Bioinformatics
Advance Access Publication Date: Day Month Year
Manuscript Category

Data and text mining

Multi-scale topology and position feature learning and relationship-aware graph reasoning for prediction of drug-related microbes

Ping Xuan [1, 2], Jing Gu [1], Hui Cui [3], Shuai Wang [4], Toshiya Nakaguchi [5], Cheng Liu [2], Tiangang Zhang [1, 6,*]

[1]School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China, [2]Department of Computer Science, Shantou University, Shantou 515063, China, [3]Department of Computer Science and Information Technology, La Trobe University, Melbourne 3083, Australia, [4]School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China, [5]Center for Frontier Medical Engineering, Chiba University, Chiba 2638522, Japan, [6]School of Mathematical Science, Heilongjiang University, Harbin 150080, China

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The human microbiome may impact the effectiveness of drugs by modulating their activities and toxicities. Predicting candidate microbes for drugs can facilitate the exploration of the therapeutic effects of drugs. Most recent methods concentrate on constructing of the prediction models based on graph reasoning. They fail to sufficiently exploit the topology and position information, the heterogeneity of multiple types of nodes and connections, and the long-distance correlations among nodes in microbe-drug heterogeneous graph.
Results: We propose a new microbe-drug association prediction model, NGMDA, to encode the position and topological features of microbe (drug) nodes, and fuse the different types of features from neighbors and the whole heterogeneous graph. First, we formulate the position and topology features of microbe (drug) nodes by t-step random walks, and the features reveal the topological neighborhoods at multiple scales and the position of each node. Second, as the features of nodes are high-dimensional and sparse, we designed an embedding enhancement strategy (ES) based on supervised fully-connected autoencoders to form the embeddings with representative features and the more discriminative node distributions. Third, we propose an adaptive neighbor feature fusion (NFF) module which fuses features of neighbors by the constructed position- and topology-sensitive heterogeneous graph neural networks (HGNN). A novel self-attention mechanism is developed to estimate the importance of the position and topology of each neighbor to a target node. Finally, a heterogeneous graph feature fusion (GFF) module is constructed to learn the long-distance correlations among the nodes in the whole heterogeneous graph by a relationship-aware graph transformer (RAGT). RAGT contains the strategy for encoding the connection relationship types among the nodes, which is helpful for integrating the diverse semantics of these connections. The extensive comparison experimental results demonstrates NGMDA's superior performance over five state-of-the-art prediction methods. The ablation experiment shows the contributions of the multi-scale topology and position feature learning, the embedding enhancement strategy, the neighbor feature fusion, and the heterogeneous graph feature fusion. Case studies over three drugs further indicate that NGMDA has ability in discovering the potential drug-related microbes.
Availability: Source codes and supplementary materials are available at https://github.com/pingxuan-hlju/NGMDA.
Contact: zhang@hlju.edu.cn
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The human microbiome is a collection of all microbiota that reside in or on human organs, including bacteria, viruses, protists, fungi, and archaea. Previous human microbiome studies demonstrated that interactions between the human microbes and corresponding hosts regulate human health, such as controlling immune function, providing resistance to pathogens, and even influencing brain physiology and behavior (Duvallet *et al.*, 2017; Zhu *et al.*, 2020). An imbalance of human microbiota and some diseases are closely related, including chronic inflammation, neurological disorders, and breast cancer (Wang *et al.*, 2019; Rackaityte and Lynch, 2020).

Microbes can change the toxicity and inhibitory activity of drugs (Nejman *et al.*, 2020; Algavi and Borenstein, 2023) and impact the effectiveness of disease treatments by biologically altering a drug's chemical structure (Yin *et al.*, 2022). Hucioglu *et al*. suggested that cooperation between Staphylococcus aureus and Candida albicans leads to drug resistance by strengthening biofilm formation (Hacioglu *et al.*, 2019). Also, the gut microbiome produces large quantities of bacterial enzymes that affect therapeutic efficacy (Zimmermann *et al.*, 2019). Therefore, discovering new microbe-drug associations is essential in drug functional studies and precision medicine.

Recently, the computational methods were proposed for predicting the drug-target interactions (Li *et al.*, 2022), incRNA-miRNA interactions (Wang *et al.*, 2022), miRNA-disease associations (Peng *et al.*, 2022a), metabolite-disease associations (Gao *et al.*, 2023), and incRNA-disease associations (Wang *et al.*, 2023a). Computational methods have also shown the ability to determine potential microbe-drug associations and identify reliable drug-related candidates for wet experiments. Microbe-drug association probabilities can be inferred by prediction models using Conditional Random Field (CRF) and Graph Convolutional Network (GCN) (Long *et al.*, 2020a). Long *et al*. proposed EGTMDA to learn node features for microbes and drugs using meta-paths and hierarchical attention mechanism (Long *et al.*, 2020b). SCSMDA enhanced representations of drugs and microbes using graph contrastive learning and elaborate meta-paths (Tian *et al.*, 2023). However, shortcomings exist in these methods. GCNMDA used vanilla homogeneous models to learn representations of drugs and microbes without considering abundantly available heterogeneous information. In addition, these methods based on meta-paths focus on neighbors originating from meta-paths while ignoring other non-neighboring nodes across the entire heterogeneous graph.

We proposed NGMDA to predict candidate microbes for drugs by learning the features of drugs and microbes from neighbors and the whole heterogeneous graph. Our contributions are summarized as follows:

- The multi-scale topology information of nodes reflects neighbor regions of different ranges, which is important for microbe-drug association prediction. Therefore, topology features of microbe (drug) nodes are designed based on t-step random walks to obtain multi-scale topological neighborhoods of nodes. We also extracted node position features to form the position of each node in the entire heterogeneous graph.
- An embedding enhancement strategy based on fully-connected autoencoders with node class labels is proposed to extract important low-dimensional features of the microbe or drug nodes. This strategy also enhances the differences of feature distributions among different types of nodes by determining the node class.
- In the microbe-drug heterogeneous graph, different neighbor nodes often have special topological neighborhoods and positional features that affect the importance of neighbors with a target node. A new position-sensitive and topology-sensitive self-attention mechanism (PTA) adaptively distinguishes the contributions of different neighbor nodes. Also, NFF models heterogeneity of the graph and aggregates the representations of nodes based on HGNN with PTA.
- A microbe or drug node may be closely related to distant nodes due to the heterogeneity of the microbe-drug graph. We have designed GFF based on a relationship-aware graph transformer (RAGT) to reveal the diverse connections between the target node and all other nodes in the heterogeneous graph. Comprehensive experiments suggest the superiority of NGMDA by comparing it with advanced methods.

2 Materials and methods

We propose a microbe-drug association prediction model called NGMDA (Figure 1) that consists of an embedding enhancement strategy (ES), neighbor feature fusion module (NFF), and heterogeneous graph feature fusion module (GFF). A heterogeneous graph is constructed to describe the diverse connectivity relationships between drugs and microbes (Figure 1 (a)). The node features of these drugs and microbes are projected into a low-dimensional feature space, and their differences are enhanced to obtain a fine node embedding (Figure 1 (a)). NFF learns similarity, position, and topology representations between nodes based on position-sensitive and topology-sensitive HGNN (Figure 1(b)). We use GFF to learn multi-modal representations of various nodes across the heterogeneous graph by a relationship-aware graph transformer (Figure 1(c)). These four representations are combined into fully-connected layers to predict microbe-drug association probabilities.

2.1 Dataset

Associations between drugs and microbes, similarities between drugs, and the attribute features of the microbes X^{micr} are collected from previously published microbe-drug association prediction work (Long *et al.*, 2020b). We extracted 2470 microbe-drug association data from the Microbe-Drug Associations Database (MDAD) (Sun *et al.*, 2018), which contains 173 microbes and 1373 drugs. Drugbank (Knox *et al.*, 2024) provides the interactions among the drugs. On the basis of the biological hypothesis that the drugs with similar treatment functions are more likely interact with the similar microbes, EGTMDA calculated the Gaussian kernel similarities of drugs based on their interactions. The structural similarity of two drugs was measured based on the common subgraphs within their chemical structures (Hattori *et al.*, 2010). The final drug similarities were obtained by the weighted sum of the drug Gaussian kernel similarities and the drug structure similarities. The sequences of microbes were extracted from NCBI database, and then principal component analysis (PCA) was utilized to obtain their important features.

2.2 Calculation of microbe similarity

As two microbes with similar gene sequences are typically similar, we calculate the cosine similarity on the attribute characteristics for each microbe. The similarity between microbe m_i and m_j is $K_{ij}^{micr} \in [0, 1]$,

$$K_{ij}^{micr} = \frac{1}{2}\left(\frac{X_i^{micr}(X_j^{micr})^T}{\|X_i^{micr}\|\|X_j^{micr}\|} + 1\right), \tag{1}$$

where X_i^{micr} is the i-th row of X^{micr}, which contains the main gene sequence characteristics of m_i, and $(X)^T$ is a transposition of X. The microbe similarities were listed in the supplementary file SF1.

2.3 Microbe-drug heterogeneous graph

We constructed a microbe-drug heterogeneous graph $G = (V, E)$ as shown in Figure 1 (a). The node set V consists the drug node subset V^{drug} and microbe node subset V^{micr} and $< i, j > \in E$ represents an edge from node v_j to v_i. The drug similarity matrix and drug-microbe association matrix are expressed as K^{drug} and $B^{bipa} \in R^{N_d \times N_m}$, respectively, where N_d (or N_m) denotes the number of drugs (or microbes). If there is a known association exists between drugs d_i and m_j, then $B_{i,j}^{bipa} = 1$. Further, $B_{i,j}^{bipa} = 0$ indicates that no connection has yet been found. There are many low similarity data in the similarity matrix, which might be noise in microbe-drug association prediction. When constructing microbe-microbe (or drug-drug) adjacent matrix, connecting edges are added between the microbe (or drug) nodes with a similarity not less than a threshold β. The adjacency matrix of the heterogeneous graph G is represented as $B^{hete} \in R^{(N_d+N_m)\times(N_d+N_m)}$, such that

$$B^{hete} = \begin{bmatrix} \tilde{K}^{drug} & B^{bipa} \\ (B^{bipa})^T & \tilde{K}^{micr} \end{bmatrix}, \tag{2}$$

where $\tilde{K}^{drug}$ (or $\tilde{K}^{micr}$) is the drug (or microbe) similarity matrix after thresholding.

2.4 Heterogeneous graph node feature construction and enhancement

2.4.1 Heterogeneous graph node feature construction

The heterogeneous graph node features are constructed by a drug-drug similarity matrix, microbe-microbe similarity matrix, and drug-microbe association matrix. The similarity feature matrix is formed by combining the drug and microbe similarities defined above as

$$H_{simi} = \begin{bmatrix} K^{drug} \\ K^{micr} \end{bmatrix}, \tag{3}$$

where K_i^{drug} (or K_j^{micr}) contains the similarities between d_i (or m_j) and other drugs (or microbes). The multi-modal feature matrix $H_{moda} \in R^{(N_d+N_m)\times(N_d+N_m)}$ can be represented as

$$H_{moda} = \begin{bmatrix} K^{drug} & B^{bipa} \\ (B^{bipa})^T & K^{micr} \end{bmatrix}, \tag{4}$$

where the i-th row in H_{moda} records the similarities between d_i and all other drugs and the associations between d_i and all other microbes. The association

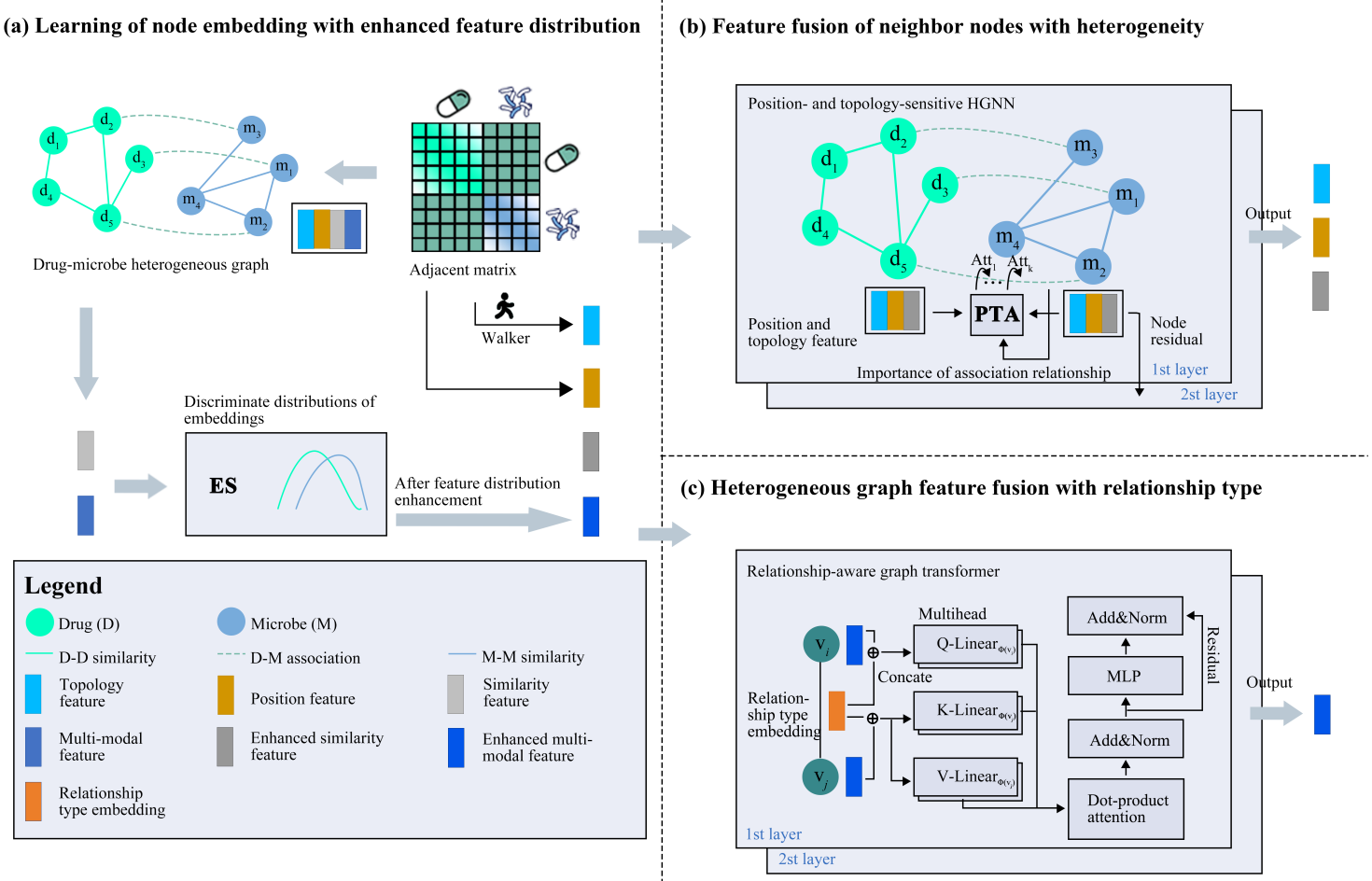

Fig. 1: Framework of the proposed NGMDA model. **(a)** construct microbe-drug heterogeneous graph and enhance similarity and multi-modal embeddings. **(b)** fuse features of neighbor nodes by position-sensitive and topology-sensitive HGNN. **(c)** learn long distance connection from the entire heterogeneous graph based on RAGT.

with drugs and similarities between microbes are contained in the $(N_d + j)$-th row. Because similarity and multi-modal features are common node attributes for microbe-drug association prediction (Peng *et al.*, 2017; Peng *et al.*, 2021; Wang *et al.*, 2023; Meng *et al.*, 2023), we designate these as the original features of the nodes. Existing GNN models fail to fully consider the position and topology information of nodes, so we construct position and topology features of the microbe and drug nodes. The position of v_i within the heterogeneous graph is determined by the connection between v_i and other nodes. The position feature matrix is defined as $H_{posi} = B^{hete}$, where the position feature of v_i is $H_{posi,i}$. A random walk of t-steps contains a t-hop topological neighborhood of nodes within a heterogeneous graph (Dwivedi *et al.*, 2022) and is defined as

$$RW^t = (B^{hete}(D^{hete})^{-1})^t, \tag{5}$$

where t is the number of walking steps and D^{hete} is degree matrix of B^{hete}. $RW^t_{i,j}$ represents the probability of visiting v_i to v_j in the t-th step random walk and contains the topological neighborhood information of the t-th step of v_i. The topology feature $H_{topo,i} \in R^t$ of v_i is defined as

$$H_{topo,i} = [RW_{i,i}, RW^2_{i,i}, \cdots, RW^t_{i,i}], \tag{6}$$

which contains the multi-scale topological neighborhood information of v_i.

2.4.2 Enhancing node embedding

The original features specified above are high-dimensional sparse and contain some noise. A projection operation maps drug and microbe node features into the same embedding space, which drops information about the differences in the embedding distributions of different types of nodes. Figure 2 outlines our node embedding enhancement strategy to learn representative embeddings and enhance the embedding distribution differences of the microbe and drug

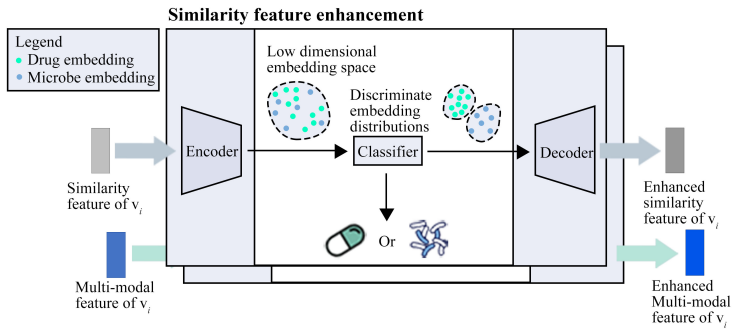

Fig. 2: Enhancing node embeddings of microbes and drugs by supervised autoencoders.

nodes. As autoencoders could effectively reduce the noise component in these embeddings, we learn important low-dimensional node embeddings based on fully-connected autoencoders. The projection and reconstruction process of multi-modal and similarity features are similar, and we use similarity features as an example to describe the process here. The similarity feature of v_i, $H_{simi,i}$, is projected into N_p dimensional space to form

$$H^{enco,1}_{simi,i} = \sigma(Linear\_enco^{(1)}_{simi,\phi(v_i)}(H_{simi,i})), \tag{7}$$

where $Linear\_enco$ is a linear layer, σ represents the nonlinear activation function ReLU, and $\phi(v_i)$ indicates the type of v_i. The similarity embedding of v_i is learned from the l-th fully-connected encoding layer as

$$H^{enco,l}_{simi,i} = \sigma(Linear\_enco^{(l)}_{simi,\phi(v_i)}(H^{enco,l-1}_{simi,i})), \quad l = 1, 2, \cdots, L_{enco}, \tag{8}$$

where L_{enco} is the total number of encoding layers. $H_{simi,i}^{enco,L_{enco}}$ is used as the input of the decoder, and the output of the l-th fully-connected decoding layer is

$$H_{simi,i}^{deco,l} = \sigma(Linear\_deco_{simi,\phi(v_i)}^{(l)}(H_{simi,i}^{deco,l-1})), \quad l = 1, 2, \cdots, L_{deco}, \tag{9}$$

where L_{deco} is total number of decoding layers, $Linear\_deco$ denotes the linear layer and $H_{simi,i}^{deco,0} = H_{simi,i}^{enco,L_{enco}}$. After projection, the multi-modal embedding $H_{moda}^{enco,L_{enco}} \in R^{(N_d+N_m)\times(N_p)}$ can be learned. The mean square error (MSE) estimates the reconstruction loss of the node similarity features as

$$\gamma_{reco,simi} = \frac{1}{|T|}\sum_{i\in T}\left\|H_{simi,i} - H_{simi,i}^{L_{deco}}\right\|^2, \tag{10}$$

where T is the batch of nodes in the training set. Similarly, the reconstruction loss of the multi-modal feature is $\gamma_{reco,moda}$.

We classify the projected node embedding to enhance the differences between the drug and microbe embedding distributions. Considering a multi-modal embedding, as an example, is the input of the classify and $label_i$ is the corresponding classification labels. The classification loss of the multi-modal embedding in the training samples is estimated by the cross-entropy loss function

$$\gamma_{clas,moda} = \frac{1}{|T|}\sum_{i\in T}\sum_{k=1}^{2} -label_{i,k}\cdot log(Linear\_clas^{moda}(H_{moda,i}^{enco,L_{enco}})_k), \tag{11}$$

where $Linear\_clas^{moda} \in R^{N_p\times 2}$. The classification loss of the similarity embedding is represented as $\gamma_{inty,simi}$. The total loss of the embedding classification of the drug and microbe nodes is

$$\gamma_{proc} = \gamma_{reco,simi} + \gamma_{reco,moda} + \gamma_{inty,simi} + \gamma_{inty,moda}. \tag{12}$$

2.5 Neighbor feature fusion

The topological neighborhood and position information of the neighboring nodes impact their importance with a target node. We propose a neighbor feature fusion (NFF) module based on heterogeneous graph neural networks (HGNN) with a position-sensitive and topology-sensitive self-attention mechanism (PTA) to learn representative similarity and the position and topology representations of each microbe and drug nodes. The relationship types between the nodes are critical auxiliary features, so we also calculate the importance of the integrated features, as shown in Figure 3.

The type of relationship between nodes contains the similarity relationship between drugs (or microbes) and the association relationship between drugs (or microbes) and microbes (or drugs). The relationship type of v_j to v_i is represented as

$$\psi(<i,j>) = \begin{cases} 0, & if\ v_i, v_j \in V^{drug} \\ 1, & if\ v_i \in V^{micr}, v_j \in V^{drug} \\ 2, & if\ v_i \in V^{drug}, v_j \in V^{micr} \\ 3, & if\ v_i, v_j \in V^{micr} \end{cases}. \tag{13}$$

Then, the importance of the relationship type of v_j to v_i is $r_{\psi(<i,j>)}^{l}$, which is learned during the training process at the l-th layer, and $r_{\psi(<i,j>)}^{0} = 1$.

Multi-head attention can reasonably stabilize the learning process of self-attention by allocating the attention value of each head (Veličković *et al.*, 2018). After obtaining the similarity representations $H_{simi,j}^{l-1}$ and $H_{simi,i}^{l-1}$ of v_i and v_j, respectively, at $l-1$-th layer, we compute the importance of the similarity representation of v_j to v_i in the next layer by

$$s_{i,j,k}^{l} = (softmax(W_{p(k)}^{l}W_{\phi(v_i),k}^{l}H_{simi,i}^{l-1}))^T norm(W_{\phi(v_j),k}^{l}H_{simi,j}^{l-1}), \quad l = 1, 2, \cdots, L_{nff}\ and\ k = 1, 2, \cdots, K_{nff}, \tag{14}$$

where L_{nff} is the total layer number, K_{nff} is the head number, $H_{simi,i}^{0} = H_{simi,i}^{enco,L_{enco}}$, $W_{\phi(v_i),k}^{l}$ and $W_{p(k)}^{l} \in R^{N_p\times N_p}$ are weight matrices, and $norm$ represents L_2 normalization. The $H_{simi,i}^{l-1}$ and $H_{simi,j}^{l-1}$ terms are transformed into latent representations $W_{\phi(v_i),k}^{l}H_{simi,i}^{l-1}$ and $W_{\phi(v_j),k}^{l}H_{simi,j}^{l-1}$. Then, $W_{p(k)}^{l}W_{\phi(v_i),k}^{l}H_{simi,i}^{l-1}$ is the distribution of the similarity representation

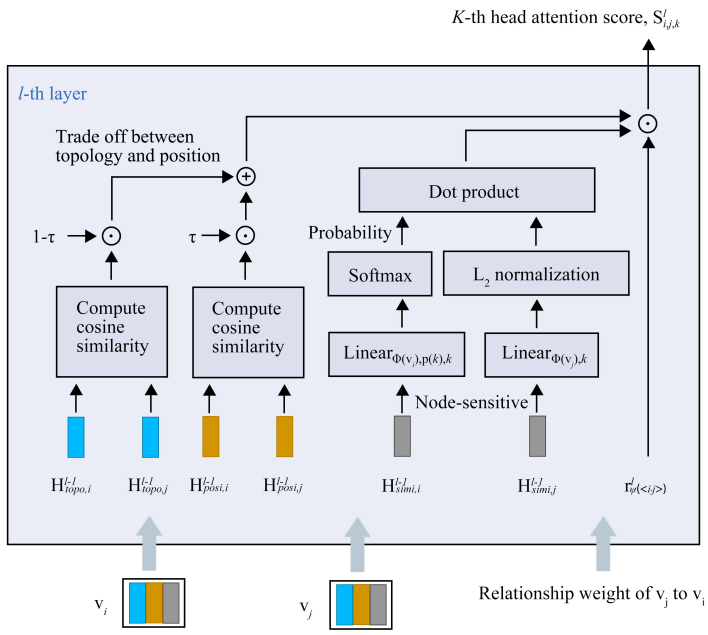

Fig. 3: Calculation of position-sensitive and topology-sensitive self-attention of v_j to v_i.

importance, which is converted to a probability distribution using the softmax function. As the magnitude of the latent representation affects the importance score of v_j to v_i, we standardize $W_{\phi(v_j),k}^{l}H_{simi,j}^{l-1}$ with L_2 normalization. The importance score of v_j to v_i is calculated by the inner product of the importance distribution of v_i and the representation of v_j. The multiple neighbors of node v_i have their various topological neighborhoods and positions, so these neighbors have different importance for v_i's feature learning. Therefore, the importance of each neighbor node for v_i was calculated before the v_i's features were updated. We calculate the importance $c_{i,j}^{l} \in [0,1]$ of the position and topology of v_j to v_i by

$$c_{i,j}^{l} = [\tau\cdot cosine(H_{posi,i}^{l-1}, H_{posi,j}^{l-1}) + (1-\tau)\cdot cosine(H_{topo,i}^{l-1}, H_{topo,j}^{l-1}) + 1]/2, \tag{15}$$

where $H_{posi,i}^{l-1}$ and $H_{topo,i}^{l-1}$ are the position and topology representations of v_i, respectively, $H_{posi,i}^{0} = H_{posi,i}$, and $H_{topo,i}^{0} = H_{topo,i}$. The parameter $\tau \in [0,1]$ balances the contributions between the position and topology representations. The importance score of v_j to v_i is then

$$a_{i,j,k}^{l} = softmax_{j\in N(i)}(r_{\psi(<i,j>)}^{l}\cdot s_{i,j,k}^{l}\cdot c_{i,j}^{l}). \tag{16}$$

Here, $a_{i,j,k}^{l}$ is position and topology sensitive by integrating the importance of the neighbor position and topology.

As residuals can alleviate over-smoothing and vanishing gradients (Lv *et al.*, 2021), we add a node residual for every attention head. The similarity representation of v_i is then updated as

$$H_{simi,i,k}^{l} = \sigma(\sum_{j\in N(i)} a_{i,j,k}^{l}\cdot W_{\phi(v_i),k}^{l}\cdot H_{simi,j}^{l-1} + W_{res(k)}^{l}\cdot H_{simi,i}^{l-1}), \tag{17}$$

where $W_{res(k)}^{l} \in R^{N_p\times N_p}$ is the weight matrix. The similarity representations of the different heads are aggregated at the l-th layer to obtain

$$H_{simi,i}^{l} = \frac{1}{K_{nff}}\sum_{k=1}^{K_{nff}} H_{simi,i,k}^{l}, \tag{18}$$

where K_{nff} is the head number of the NFF. As the similarity, position, and topology representations have the same update procedure, the position and topology representations are updated as $H_{posi,i}^{l}$ and $H_{topo,i}^{l}$ in the l-th layer, respectively.

2.6 Heterogeneous graph feature fusion

2.6.1 Relationship type encoding

Relationship types of similarity and association can reflect diverse semantic connections between drug and microbe nodes. The relationship type $\psi(< i, j >)$ of v_j to v_i is represented as a one-hot vector that is linearly transformed to obtain the embedding of the relationship type $e_{\psi(<i,j>)} \in R^{N_p}$.

2.6.2 Relationship-aware graph transformer

To capture the connection between the target node and distant nodes, we designed a heterogeneous graph feature fusion (GFF) module presented in Figure 1 (c). A relationship-aware graph transformer (RAGT) is proposed within the GFF and inspired by these methods (Peng *et al.*, 2022b; Peng *et al.*, 2022c; Diao and Loynd, 2022). To embed the relationship type $\psi(< i, j >)$ of v_j to v_i into the query, key, and value vectors, we concatenate a multi-modal feature of v_i (or v_j) and the relationship type embedding $e_{\psi(<i,j>)}$. We obtain the query, key, and value vectors of the h-th head in the l-th layer by linear transformations such that

$$\begin{aligned} q^l_{i,j,h} &= [H^{enco,l-1}_{moda,i}, e_{\psi(<i,j>)}]W^l_{\phi(v_i),Q,h}, \\ k^l_{i,j,h} &= [H^{enco,l-1}_{moda,j}, e_{\psi(<i,j>)}]W^l_{\phi(v_j),K,h}, \\ v^l_{i,j,h} &= [H^{enco,l-1}_{moda,j}, e_{\psi(<i,j>)}]W^l_{\phi(v_j),V,h}, \\ l = 1, 2, \cdots, L_{gff} &and\ h = 1, 2, \cdots, K_{gff}, \end{aligned} \tag{19}$$

where $W^l_{\phi(v_i),Q,h}, W^l_{\phi(v_i),K,h}, W^l_{\phi(v_i),V,h} \in R^{2\times N_p, \frac{N_p}{K_{gff}}}$ are weight matrices, $H^{G,0}_{enc,i} = H^G_{enc,i}$, L_{gff} is the total layer number, and K_{gff} indicates the head number of GFF. The importance of v_j to v_i is

$$f^l_{i,j,h} = softmax_{j\in V}(\frac{(q^l_{i,j,h})^T k^l_{i,j,h}}{\sqrt{N_p/K_{gff}}}). \tag{20}$$

After aggregating the neighbor representations of v_i in the h-th head, we concatenate the representations from each head to form

$$m^l_i = \Big\Vert_{h=1}^{K_{gff}} (\sum_{j=1}^{|V|} f^l_{i,j,h} \cdot v^l_{i,j,h}), \tag{21}$$

where $\Vert$ is the concatenation operation. The application of layer normalization (LayerNorm) is crucial for this training process and for expressing the capacity of attention (Brody *et al.*, 2023). The multi-modal representation $H^l_{moda,i}$ of v_i is updated based on LayerNorm, such that

$$\begin{aligned} n^l_i &= LayerNorm(W^l_1 m^l_i + H^{l-1}_{moda,i}), \\ H^l_{moda,i} &= LayerNorm(\sigma(W^l_2 n^l_i)W^l_3 + n^l_i), \end{aligned} \tag{22}$$

where $W^l_1, W^l_2, W^l_3 \in R^{N_p\times N_p}$ represent weight matrices.

2.7 Representation integration and optimization

The original features and representations learned from shallower layers retain the detailed information of the nodes, along with more abstract information are learned from deeper layers. By concatenating the original features and representations from each layer of NFF and GFF, the final representation of the drug d_i is are formed as

$$\tilde{H}_i = [H_{simi,i}, \Big\Vert_{l=0}^{L_{nff}} [H^l_{posi,i}, H^l_{topo,i}, H^l_{simi,i}], H_{moda,i}, \Big\Vert_{l=0}^{L_{gff}} H^l_{moda,i}]. \tag{23}$$

Likewise, we obtain the final representation $\tilde{H}_{j+N_d}$ of microbe m_j. Following the stack of linear layers and the non-linear activation function, $\tilde{H}_i$ and $\tilde{H}_{j+N_d}$ are combined to compute the association prediction score $pred_{i,j} \in R^2$, such that

$$\begin{aligned} u_{i,j} &= \sigma(W_{attr}[\tilde{H}_i, \tilde{H}_{j+N_d}] + b_{attr}), \\ pred_{i,j} &= softmax(W_{pred}u_{i,j} + b_{pred}), \end{aligned} \tag{24}$$

where W_{attr} (or W_{pred}) is the weight matrix and b_{attr} (or b_{pred}) is a bias vector. Then, $pred_{i,j} = [(pred_{i,j})_0, (pred_{i,j})_1]$, where $(pred_{i,j})_0$ represents the unrelated probability between d_i and m_j, and the associated probability is $(pred_{i,j})_1$. The loss of the association prediction is represented by

$$\gamma_{pred} = \frac{1}{|B|} \sum_{(i,j)\in B} \sum_{k=1}^{2} -(label(i,j)_k \cdot log((pred_{i,j})_k), \tag{25}$$

where B is the training example set and $label(i, j)$ is the association label of d_i and m_j. The final loss of NGMDA γ is the weighted sum of the ES loss γ_{proc} and the association prediction loss γ_{pred}, such that

$$\gamma = \varepsilon \cdot \gamma_{proc} + (1 - \varepsilon) \cdot \gamma_{pred}, \tag{26}$$

where the balance factor $\varepsilon \in [0, 1]$ is a hyper parameter.

3 Experimental Evaluation and Discussion

3.1 Evaluation metrics

The performance of NGMDA and other comparison methods is evaluated with five-fold cross-validation. All known associations between drugs and microbes are classified as positive samples, with associations equally divided into five parts. All unobserved microbe-drug associations are taken as negative samples to form a set of negative samples. Four positive examples and an equal number of negative examples are randomly selected from the negative sample to be utilized for training, and the remainder are test examples.

The area under the receiver operating characteristic curve (AUC) (Huang and Ling, 2005), the area under the precision-recall curve (AUPR) (Saito and Rehmsmeier, 2015), and the recall rate of the top k candidate microbes associated with drugs are selected as our evaluation indicators. If the association score between d_i and m_j is less than a threshold θ, then it is considered a negative sample. Otherwise, it is identified as a positive sample. The TPRs, FPRs, precisions, and recalls of each drug were calculated at different threshold θ, we calculated the average AUCs and average AUPRs of 1373 drugs for each fold. The five fold AUCs (or AUPRs) were averaged as the final AUC (or AUPR). Considering that high-ranking candidates may be chosen by biologists for humidity experiments, more positive samples are expected to appear as top-rank candidates. Hence, we compute a recall rate of the top k candidate microbes of drug d_i.

3.2 Parameter settings

NGMDA runs on a 2080ti server based on the PyTorch framework and is optimized with the Adam algorithm. The proposed model has some hyper-parameters including the steps of random walking, the layer numbers of NFF and that of GFF, and the balance factor of loss ϵ. We firstly establish the variation range for each hyper-parameter, and then select the value which obtains the best performance for the model as the final value of the hyper-parameter. To assess the effect of random walk step size on the prediction performance, the step size was selected from {1, 2, 4, 8, 16, 32}. The model achieves the highest AUC (AUC=0.944) and AUPR (AUPR=0.728) when step size is 2 (supplementary table ST1). The random walk steps are set to 2 for the topological embedding formation. For NFF and GFF, we fine-tuned the layer number within a range, {1, 2, 3}, and performed all the combinations of the layer number of NFF and GFF. As shown in the supplementary table ST2, the model gets the best performance when their layer numbers are 2. The balance factor ε regulates the importance of the loss of embedding enhancement strategy and that of the association prediction loss, and it was chosen from the range of $\{0, 0.1, \cdots, 0.5\}$. The supplementary table ST3 demonstrated the corresponding results and ε was set to 0.2 finally. The drug (microbe) similarity threshold, β, was selected from $\{0.5, 0.6, \cdots, 0.9\}$, and it was set to 0.9 in our experiment (supplementary table ST4). Parameter τ is utilized to balance the importance of the topology and position features, and τ varies from 0 to 1 with a step size of 0.2. The supplementary table ST5 indicates τ value of 0.4 is more favorable for the prediction performance of the model.

3.3 Ablation experiments

We perform ablation experiments to evaluate the contributions of position and topology feature learning (PTL), ES, NFF and GFF as listed in Table 1. For NGMDA without GFF, the AUC and AUPR metrics drop by 1.1% and 6.0%, respectively. The AUC and AUPR of NGMDA without NFF decrease by 1.0%

and 5.3%, respectively, compared to the whole model. The AUC and AUPR decrease by 0.6% and 4.6%, respectively, if NGMDA has no ES. The AUC and AUPR of our model achieve 0.9% and 1.5%, respectively, higher than NGMDA without PTL. We built the prediction model without multi-scale topological feature learning and the one without position feature learning, respectively. Their AUCs decreased by 0.8% and 0.4%, and their AUPR decreased by 1% and 0.5%, respectively. After the relationship type integration was eliminated from the prediction model, its AUC and AUPR decreased by 0.6% and 2.7%.

Table 1. Results of the ablation studies.

Networks	Average AUC	Average AUPR
NGMDA	0.944	0.728
NGMDA w/o PTL	0.935	0.713
NGMDA w/o ES	0.938	0.682
NGMDA w/o NFF	0.934	0.675
NGMDA w/o GFF	0.933	0.668
NGMDA w/o Topo	0.936	0.718
NGMDA w/o Posi	0.94	0.723
NGMDA w/o Rel	0.938	0.701

The ablation experiments indicate that merging node features of the heterogeneous graph contributes the most to model performance (Table 1). A possible reason is that some non-neighbouring nodes exist across the entire heterogeneous graph that are also closely related to the target node. NFF achieves the second most significant contribution, suggesting that the neighbouring node information of the target node is also important. The embedding enhancement strategy boosts the prediction performance, which suggests its value in reducing noise in the node embeddings and enhancing differences in the node distributions. Multi-scale topology and position features indicated the neighbors with multiple ranges and the location information of each node were important for the improved prediction performance. The experimental results (Table 1) also demonstrated the relationship type integration is helpful for improving the prediction performance.

3.4 Comparison with other methods

NGMDA is compared with five state-of-the-art microbe-drug association prediction methods, including GCNMDA (Long *et al.*, 2020a), EGATMDA (Long *et al.*, 2020b), GSAMDA (Tan *et al.*, 2022), GACNNMDA (Ma *et al.*, 2023), and SCSMDA (Tian *et al.*, 2023). NGMDA and five compared methods were trained and tested by using the same data separation during five-fold cross-validation. The hyper-parameters of these methods are set according to their corresponding literature. We briefly describe these comparison methods in the following.

- GCNMDA (Long *et al.*, 2020a): It established a microbe-drug heterogeneous network and integrated multiple kinds of similarities. These similarities were measured based on the chemical structures of drugs, the Gaussian interaction profiles of drugs (microbes), and the microbe sequences. The prediction model was constructed based on GCN and CRF.
- EGATMDA (Long *et al.*, 2020b): It constructed a microbe-disease-drug network and then inferred the microbe-drug associations by a hierarchical attention mechanism.
- GSAMDA (Tan *et al.*, 2022): The model calculated the drug (microbe) similarities based on the Gaussian interaction profiles and Hamming interaction profiles of drugs (or microbes), and learned the node features by the graph attention networks and sparse auto-encoder.
- GACNNMDA (Ma *et al.*, 2023): The multiple microbe-drug heterogeneous networks were constructed based on the Gaussian interaction and Hamming interaction profiles of drugs (microbes). The potential microbe-drug associations were identified by the convolutional neural networks.
- SCSMDA (Tian *et al.*, 2023): The model constructed the microbe-drug networks based on the microbe gene sequence information, the Gaussian kernel interaction profiles of drugs (or microbes), and the chemical structures of drugs. It learned the features of the microbe and drug nodes by graph contrastive learning.

Table 2. AUCs and AUPRs of different methods in comparison all the 1373 drugs.

Networks	AUC	AUPR
NGMDA	94.4%	72.8%
SCSMDA	91.6%	34.0%
GSAMDA	90.2%	24.7%
GACNNMDA	84.3%	19.6%
EGATMDA	94.0%	30.7%
GCNMDA	90.3%	31.5%

We first compute the AUC and AUPR and then calculate the average AUC and AUPR over 1373 drugs. As shown in Table 2, NGMDA achieves the best average AUC of 0.944, which is 0.4% higher than the second-best EGATMDA model, 4.1% better than GCNMDA, 10.1% over GACNNMDA, 4.2% superior to GSAMDA, and 2.8% greater than SCSMDA. NGMDA also produces the best average AUPR of 72.8%, which is 38.8%, 41.3%, 42.1%, 53.2%, and 48.1% better than SCSMDA, GCNMDA, EGATMDA, GACNNMDA, and GASMDA, respectively. We compute the average AUCs (AUPRs) of 1373 drugs per fold for NGMDA and each of the compared methods. To observe whether NGMDA's prediction performance is significantly higher than each compared method, the statistical test was conducted. NGMDA has 1373 AUCs (AUPRs) for the 1373 drugs, and the compared methods also have 1373 AUCs (AUPRs) for these drugs. The paired wilcoxon test was executed on NGMDA's AUCs (AUPRs) and the AUCs (AUPRs) of the compared methods (Table 3). The results indicated NGMDA obtained the significantly higher prediction performance than all the compared methods.

The performances of GCNMDA, GACNNMDA, and GSAMDA are not as good as NGMDA, EGATMDA, and SCSMDA. This outcome is likely because these learn node representations using simple models (e.g., GCN and GAT) without considering node or edge types in the microbe-drug heterogeneous graph. EGATMDA and SCSMDA learn the features of drugs and microbes from semantic information based on meta-paths. These models only focus on learning features of the neighbour nodes derived from meta-paths and do not consider the remaining nodes across the entire heterogeneous graph.

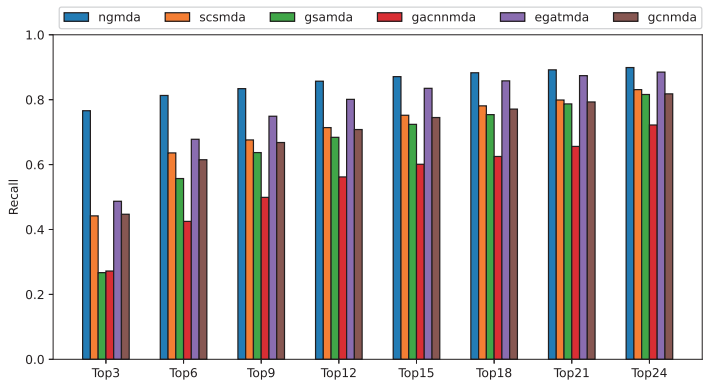

Fig. 4: The average recalls of drugs at different top k settings.

The average recalls under different top-k candidate microbes for all drugs are presented in Figure 4. NGMDA outperforms all other methods at different top cutoffs due to its enhanced embedding of the nodes and fusing the features of neighbor nodes and the whole heterogeneous graph. When k = 3, our model achieves the highest recall rate of 76.6%, where the second-best 48.7% is attained by EGATMDA. SCSMDA achieves the fourth-best result with a recall rate of 44.7%, which is 0.5% below GCNMDA. When k is 6, 9, and 12, NGMDA maintains the best recall values of 81.3%, 83.4%, and 85.7%, respectively. The second performer is EGATMDA with recall rates of 67.8%, 74.9% and 80.1%, respectively. SCSMDA surpasses GCNMDA with recall rates of 63.6%, 67.6%, and 71.4%, respectively, while the recall rates of GCNMDA are lower at 61.5%, 66.8%, and 70.8%, respectively. GSAMDA does not perform well with recall rates of 55.7%, 63.7%, and 68.4%, respectively, while still being consistently higher than GACNNMDA, which obtained the lowest recall rates of 42.5%, 49.9% and 56.2%, respectively.

Table 3. The paired Wilcoxon test result on AUCs and AUPRs of 1373 drugs comparing NGMDA with other compared methods.

	GCNMDA	EGATMDA	GACNNMDA	GSAMDA	SCSMDA
p-value of AUCs	4.21e-155	1.70e-59	7.26e-159	4.38e-150	1.40e-151
p-value of AUPRs	1.47e-186	2.26e-156	6.63e-196	1.30e-186	1.82e-215

Table 4. The top-20 candidate microbes of Ciprofloxacin.

Rank	Microbe name	Evidence	Rank	Microbe name	Evidence
1	Candida albicans	PMID: 31471074	11	Bacillus subtilis	MDAD
2	Pseudomonas aeruginosa	aBiofilm, MDAD	12	Actinomyces oris	Unconfirmed
3	Staphylococcus aureus	aBiofilm, MDAD	13	Human immunodeficiency virus 1	PMID: 9566552
4	Escherichia coli	aBiofilm, MDAD	14	Streptococcus sanguis	PMID: 11347679
5	Streptococcus mutans	PMID: 30468214	15	Stenotrophomonas maltophilia	aBiofilm, MDAD
6	Staphylococcus epidermis	PMID: 10632381	16	Haemophilus influenzae	MDAD
7	Staphylococcus epidermidis	PMID: 28481197	17	Listeria monocytogenes	PMID: 28355096
8	Salmonella enterica	PMID: 26933017	18	Burkholderia cenocepacia	PMID: 27799222
9	Vibrio harveyi	PMID: 27247095	19	Streptococcus pneumoniae	PMID: 26100702
10	Enterococcus faecalis	PMID: 27790716	20	Serratia marcescens	PMID: 23751969

Table 5. The top-20 candidate microbes of Moxifloxacin.

Rank	Microbe name	Evidence	Rank	Microbe name	Evidence
1	Pseudomonas aeruginosa	PMID: 31691651	11	Staphylococcus epidermidis	PMID: 11249827
2	Staphylococcus aureus	PMID: 31689174	12	Candida albicans	aBiofilm, MDAD
3	Escherichia coli	PMID: 31542319	13	Streptococcus pneumoniae	PMID: 22407042
4	Vibrio harveyi	Unconfirmed	14	Serratia marcescens	Unconfirmed
5	Bacillus subtilis	PMID: 30036828	15	Acinetobacter baumannii	PMID: 12951327
6	Listeria monocytogenes	PMID: 28739228	16	Actinomyces oris	PMID: 26538502
7	Salmonella enterica	PMID: 22151215	17	Clostridium perfringens	PMID: 29486533
8	Stenotrophomonas maltophilia	aBiofilm, MDAD	18	Klebsiella pneumoniae	PMID: 27257956
9	Burkholderia cenocepacia	PMID: 28355096	19	Burkholderia pseudomallei	PMID: 15731198
10	Burkholderia multivorans	Unconfirmed	20	Haemophilus influenzae	MDAD

Table 6. The top-20 candidate microbes of Vancomycin.

Rank	Microbe name	Evidence	Rank	Microbe name	Evidence
1	Staphylococcus aureus	MDAD;aBiofilm	11	Streptococcus mutans	PMID: 464571
2	Pseudomonas aeruginosa	PMID: 26980934	12	Stenotrophomonas maltophilia	Unconfirmed
3	Escherichia coli	PMID: 33468474	13	Streptococcus pneumoniae	PMID: 10376600
4	Staphylococcus epidermidis	PMID: 20685088	14	Acinetobacter baumannii	PMID: 23422916
5	Bacillus subtilis	PMID: 14165485	15	Actinomyces oris	PMID: 26538502
6	Enterococcus faecalis	PMID: 15528891	16	Salmonella enterica	Unconfirmed
7	Vibrio harveyi	PMID: 25066453	17	Klebsiella pneumoniae	Unconfirmed
8	Listeria monocytogenes	PMID: 10588323	18	Clostridium perfringens	PMID: 16870765
9	Burkholderia cenocepacia	Unconfirmed	19	Serratia marcescens	Literature (Ali, 2018)
10	Burkholderia multivorans	Unconfirmed	20	Streptococcus sanguis	PMID: 7287904

3.5 Case studies on three drugs

To confirm NGMDA's discovery potential of drug related microbial candidates, case studies with Ciprofloxacin, Moxifloxaxin, and Vancomycin are performed. Ciprofloxacin treats skin infections, typhoid fever, pneumonia, endocarditis, and other bacterial infections. Moxifloxacin treats pneumonia, tuberculosis, sinusitis, and chronic bronchitis. Vancomycin is an antibiotic that treats bloodstream infections, endocarditis, and orthopedic infections. All the known microbe-drug associations and the randomly selected equal number of unobserved microbe-drug associations were utilized to train the model for case studies. Candidate microbes are obtained for each of these drugs, and we collected the top 20 candidates, as listed in Table 4, 5, and 6.

The MDAD (Sun *et al.*, 2018) provides microbe-drug associations that were verified by experimental or clinical studies. The aBiofilm database (Rajput *et al.*, 2018) organizes data on anti-biofilm agents disrupting biofilms, covering 1720 drugs and 140 microbes. We use MDAD, aBiofilm database, and literature to verify the microbe-drug association prediction results of NGMDA. Among the top 20 candidate microbes related to Ciprofloxacin, six are recorded by MDAD, and four are contained in the aBiofilm database, which suggests that these microbes are indeed associated with the drug Ciprofloxacin, and these 13 candidates are further confirmed by the literature. For example, several microbes, including Candida albicans, Human immunodeficiency virus 1, Streptococcus mutans, and Streptococcus pneumoniae, are inhibited (or killed) by Ciprofloxacin (Gollapudi *et al.*, 1998; Dridi *et al.*, 2015; Zhang *et al.*, 2019; Hacioglu *et al.*, 2019). The two microbes, Staphylococcus epidermidis and Salmonella enterica, were validated to be highly susceptible to Ciprofloxacin (Eibach *et al.*, 2016; Szczuka *et al.*, 2017). In addition, Vibrio harveyi, Enterococcus faecalis, and Listeria monocytogenes are identified as Ciprofloxacin-resistant microbes (Stalin and Srinivasan, 2016; Kim and Woo, 2017; Escolar *et al.*, 2017). For the microbe candidates related to Moxifloxacin in Table 5, three candidates are included in MDAD, two in the aBiofilm database and 14 candidates are supported by the literature. Considering the candidate microbes of Vancomycin in Table 6, Staphylococcus is confirmed by the MDAD and aBiofilm databases, and 14 candidates are supported by literature. Among all 60 microbe candidates, nine are unconfirmed, which indicates that no relevant evidence is found to support their association. The above analysis demonstrates that NGMDA can discover potential candidate microbes for target drugs under study.

3.6 Prediction of novel microbe-drug associations

NGMDA is implemented to predict the potential candidate microbes for all drugs. The top-ranked 20 microbe candidates are listed in the supplementary file SF2, which can be leveraged by biologists to screen reliable candidate microbes.

4 Conclusion

We proposed a novel microbe-drug association prediction model to encode node neighborhood topologies across multiple scales and perform graph inference by propagating different types of connections and information about the nodes. The multi-scale topology feature is formed by estimating the probability that a random walker accesses itself in different steps. The established node embedding strategy enhances the representations of microbe and drug nodes that form the specific distribution of the corresponding node. The NFF combines the features of different types of neighbors and target nodes by adaptively evaluating the weights of the position features, topology features, and original features of the neighbor nodes. The long-distance connections and encoding of the relationship types between the nodes through GFF enable the knowledge propagation of the entire graph and the capture of diverse relationships. Cross-validation experimental results on public datasets suggest the superiority and effectiveness of NGMDA. The average recall rate of drugs and case analyses of experimental results further demonstrate that NGMDA provides reliable microbe candidates for related drugs under investigation.

Funding

This work is supported by the Natural Science Foundation of China (62372282, 61972135, 62172143), STU Scientific Research Initiation Grant (NTF22032), and the Natural Science Foundation of Heilongjiang Province (LH2023F044).

References

Algavi, Y. M. and Borenstein, E. A data-driven approach for predicting the impact of drugs on the human microbiome. *Nature Communications*, 14(1): 3614, 2023.

Ali, A. Bacterial sensitivity of serratia marcescens against antibiotics. *International Journal of Scientific & Engineering Research*, 9(9):361–363, 2018.

Brody, S., Alon, U., and Yahav, E. On the expressivity role of LayerNorm in transformers' attention. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 14211–14221, Toronto, Canada, jul 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.895.

Diao, C. and Loynd, R. Relational attention: Generalizing transformers for graph-structured tasks. *arXiv preprint arXiv:2210.05062*, 2022.

Dridi, B., Lupien, A., Bergeron, M. G., Leprohon, P., and Ouellette, M. Differences in antibiotic-induced oxidative stress responses between laboratory and clinical isolates of streptococcus pneumoniae. *Antimicrobial agents and chemotherapy*, 59(9):5420–5426, 2015.

Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., and Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nature communications*, 8(1):1784, 2017.

Dwivedi, V. P., Luu, A. T., Laurent, T., Bengio, Y., and Bresson, X. Graph neural networks with learnable structural and positional representations. In *International Conference on Learning Representations*, 2022.

Eibach, D., Al-Emran, H. M., Dekker, D. M., Krumkamp, R., Adu-Sarkodie, Y., Cruz Espinoza, L. M., Ehmen, C., Boahen, K., Heisig, P., Im, J., et al. The emergence of reduced ciprofloxacin susceptibility in salmonella enterica causing bloodstream infections in rural ghana. *Clinical Infectious Diseases*, 62(suppl_1):S32–S36, 2016.

Escolar, C., Gómez, D., Del Carmen Rota García, M., Conchello, P., and Herrera, A. Antimicrobial resistance profiles of listeria monocytogenes and listeria innocua isolated from ready-to-eat products of animal origin in spain. *Foodborne pathogens and disease*, 14(6):357–363, 2017.

Gao, H., Sun, J., Wang, Y., Lu, Y., Liu, L., Zhao, Q., and Shuai, J. Predicting metabolite–disease associations based on auto-encoder and non-negative matrix factorization. *Briefings in Bioinformatics*, 24(5):bbad259, 2023.

Gollapudi, S., Kim, C. H., Roshanravan, B., and Gupta, S. Ciprofloxacin inhibits activation of latent human immunodeficiency virus type 1 in chronically infected promonocytic u1 cells. *AIDS research and human retroviruses*, 14 (6):499–504, 1998.

Hacioglu, M., Haciosmanoglu, E., Birteksoz-Tan, A. S., Bozkurt-Guzel, C., and Savage, P. B. Effects of ceragenins and conventional antimicrobials on candida albicans and staphylococcus aureus mono and multispecies biofilms. *Diagnostic microbiology and infectious disease*, 95(3):114863, 2019.

Hattori, M., Tanaka, N., Kanehisa, M., and Goto, S. Simcomp/subcomp: chemical structure search servers for network analyses. *Nucleic acids research*, 38(suppl_2):W652–W656, 2010.

Huang, J. and Ling, C. X. Using auc and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3): 299–310, 2005.

Kim, M.-C. and Woo, G.-J. Characterization of antimicrobial resistance and quinolone resistance factors in high-level ciprofloxacin-resistant enterococcus faecalis and enterococcus faecium isolates obtained from fresh produce and fecal samples of patients. *Journal of the Science of Food and Agriculture*, 97 (9):2858–2864, 2017.

Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., Pon, A., Cox, J., Chin, N. E., Strawbridge, S. A., et al. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic Acids Research*, 52(D1):D1265–D1275, 2024.

Li, F., Zhang, Z., Guan, J., and Zhou, S. Effective drug–target interaction prediction with mutual interaction neural network. *Bioinformatics*, 38(14): 3582–3589, 2022.

Long, Y., Wu, M., Kwoh, C. K., Luo, J., and Li, X. Predicting human microbe–drug associations via graph convolutional network with conditional random field. *Bioinformatics*, 36(19):4918–4927, 2020a.

Long, Y., Wu, M., Liu, Y., Kwoh, C. K., Luo, J., and Li, X. Ensembling graph attention networks for human microbe–drug association prediction. *Bioinformatics*, 36(Supplement_2):i779–i786, 2020b.

Lv, Q., Ding, M., Liu, Q., Chen, Y., Feng, W., He, S., Zhou, C., Jiang, J., Dong, Y., and Tang, J. Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 1150–1160, 2021.

Ma, Q., Tan, Y., and Wang, L. Gacnnmda: a computational model for predicting potential human microbe-drug associations based on graph attention network and cnn-based classifier. *BMC bioinformatics*, 24(1):35, 2023.

Meng, R., Yin, S., Sun, J., Hu, H., and Zhao, Q. scaaga: Single cell data analysis framework using asymmetric autoencoder with gene attention. *Computers in Biology and Medicine*, 165:107414, 2023.

Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., et al. The human tumor microbiome is composed of tumor type–specific intracellular bacteria. *Science*, 368(6494):973–980, 2020.

Peng, W., Lan, W., Zhong, J., Wang, J., and Pan, Y. A novel method of predicting microrna-disease associations based on microrna, disease, gene and environment factor networks. *Methods*, 124:69–77, 2017.

Peng, W., Chen, T., and Dai, W. Predicting drug response based on multi-omics fusion and graph convolution. *IEEE Journal of Biomedical and Health Informatics*, 26(3):1384–1393, 2021.

Peng, W., Che, Z., Dai, W., Wei, S., and Lan, W. Predicting mirna-disease associations from mirna-gene-disease heterogeneous network with multi-relational graph convolutional network model. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022a.

Peng, W., Liu, H., Dai, W., Yu, N., and Wang, J. Predicting cancer drug response using parallel heterogeneous graph convolutional networks with neighborhood interactions. *Bioinformatics*, 38(19):4546–4553, 2022b.

Peng, W., Tang, Q., Dai, W., and Chen, T. Improving cancer driver gene identification using multi-task learning on graph convolutional network. *Briefings in Bioinformatics*, 23(1):bbab432, 2022c.

Rackaityte, E. and Lynch, S. V. The human microbiome in the 21st century. *Nature communications*, 11(1):5256, 2020.

Rajput, A., Thakur, A., Sharma, S., and Kumar, M. abiofilm: a resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance. *Nucleic acids research*, 46(D1):D894–D900, 2018.

Saito, T. and Rehmsmeier, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.

Stalin, N. and Srinivasan, P. Molecular characterization of antibiotic resistant vibrio harveyi isolated from shrimp aquaculture environment in the south east coast of india. *Microbial pathogenesis*, 97:110–118, 2016.

Sun, Y.-Z., Zhang, D.-H., Cai, S.-B., Ming, Z., Li, J.-Q., and Chen, X. Mdad: a special resource for microbe-drug associations. *Frontiers in cellular and infection microbiology*, 8:424, 2018.

Szczuka, E., Jabłońska, L., and Kaznowski, A. Effect of subinhibitory concentrations of tigecycline and ciprofloxacin on the expression of biofilm-associated genes and biofilm structure of staphylococcus epidermidis. *Microbiology*, 163(5):712, 2017.

Tan, Y., Zou, J., Kuang, L., Wang, X., Zeng, B., Zhang, Z., and Wang, L. Gsamda: a computational model for predicting potential microbe–drug associations based on graph attention network and sparse autoencoder. *BMC bioinformatics*, 23(1):492, 2022.

Tian, Z., Yu, Y., Fang, H., Xie, W., and Guo, M. Predicting microbe–drug associations with structure-enhanced contrastive learning and self-paced negative sampling strategy. *Briefings in Bioinformatics*, 24(2):bbac634, 2023.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. *International Conference on Learning Representations*, 2018.

Wang, D., Guo, Q., Yuan, Y., and Gong, Y. The antibiotic resistance of helicobacter pylori to five antibiotics and influencing factors in an area of china with a high risk of gastric cancer. *BMC microbiology*, 19:1–10, 2019.

Wang, S., Hui, C., Zhang, T., Wu, P., Nakaguchi, T., and Xuan, P. Graph reasoning method based on affinity identification and representation decoupling for predicting lncrna-disease associations. *Journal of Chemical Information and Modeling*, 63(21):6947–6958, 2023a.

Wang, T., Sun, J., and Zhao, Q. Investigating cardiotoxicity related with herg channel blockers using molecular fingerprints and graph attention mechanism. *Computers in Biology and Medicine*, 153:106464, 2023.

Wang, W., Zhang, L., Sun, J., Zhao, Q., and Shuai, J. Predicting the potential human lncrna–mirna interactions based on graph convolution network with conditional random field. *Briefings in Bioinformatics*, 23(6):bbac463, 2022.

Yin, B., Wang, X., Yuan, F., Li, Y., and Lu, P. Research progress on the effect of gut and tumor microbiota on antitumor efficacy and adverse effects of chemotherapy drugs. *Frontiers in microbiology*, 13:3784, 2022.

Zhang, R., Jones, M. M., Moussa, H., Keskar, M., Huo, N., Zhang, Z., Visser, M. B., Sabatini, C., Swihart, M. T., and Cheng, C. Polymer–antibiotic conjugates as antibacterial additives in dental resins. *Biomaterials science*, 7(1):287–295, 2019.

Zhu, S., Jiang, Y., Xu, K., Cui, M., Ye, W., Zhao, G., Jin, L., and Chen, X. The progress of gut microbiome research related to brain disorders. *Journal of neuroinflammation*, 17:1–20, 2020.

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*, 570(7762):462–467, 2019.