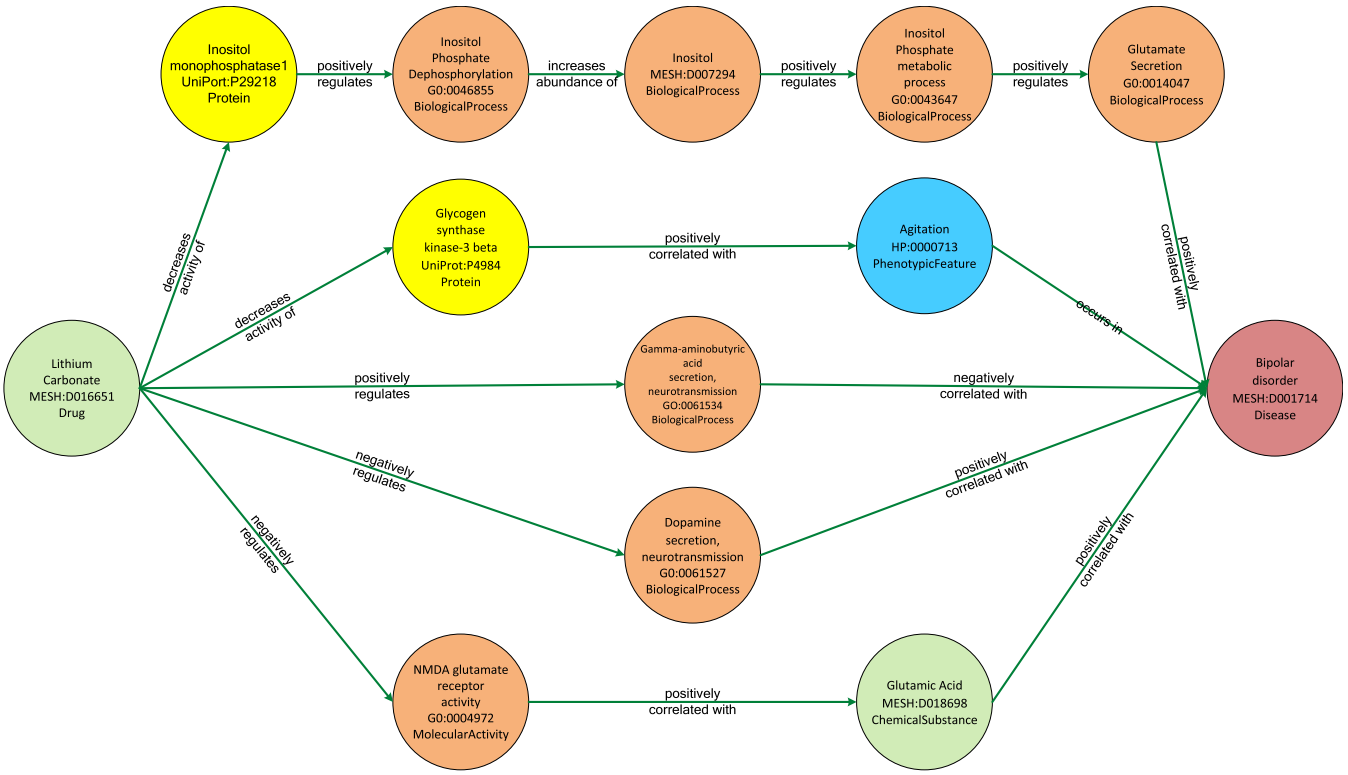


Figure 5. DrugMechDB-based mechanism of action path from lithium to a bipolar disorder.

Table 6. Top 5 Predicted Diseases Potentially Treatable by Quetiapine

rank	disease	MESH	evidence
1	Depressive Disorder	D003866	51
2	Tourette Syndrome	D005879	52
3	Conduct Disorder	D019955	53
4	Parkinson's Disease	D010300	54, 56
5	Panic Disorder	D016584	55

ing its ability to leverage neighborhood information more comprehensively. Furthermore, the ablation experiment results demonstrate that our designed Gumbel-Softmax-based heuristic search module can effectively introduce randomness, expanding the model's search space and finding better representations of drugs and diseases. However, this module still has limitations. First, it cannot guarantee finding the global optimum. Second, it relies on parameter settings, such as initial temperature and temperature drop rate.

CONCLUSION

In this article, we proposed a novel knowledge graph convolutional network with a heuristic search for drug repositioning, KGCNH. KGCNH incorporates relation features into the attention score computation, enabling it to effectively capture the importance between entities. To mitigate the influence of random initial embedding on model performance, we employed a heuristic search strategy based on Gumbel-Softmax, which can effectively expand the model's search space and enable it to learn better representations of drugs and diseases. We conducted extensive experiments to evaluate the effectiveness of KGCNH in drug repositioning tasks and compare it with several competing methods. The results demonstrate that KGCNH significantly outperforms the baselines. Furthermore, case studies suggest the ability of KGCNH to predict potential drug−disease associations for specific drugs.

Despite the superior performance of KGCNH, there is still room for improvement. First, the model relies on transductive learning, requiring retraining whenever new drugs or diseases are introduced to the knowledge graph. In future work, we plan to incorporate more contextual information about drugs and diseases and integrate inductive learning techniques. Second, our current model training involves the entire knowledge graph, necessitating substantial computational resources. With the default parameters that we have configured, employing an NVIDIA A800 GPU, the training duration is approximately 1 h for the Hetionet data set and 3.5 h for the DrugBank data set. In future work, one can develop a sampler capable of effectively filtering out irrelevant triples within the knowledge graph.

ASSOCIATED CONTENT

Data Availability Statement

The source code of KGCNH and the data underlying this article are available in our github repository at https://github.com/xiang-Du/KGCNH. The original data are available free of charge. Hetionet: https://github.com/hetio/hetionet; DrugBank (DRKG): https://github.com/gnn4dr/DRKG.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c00737.

Performance of KGCNH under the various ratios of positive to negative samples, statistics of data sets, the hyperparameter settings for the baseline models, and performance of KGCNH on the KEGG data set (PDF)

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937

AUTHOR INFORMATION

Corresponding Author

Min Li − *School of Computer Science and Engineering, Central South University, Changsha, Hunan 410083, China;* orcid.org/0000-0002-0188-1394; Email: limin@mail.csu.edu.cn

Authors

Xiang Du − *School of Computer Science and Engineering, Central South University, Changsha, Hunan 410083, China; School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China*

Xinliang Sun − *School of Computer Science and Engineering, Central South University, Changsha, Hunan 410083, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c00737

Funding

This work was supported by the National Natural Science Foundation of China [62225209] and the science and technology innovation program of Hunan Province [2021RC4008].

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We wish to thank the High Performance Computing Center of Central South University for supporting this work.

REFERENCES

(1) Ashburn, T. T.; Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discovery* **2004**, *3*, 673−683.

(2) Zhou, Y.; Wang, F.; Tang, J.; Nussinov, R.; Cheng, F. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digital Health* **2020**, *2*, e667−e676.

(3) Yang, Z.; Liu, J.; Zhu, X.; Yang, F.; Zhang, Q.; Shah, H. A. FragDPI: a novel drug-protein interaction prediction model based on fragment understanding and unified coding. *Front. Comput. Sci.* **2023**, *17*, 175903.

(4) Goodwin, S.; McPherson, J. D.; McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333−351.

(5) Pontén, F.; Jirström, K.; Uhlen, M. The Human Protein Atlas—a tool for pathology. *J. Pathol.* **2008**, *216*, 387−393.

(6) The GTEx Consortium; Ardlie, K. G.; Deluca, D. S.; Segrè, A. V.; Sullivan, T. J.; Young, T. R.; Gelfand, E. T.; Trowbridge, C. A.; Maller, J. B.; Tukiainen, T.; et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **2015**, *348*, 648−660.

(7) Wu, X.; Duan, J.; Pan, Y.; Li, M. Medical knowledge graph: Data sources, construction, reasoning, and applications. *Big Data Min. Anal.* **2023**, *6*, 201−217.

(8) MacLean, F. Knowledge graphs and their applications in drug discovery. *Expert Opin. Drug Discovery* **2021**, *16*, 1057−1069.

(9) Rossi, A.; Barbosa, D.; Firmani, D.; Matinata, A.; Merialdo, P. Knowledge graph embedding for link prediction: A comparative analysis. *ACM Trans. Knowledge Knowl. Discovery Data* **2021**, *15*, 1−49.

(10) Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Adv Neural Inf. Process. Syst.* **2013**, *26*, 2787.

(11) Wang, Z.; Zhang, J.; Feng, J.; Chen, Z. Knowledge graph embedding by translating on hyperplanes. *Proceedings of the AAAI conference on artificial intelligence*; AAAI Press: Washington, DC, 2014.

(12) Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; Zhu, X. Learning entity and relation embeddings for knowledge graph completion. *Proceedings of the AAAI conference on artificial intelligence*; AAAI Press: Washington, DC, 2015.

(13) Ji, G.; He, S.; Xu, L.; Liu, K.; Zhao, J. Knowledge graph embedding via dynamic mapping matrix. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing*; Association for Computer Linguistics, 2015; Vol *1* (Long papers), pp 687−696.

(14) Sun, Z.; Deng, Z.; Nie, J.; Tang, J. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. *7th International Conference on Learning Representations, ICLR 2019*; 2019.

(15) Choudhary, S.; Luthra, T.; Mittal, A.; Singh, R. A survey of knowledge graph embedding and their applications. *arXiv* **2021**, No. arXiv:2107.07842.

(16) Nickel, M.; Tresp, V.; Kriegel, H.-P. A three-way model for collective learning on multi-relational data. *ICML'11 Proceedings of the 28th International Conference on Machine Learning* **2011**, 809−816.

(17) Yang, B.; Yih, W.; He, X.; Gao, J.; Deng, L. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7−9, 2015, Conference Track Proceedings*; 2015.

(18) Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; Bouchard, G. Complex embeddings for simple link prediction. *International conference on machine learning*; Association for Computing Machinery, 2016; pp 2071−2080.

(19) Zhang, S.; Tay, Y.; Yao, L.; Liu, Q. Quaternion knowledge graph embeddings. *Adv Neural Inf. Process. Syst.* **2019**, *32*, 2735.

(20) Ge, X.; Wang, Y.-C.; Wang, B.; Kuo, C.-C. J. Knowledge Graph Embedding: An Overview. *arXiv* **2023**, No. arXiv:2309.12501.

(21) Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Networks* **2009**, *20*, 61−80.

(22) Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral Networks and Locally Connected Networks on Graphs. *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14−16, 2014, Conference Track Proceedings*; 2014.

(23) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24−26, 2017, Conference Track Proceedings*; 2017.

(24) He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; Wang, M. Lightgcn: Simplifying and powering graph convolution network for recommendation. *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*; Association for Computing Machinery, 2020; pp 639−648.

(25) Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*; 2018.

(26) Wu, Y.; Gao, M.; Zeng, M.; Zhang, J.; Li, M. BridgeDPI: a novel graph neural network for predicting drug−protein interactions. *Bioinformatics* **2022**, *38*, 2571−2578.

(27) Wang, K.; Li, M. Fusion-Based Deep Learning Architecture for Detecting Drug-Target Binding Affinity Using Target and Drug Sequence and Structure. *IEEE J. Biomed. Health Inf* **2023**, *27*, 6112.

(28) Wang, K.; Zhou, R.; Tang, J.; Li, M. GraphscoreDTA: optimized graph neural network for protein−ligand binding affinity prediction. *Bioinformatics* **2023**, *39*, btad340.

(29) Guo, Y.; Zhou, D.; Ruan, X.; Cao, J. Variational gated autoencoder-based feature extraction model for inferring disease-miRNA associations based on multiview features. *Neural Networks* **2023**, *165*, 491−505.

(30) Sun, X.; Wang, B.; Zhang, J.; Li, M. Partner-Specific Drug Repositioning Approach Based on Graph Convolutional Network. *IEEE J. Biomed. Health Inf.* **2022**, *26*, 5757−5765.

(31) Jia, X.; Sun, X.; Wang, K.; Li, M. DRGCL: Drug Repositioning via Semantic-enriched Graph Contrastive Learning. *IEEE J. Biomed. Health Inf.* **2024**, 3372527.
(32) Sun, X.; Jia, X.; Lu, Z.; Tang, J.; Li, M. Drug repositioning with adaptive graph convolutional networks. *Bioinformatics* **2024**, *40*, btad748.
(33) Wang, H.; Zhao, M.; Xie, X.; Li, W.; Guo, M. Knowledge graph convolutional networks for recommender systems. *The world wide web conference*; Association for Computing Machinery, 2019; pp 3307−3313.
(34) Wang, X.; He, X.; Cao, Y.; Liu, M.; Chua, T.-S. Kgat: Knowledge graph attention network for recommendation. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*; Association for Computing Machinery, 2019; pp 950−958.
(35) Wang, H.; Xu, Y.; Yang, C.; Shi, C.; Li, X.; Guo, N.; Liu, Z. Knowledge-adaptive contrastive learning for recommendation. *Proceedings of the sixteenth ACM international conference on web search and data mining*; Association for Computing Machinery, 2023; pp 535−543.
(36) Lin, X.; Quan, Z.; Wang, Z.-J.; Ma, T.; Zeng, X. KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence* **2020**, 2739−2745.
(37) Bang, D.; Lim, S.; Lee, S.; Kim, S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. *Nat. Commun.* **2023**, *14*, 3570.
(38) Zhang, X.; Che, C. Drug repurposing for parkinson's disease by integrating knowledge graph completion model and knowledge fusion of medical literature. *Future Internet* **2021**, *13*, 14.
(39) Zhang, R.; Hristovski, D.; Schutte, D.; Kastrin, A.; Fiszman, M.; Kilicoglu, H. Drug repurposing for COVID-19 via knowledge graph completion. *J. Biomed. Inf.* **2021**, *115*, 103696.
(40) Himmelstein, D. S.; Lizee, A.; Hessler, C.; Brueggeman, L.; Chen, S. L.; Hadley, D.; Green, A.; Khankhanian, P.; Baranzini, S. E. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife* **2017**, *6*, No. e26726.
(41) Ioannidis, V. N.; Song, X.; Manchanda, S.; Li, M.; Pan, X.; Zheng, D.; Ning, X.; Zeng, X.; Karypis, G. *DRKG - Drug Repurposing Knowledge Graph for Covid-19*. https://github.com/gnn4dr/DRKG/, 2020.
(42) Jang, E.; Gu, S.; Poole, B. Categorical Reparameterization with Gumbel-Softmax. *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24−26, 2017, Conference Track Proceedings*; 2017.
(43) Rendle, S.; Freudenthaler, C.; Gantner, Z.; Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. *arXiv* **2012**, No. arXiv:1205.2618.
(44) Davis, A. P.; Grondin, C. J.; Johnson, R. J.; Sciaky, D.; Wiegers, J.; Wiegers, T. C.; Mattingly, C. J. Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res.* **2021**, *49*, D1138−D1143.
(45) Ursu, O.; Holmes, J.; Knockel, J.; Bologa, C. G.; Yang, J. J.; Mathias, S. L.; Nelson, S. J.; Oprea, T. I. DrugCentral: online drug compendium. *Nucleic Acids Res.* **2017**, D932.
(46) Simhandl, C.; Mersch, J. Lithium and bipolar disorder−a renaissance? *Neuropsychiatrie: Klinik, Diagnostik, Therapie und Rehabilitation: Organ der Gesellschaft Osterreichischer Nervenarzte und Psychiater* **2007**, *21*, 121−130.
(47) Erickson, H.; Goggin, J.; Messiha, F. Comparison of lithium and haloperidol therapy in Gilles de la Tourette syndrome. *Adv. Exp. Med. Biol.* **1977**, *90*, 197−205.
(48) Coppen, A. Lithium in unipolar depression and the prevention of suicide. *J. Clin. Psychiatry* **2000**, *61*, 52−56.
(49) Wang, J.; Xu, C.; Liu, C.; Zhou, Q.; Chao, G.; Jin, Y. Effects of different doses of lithium on the central nervous system in the rat valproic acid model of autism. *Chem.-Biol. Interact.* **2023**, *370*, 110314.
(50) Toledo, E.; Inestrosa, N. Activation of Wnt signaling by lithium and rosiglitazone reduced spatial memory impairment and neurodegeneration in brains of an APPswe/PSEN1ΔE9 mouse model of Alzheimer's disease. *Mol. Psychiatry* **2010**, *15*, 272−285.
(51) Baune, B. T.; Caliskan, S.; Todder, D. Effects of adjunctive antidepressant therapy with quetiapine on clinical outcome, quality of sleep and daytime motor activity in patients with treatment-resistant depression. *Hum. Psychopharmacol.* **2007**, *22*, 1−9.
(52) Matur, Z.; Üçok, A. Quetiapine treatment in a patient with Tourette's syndrome, obsessive-compulsive disorder and drug-induced mania. *Isr. J. Psychiatry Relat. Sci.* **2003**, *40*, 150.
(53) Barzman, D. H.; DelBello, M. P.; Adler, C. M.; Stanford, K. E.; Strakowski, S. M. The efficacy and tolerability of quetiapine versus divalproex for the treatment of impulsivity and reactive aggression in adolescents with co-occurring bipolar disorder and disruptive behavior disorder (s). *J. Child Adolescent Psychopharmacol.* **2006**, *16*, 665−670.
(54) Juri, C.; Chaná, P.; Tapia, J.; Kunstmann, C.; Parrao, T. Quetiapine for insomnia in Parkinson disease: results from an open-label trial. *Clin. Neuropharmacol.* **2005**, *28*, 185−187.
(55) Takahashi, H.; Sugita, T.; Yoshida, K.; Higuchi, H.; Shimizu, T. Effect of quetiapine in the treatment of panic attacks in patients with schizophrenia: 3 case reports. *J. Neuropsychiatry Clin. Neurosci.* **2004**, *16*, 113−115.
(56) Cannas, A.; Solla, P.; Floris, G.; Tacconi, P.; Loi, D.; Marcia, E.; Marrosu, M. G. Hypersexual behaviour, frotteurism and delusional jealousy in a young parkinsonian patient during dopaminergic therapy with pergolide: A rare case of iatrogenic paraphilia. *Prog. Neuropsychopharmacol Biol. Psychiatry* **2006**, *30*, 1539−1541.

https://doi.org/10.1021/acs.jcim.4c00737
J. Chem. Inf. Model. 2024, 64, 4928−4937