Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category

OXFORD

Systems Biology

Ensembling graph attention networks for human microbe-drug association prediction

Yahui Long [1,2], Min Wu [3], Yong Liu [4], Chee Keong Kwoh [2], Jiawei Luo [1,*] and Xiaoli Li [3,*]

[1]College of Computer Science and Electronic Engineering, Hunan University, Changsha 410000, China
[2]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[3]Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 138632, Singapore
[4]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY), Nanyang Technological University, Singapore 639798, Singapore

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Human microbes get closely involved in an extensive variety of complex human diseases and become new drug targets. *In silico* methods for identifying potential microbe-drug associations provide an effective complement to conventional experimental methods, which can not only benefit screening candidate compounds for drug development, but also facilitate novel knowledge discovery for understanding microbe-drug interaction mechanisms. On the other hand, the recent increased availability of accumulated biomedical data for microbes and drugs provides a great opportunity for a machine learning approach to predict microbe-drug associations. We are thus highly motivated to integrate these data sources to improve prediction accuracy. In addition, it is extremely challenging to predict interactions for new drugs or new microbes, which have no existing microbe-drug associations.

Results: In this work, we leverage various sources of biomedical information and construct multiple networks (graphs) for microbes and drugs. Then, we develop a novel ensemble framework of graph attention networks with a hierarchical attention mechanism for microbe-drug association prediction from the constructed multiple microbe-drug graphs, denoted as EGATMDA. In particular, for each input graph, we design a graph convolutional network with node-level attention to learn embeddings for nodes (i.e., microbes and drugs). To effectively aggregate node embeddings from multiple input graphs, we implement graph-level attention to learn the importance of different input graphs. Experimental results under different cross-validation settings (e.g., the setting for predicting associations for new drugs) showed that our proposed method outperformed seven state-of-the-art methods. Case studies on predicted microbe-drug associations further demonstrated the effectiveness of our proposed EGATMDA method.

Availability: Source codes and supplementary materials are available at: `https://github.com/longyahui/EGATMDA/`

Contact: luojiawei@hnu.edu.cn and xlli@i2r.a-star.edu.sg

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Accumulated clinical and experimental reports confirm that human microbes residing in and on the human body have close interactions with human hosts (Huttenhower *et al.*, 2012; Sommer and Bäckhed, 2013). Microbe communities, mainly comprised of bacteria, viruses, archaea, fungi, and protozoa, are shown to play a fundamental role in maintaining human health, such as facilitating the metabolism (Ventura *et al.*, 2009), producing essential vitamins and gene products (Kau *et al.*, 2011), and

protecting against invasion from pathogens (Sommer and Bäckhed, 2013). Therefore, the dysbiosis or imbalance of microbe communities can lead to various human infection diseases (Huttenhower *et al.*, 2012; Sommer and Bäckhed, 2013), such as obesity (Zhang *et al.*, 2009), diabetes (Wen *et al.*, 2008), systemic inflammatory response syndrome (Mshvildadze *et al.*, 2010) and even cancer (Schwabe and Jobin, 2013). As such, microbe is considered as a new therapeutic target for precision medicine (Kashyap *et al.*, 2017).

However, with the increasing emergence of drug-resistant microbes, there is an urgent need to identify microbe-drug associations on a large scale for drug development. Recent studies have shown that microbes play an important role in modulating drug activity and toxicity (Zimmermann *et al.*, 2019; Lewis and Strandwitz, 2019), and drugs can also, in turn, change the diversity and function of microbe communities. Furthermore, more and more microbe-drug associations have been reported in the literature. For example, Haiser *et al.* (2013) demonstrated that gut Actinobacterium *Eggerthella lenta* can result in the inactivation of the cardiac drug *digoxin*. In addition, the microbial β-glucuronidases in the gut assisted the treatment of irinotecan for colorectal cancer by reactivating the excreted, inactive metabolite (Guthrie *et al.*, 2017). Zimmermann *et al.* (2019) revealed that gut bacterium *Bacteroides thetaiotaomicron* is a prolific drug metabolizer, which can metabolize multiple kinds of drugs, such as *diltiazem*. While these microbe-drug associations are detected based on experimental methods, it is actually very difficult for them to select target microbes, leading to slow progress for developing new drugs. To tackle this problem, most efforts have been devoted to the optimization or combination of already known compounds (i.e. drug repurposing and drug combination) (Durand *et al.*, 2019). However, the emerging of drug-resistance brings a new challenge for drug development. It is thus highly desired to develop an effective method to infer candidate target microbes for new drugs, which is essential for drug discovery and repositioning, as well as personalized medicine. As conventional wet-lab experiments are time-consuming, labor-intensive and expensive, *in silico* methods can thus serve as promising complements to computationally provide accurate predictions of microbe-drug associations.

Recently, a database called MDAD has been curated for clinically and experimentally verified microbe-drug associations (Sun *et al.*, 2018). In addition, we can further derive potential microbe-drug associations by linking the microbe-disease associations and drug-disease associations from public databases, such as DrugBank and Disbiome. As graphs are well-known structure to capture different kinds of relationships, we can use different graphs to model the microbe-drug associations derived from different sources. The above graph data for microbes and drugs provide a golden opportunity for us to leverage graph-based deep learning techniques for predicting their associations. In particular, graph attention network (GAT) (Veličković *et al.*, 2017) shows great potential in modeling complex graph data, which has been successfully applied for node classification (Wang *et al.*, 2019), social influence analysis (Qiu *et al.*, 2018) and recommender system (Wu *et al.*, 2019). It is thus natural for us to customize GAT for novel microbe-drug association prediction. However, there currently exist two main challenges in this important task. Firstly, with the limitation of screening technologies, many drugs or microbes do not have known microbe-drug associations, which are denoted as *new* drugs or *new* microbes. As we have no training data for these new drugs or microbes, it is thus very challenging for the trained model (e.g., GAT) to predict their associations. Secondly, as we mentioned above, we construct multiple graphs for microbes and drugs. Different graphs may have different biological meanings and the same node (i.e., microbe or drug) may play different roles in different graphs. How to effectively integrate multiple graphs remains a computational challenge.

To address the above issues, we propose a novel ensemble framework of graph attention networks for microbe-drug association prediction, named EGATMDA as shown in Figure 1. First, we derive comprehensive features for both microbes and drugs. More importantly, we extract potential or virtual microbe-drug associations for new drugs (microbes) based on the meta-paths in different input graphs. For example, we generate virtual microbe-drug associations using the meta-path 'microbe-disease-drug' in microbe-disease-drug network as shown in Figure 1. With the derived features and virtual interactions for new drugs (microbes), GAT is thus able to propagate the information from local neighbors to learn their representations and then make reasonable predictions for them. Second, we develop a hierarchical attention mechanism, i.e., node-level attention and graph-level attention, to learn node representations from multiple graphs. In particular, we design a graph attention network with node-level attention to learn representations for nodes (i.e., microbes and drugs) in each input graph. To effectively aggregate the node representations from multiple input graphs, we further implement graph-level attention to learn the importance of different input graphs. Experimental results under different cross-validation settings showed that our method consistently outperformed seven state-of-the-art methods.

Overall, our main contributions are summarized as follows.

- We constructed *three different genres of networks* and also derived comprehensive features for microbes and drugs, enabling accurate predictions for *new* drugs and *new* microbes.
- We proposed a novel *ensemble framework of graph attention networks* for predicting microbe-drug associations. To the best of our knowledge, this is the first attempt to adopt graph attention network (GAT) to tackle this important problem.
- We designed a *hierarchical attention mechanism* in our ensemble framework to effectively learn node embeddings from multiple input graphs for microbe-drug association prediction.
- Our comprehensive experimental results and case studies demonstrated the proposed EGATMDA method outperformed seven state-of-the-art methods significantly on the benchmark MDAD dataset.

2 Related Work

In this section, we first present graph neural networks, including graph convolutional networks (GCN) and graph attention networks (GAT), and their applications in bioinformatics. To our best knowledge, so far no work has used GCN or GAT for predicting microbe-drug associations.

Graph Convolutional Network (GCN) (Kipf and Welling, 2016), which aims to learn node embeddings/representations by implementing convolution operation on a graph based on the properties of neighborhood nodes, has recently drawn extensive attention and demonstrated superior performance in various tasks, such as text classification (Yao *et al.*, 2019), recommender system (Liu *et al.*, 2020) and relation extraction (Cai *et al.*, 2020; Zhang *et al.*, 2018). Graph Attention Networks (GAT) (Veličković *et al.*, 2017; Wang *et al.*, 2019) is an extension of graph convolutional operations, which assigns different weights to different neighbors with masked self-attentional layers. This operation enables the model to filter our noise and focus on more important neighbors. Due to the powerful capability, graph attention networks have been successfully applied for node/text classification (Wang *et al.*, 2019; Linmei *et al.*, 2019), social influence analysis (Qiu *et al.*, 2018) and recommender system (Wu *et al.*, 2019).

Recently, researchers have developed numerous GCN/GAT-based approaches to tackle various bioinformatics tasks. For example, Zitnik *et al.* (2018) used a graph convolutional network for predicting polypharmacy side effects based on multimodal data. Additionally, Zhao *et al.* (2019) proposed a novel framework of graph convolutional attention network (GCAN) to predict potential disease-RNAs associations. Very

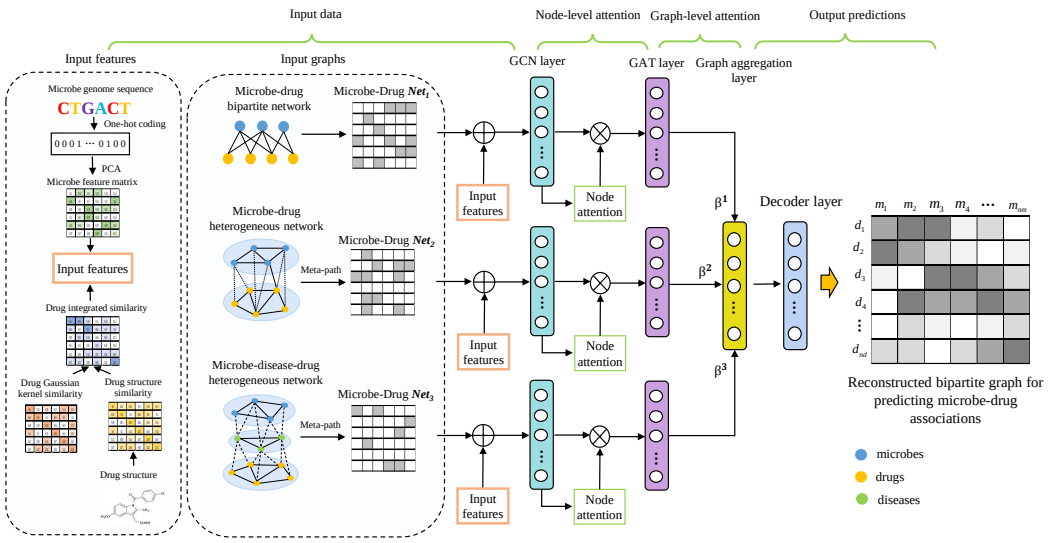


Fig. 1. The overall architecture of EGATMDA for microbe-drug association prediction.

recently, Han *et al.* (2019) developed a new framework named GCN-MF for the identification of disease-gene associations by incorporating graph convolution network with matrix factorization. Ravindra *et al.* (2020) leveraged graph attention networks to deal with the problem of disease state prediction from single-cell data. While above methods achieved relatively good prediction performance, they failed to consider abundant prior biological knowledge, which includes rich semantic information of nodes. Furthermore, most only focused on the importance of immediate neighbors (i.e., the first-order neighbor) and ignored the importance of high-order neighbors in existing GCN/GAT based methods.

3 Materials

3.1 Re-construction of three networks

We collect known microbe-drug associations from the MDAD database (`http://www.chengroup.cumt.edu.cn/MDAD/`) (Sun *et al.*, 2018), where there are 5505 clinically reported or experimentally validated microbe-drug associations between 1388 drugs and 174 microbes. After removing redundant information, we finally derive a *microbe-drug bipartite network* Net_1 (shown at the top of the second column of Figure 1), involving 2470 associations between 1373 drugs and 173 microbes.

We further derive two heterogeneous networks, namely *microbe-drug heterogeneous network* and *microbe-disease-drug network*, from multiple databases, such as DrugBank (Wishart *et al.*, 2018), HMDAD (Ma *et al.*, 2017) and CTD (Davis *et al.*, 2019). In particular, microbe-drug heterogeneous network contains drug-drug interactions and microbe-microbe interactions and microbe-drug associations. Based on the meta-paths 'drug-drug-microbe' and 'microbe-microbe-drug', we can obtain virtual microbe-drug associations and the corresponding network is denoted as Net_2. On the other hand, microbe-disease-drug network contains drug-disease associations, microbe-disease associations, and disease-disease relationships. Similarly, we derive corresponding microbe-drug network Net_3 based on the meta-path 'microbe-disease-drug'. Net_2 and Net_3 with virtual microbe-drug associations can help to better learn the representations for microbes and drugs. Overall, the statistics of the three microbe-drug networks above are shown in Table 1. More information on network construction could be found in the Supplementary Materials.

Table 1. The statistics for each microbe-drug network.

	# Microbes	# Drugs	# Associations
Net_1	173	1373	2470
Net_2	123	1228	17182
Net_3	29	92	394

For each graph, we define a binary matrix $I \in \mathbb{R}^{nd \times nm}$ to represent microbe-drug associations, with nd and nm representing the numbers of drugs and microbes respectively. If drug d_i is associated with microbe m_j, I_{ij} is equal to 1; 0 otherwise. Taking Net_1 as an example, we define its adjacent matrix $A \in \mathbb{R}^{(nd+nm)\times(nd+nm)}$ as follows :

$$A = \begin{bmatrix} 0 & I \\ I^T & 0 \end{bmatrix}. \tag{1}$$

3.2 Features for drugs and microbes

We downloaded genome sequences in FASTA format in database NCBI (`https://www.ncbi.nlm.nih.gov/genome/`) for 131 out of 173 microbes. In this work, we use one-hot coding to encoder the raw genome sequences and align each sequence with the longest one with 0 as padding (without losing information). For those microbes without sequence information available, we define their feature values as the average ones of all other known microbes. Then, Principal Component Analysis (PCA) (Chen *et al.*, 2002) is deployed on the binary matrix to extract more useful features and reduce dimension. We denote the microbe feature matrix as $F_m \in \mathbb{R}^{nm \times k}$ with k representing the dimension of microbe features. For drugs, we treat the integrated similarity, obtained by aggregating drug structure similarity and Gaussian kernel drug similarity, as drug features. Then, we obtain a drug feature matrix $F_d \in \mathbb{R}^{nd \times nd}$. The whole process to generate drug features is shown in the first column of Figure 1 or find more details about drug feature extraction in the section 1.8 in the supplementary materials. In consistent with the bipartite network in Equation 1, the feature matrix $X \in \mathbb{R}^{(nd+nm)\times(nd+k)}$ for microbes and drugs is described as follows:

$$X = \begin{bmatrix} F_d & 0 \\ 0 & F_m \end{bmatrix}. \tag{2}$$

It should be noted that this feature matrix is shared across three input networks as shown in Figure 1.

4 Methods

Here we present our proposed EGATMDA framework, which consists of three steps as shown in the right part of Figure 1. Firstly, we learn graph-specific node embeddings from each input microbe-drug network. Secondly, we aggregate the learned node embeddings and focus on important information (remove irrelevant noise) via graph-level attention. Finally, we learn a decoder for microbe-drug graph reconstruction based on the learned representations to predict novel microbe-drug associations. Next, we introduce each of the above steps in detail.

4.1 Node-level attention for node representation learning

After obtaining the adjacent matrix A in Section 3.1 and feature matrix X in Section 3.2, we can utilize them to learn node representations. Graph convolutional network (GCN) is an effective tool for graph-structured data and successfully applies to various real-world applications. Here, we first leverage GCN to learn the node representations by aggregating representations of their immediate neighbors. Suppose that every node is connected to itself (i.e. self-loop), the normalized adjacent matrix $\widetilde{A}$ of A could be defined as $\widetilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, where D is a diagonal matrix with diagonal elements being $D_{ii} = \sum_{j=1}^{n} A_{ij}$. Feature matrix X is normalized to avoid bias introduced by different nodes. After that, the graph convolutional layer, i.e., the first layer, is formulated as follows:

$$H^{\Phi} = \mathrm{ReLU}(\widetilde{A}XW_c + B), \tag{3}$$

where $H^{\Phi} \in \mathbb{R}^{(nd+nm)\times l}$ is representation matrix of graph Φ, with l representing embedding dimension; $W_c \in \mathbb{R}^{(nd+nm)\times l}$ and $B \in \mathbb{R}^{(nd+nm)\times l}$ are trainable parameters and bias matrices, respectively.

After the graph convolution layer, we derive the node representations in Equation 3. We further introduce a graph attention layer to update the node representations based on graph attention network (GAT) (Wang *et al.*, 2019; Linmei *et al.*, 2019), which aims to preserve the importance of the neighbors for node representation learning. Given a node, GAT first learns the importance of its neighbors, and subsequently fuse the features of all the neighbors according to their attention scores. In particular, the attention score e_{ij}^{Φ} for an association pair between drug d_i and microbe m_j is computed by a fully connected neural network in Equation 4:

$$e_{ij}^{\Phi} = (W_t h_j)^T \tanh\Big((W_t h_i + b)\Big), \tag{4}$$

where h represents the node representation derived from the graph convolutional layer; W_t and b are trainable weight and bias parameters respectively, which are both shared for all graph-specific microbe-drug pairs. We further normalized the attention scores using the following softmax function, where $\mathcal{N}_i^{\Phi}$ is the set of neighbors of node i.

$$\alpha_{ij}^{\Phi} = \frac{\exp(e_{ij}^{\Phi})}{\sum_{k\in\mathcal{N}_i^{\Phi}} \exp(e_{ik}^{\Phi})}. \tag{5}$$

Eventually, we derive $Z^{\Phi} \in \mathbb{R}^{(nd+nm)\times l}$ as the representation matrix of graph Φ, where the graph-specific representation of node i, z_i^{Φ}, is derived as follows:

$$z_i^{\Phi} = \sigma\left(\sum_{j\in\mathcal{N}_i^{\Phi}} \alpha_{ij}^{\Phi} \cdot h_j\right), \tag{6}$$

where σ denotes the nonlinear activation function, i.e., ReLU.

4.2 Graph-level attention for representation aggregation

Each node (i.e. microbe and drug) in different graphs may include diverse semantic information. In order to effectively integrate the information and remove noise from different graphs, we propose a graph-level attention mechanism to aggregate multiple graph-specific representations for each node. Given a node, it has an input feature vector as shown in Equation 2 and a graph-specific representation in Equation 6. Empirically, greater relevance between these two types of features/representations indicate that the graph would play a more important role in driving the final representation for the node. Therefore, we learn the importance of each graph according to the relevance between the above two types of features for all the nodes. The attention score is defined as follows.

$$w_i^{\Phi} = \sum_i v^T \tanh(W_z \cdot z_i^{\Phi} + W_x \cdot x_i), \tag{7}$$

where z_i^{Φ} and x_i are the graph-specific representation and input feature vector for node i, respectively. W_z and W_x are trainable parameter matrices and v is also a trainable vector. w_i^{Φ} is the attention score of graph Φ, indicating the importance of the representation z_i^{Φ} to the final representation of node i . To make coefficients of different graphs comparable, we normalize the attention scores for all the graphs using the softmax function in Equation 8.

$$\beta_i^{\Phi} = \frac{\exp(w_i^{\Phi})}{\sum_{\varphi=1}^{T} \exp(w_i^{\varphi})}, \tag{8}$$

where T denotes the number of graphs. We then obtain the final representation matrix $Y \in \mathbb{R}^{(nd+nm)\times l}$ for each node by aggregating the graph-specific representations as follows:

$$Y = \begin{bmatrix} Y_d \\ Y_m \end{bmatrix} = \sum_{\Phi=1}^{T} \beta^{\Phi} \cdot Z^{\Phi}. \tag{9}$$

4.3 Decoder for microbe-drug association reconstruction

We attain the learned feature matrices $Y_m \in \mathbb{R}^{nm\times l}$ for microbes and $Y_d \in \mathbb{R}^{nd\times l}$ for drugs in Equation 9. Inspired by inductive matrix completion (Jain and Dhillon, 2013), we reconstruct an adjacent matrix for microbe-drug associations in Equation and define the loss function in Equation 11:

$$A^{'} = Y_d W_d (Y_m W_m)^T, \tag{10}$$

$$\mathcal{L}_{REC} = \sum_{(i,j)\in P\cup N} \Theta(A_{ij}^{'}, A_{ij}), \tag{11}$$

where $W_d \in \mathbb{R}^{nd\times r}$ and $W_m \in \mathbb{R}^{nm\times r}$ are trainable latent factors that are used to project learned embeddings back to original feature space for drugs and microbes. In addition, Θ is the MSE loss (i.e., mean square error), and P and N denote the sets of positive samples and negative samples, respectively.

4.4 Overall loss and optimization

Our EGATMDA model has a few parameters, such as W_d, W_m, W_c and B. To limit their impact on the model, we add a regularization term denoted as $\mathcal{L}_{\Omega}$ in Equation 12. Therefore, the overall loss function $\mathcal{L}_{Total}$ is defined in Equation 13.

$$\mathcal{L}_{\Omega} = \|W_d\|^2 + \|W_m\|^2 + \|W_c\|^2 + \|B\|^2, \tag{12}$$

$$\mathcal{L}_{Total} = \mathcal{L}_{REC} + \gamma\mathcal{L}_{\Omega}, \tag{13}$$

where γ represents a weight factor. In this work, we deploy the Adam optimizer (Kingma and Ba, 2019) for the optimization. Finally, we use the scores in the reconstructed matrix $A^{'}$ to prioritize the unknown pairs for microbe-drug association prediction.

Table 2. The AUC and AUPR obtained under CVS1, CVS2 and CVS3 settings in 5-fold CV. The best results are marked in bold and the second best is underlined.

Methods	CVS1		CVS2		CVS3	
	AUC	AUPR	AUC	AUPR	AUC	AUPR
HMDAKATZ	0.9365± 0.0073	0.9305± 0.0064	0.9146± 0.0246	0.9319± 0.0142	0.5376± 0.0448	0.5687± 0.0598
IMCMDA	0.7334± 0.0185	0.8038± 0.0215	0.6933± 0.0216	0.7692± 0.0321	0.5281± 0.0321	0.5272± 0.0412
NTSHMDA	0.8993± 0.0137	0.8965± 0.0149	0.9259± 0.0149	0.9347± 0.0085	0.5732± 0.0296	0.6533± 0.0299
GCMDR	0.8938± 0.0137	0.8956± 0.0142	0.8665± 0.0134	0.8486± 0.0152	0.5234± 0.0312	0.5032± 0.0123
NetLapRLS	0.9372±0.0078	0.9381±0.0085	0.9263±0.0125	0.9467±0.0086	0.5483±0.0554	0.5622±0.0569
BLM-NII	0.9136±0.0484	0.9394±0.0299	0.9488±0.0090	**0.9697±0.0056**	0.6459±0.0541	0.6789±0.0637
WNN-GIP	0.7799±0.0677	0.8587±0.0456	0.9356±0.0170	0.9445±0.0178	0.7503±0.0159	0.7536±0.0163
EGATMDA	**0.9586±0.0083**	**0.9460±0.0112**	**0.9562±0.0088**	0.9386±0.0179	**0.8232±0.0671**	**0.7655±0.0534**

5 Experimental Results

Here extensive experiments have been carried out to evaluate the performance of our proposed EGATMDA model on MDAD database. Next, we first briefly introduce the experimental setup and then demonstrate the performance of our model by comparing it with seven state-of-the-art methods, under three different cross-validation settings.

5.1 Experimental setup

In this work, we conducted standard 5-fold cross-validation (CV) under the following three different settings:

- CVS1 (overall testing): CV on microbe-drug pairs—random known entries in A (i.e., microbe-drug pairs) are selected for testing.
- CVS2 (horizontal testing for drugs): CV on drugs—random rows in A (i.e., drugs) are blinded for testing.
- CVS3 (vertical testing for microbes): CV on microbes—random columns in A (i.e., microbes) are blinded for testing.

For CVS1, we randomly divide known microbe-drug associations pairs into five groups. For each round, one group of microbe-drug associations (i.e., positive samples) with an equal-size set of unknown randomly sampled pairs (i.e., negative samples) are treated as test samples in turn. And the remaining four groups microbe-drug pairs together with the same number of unknown pairs are used for training. Similarly for CVS2 and CVS3, we randomly select 20% rows and columns as test data respectively. Then, the performance is evaluated by two well-known metrics that are extensively utilized for link prediction, namely, area under ROC curve (AUC) and area under precision-recall curve (AUPR). For a fair comparison, each experiment is conducted for 10 times, and the final AUC and AUPR scores are calculated by the average over the 10 repetitions. Note that the CV settings CVS2 and CVS3 are designed to evaluate the capability of a method to identify the microbe-drug associations for new drugs and new microbes respectively.

5.2 Comparison with state-of-the-art methods

As microbe-drug association prediction is a new problem, few computational approaches have been presented for this important task. We compare our method with seven state-of-the-art methods that were proposed for different link prediction problems in the field of computational biology.

- HMDAKATZ (Zhu *et al.*, 2019) is *KATZ measure based computational method*, developed for microbe-drug prediction.
- NTSHMDA (Luo and Long, 2018) is a *random walk with restart based model*, proposed to predict microbe-disease associations.
- IMCMDA (Chen *et al.*, 2018) is a *matrix completion based model* for microRNA-disease association prediction.
- GCMDR (Huang *et al.*, 2019) is a *graph convolution network based model* for identifying miRNA-drug resistance relationships.
- NetLapRLS (Xia *et al.*, 2010) is a *Laplacian regularized least squares (LapRLS) based method* for drug-target interaction prediction.
- BLM-NII (Mei *et al.*, 2012) is a *bipartite local model with Neighbor-based Interaction profile Inferring* for drug-target interaction prediction.
- WNN-GIP (Van Laarhoven and Marchiori, 2013) is a *weighted nearest neighbor-Gaussian interaction profile model*, developed for drug-target interaction prediction.

For a fair comparison, we ran seven state-of-the-art methods on MDAD dataset with their default parameters. For CVS1, our EGATMDA model achieves the best performance in terms of both AUC and AUPR as shown in Table 2, indicating it is effective for identifying novel microbe-drug associations. For CVS2, our method achieves the best average AUC score, while it achieves a lower AUPR than NetLapRLS and BLM-NII. Note that CVS3 simulates the microbe-drug association prediction for new microbes. Under this scenario, our EGATMDA model attains the best AUC value of 0.8232 and AUPR value of 0.7655, which are 9.72% and 1.58% better than the second best method WNN-GIP. While all the above results are based on 5-fold CV, we also report the performance of various methods using 2-fold CV and 10-fold CV in Supplementary Table S1. Overall, our method outperforms other methods for microbe-drug association prediction under different scenarios.

We can observe that the performance of various methods under the CVS2 setting is significantly better than that under CVS3 as shown in Table 2. As the number of microbes (173) is much smaller than that of drugs (1373), the drug similarity matrix (1373×1373) is thus more informative than the microbe similarity matrix (173×173). Hence, the information propagated from neighbors to new drugs is expected to be more abundant and accurate than new microbes. In addition, the performance of various methods under CVS1 is generally superior to that under CVS2 and CVS3 except BLM-NII and WNN-GIP. For new drugs and new microbes, we have no microbe-drug associations in training data for them, which results in lower performance under CSV2 and CSV3.

5.3 The influence of different data sources

Recall that we construct three different genres of microbe-drug networks as shown in Table 1, as inputs to collectively learn node representations.

Here, we conduct an ablation study to evaluate the impact of each network for microbe-drug association prediction. Specifically, we evaluate the performance of the model in 5-fold CV by leveraging diverse combination of three networks as inputs. As shown in Table 3, we can uncover that the best performance reaches when three networks are simultaneously fed to the model, indicating that all of three different sources of existing biomedical data are useful and can boost the prediction performance.

The original microbe-drug bipartite network Net_1 plays the most important role, as it achieves higher AUC and AUPR values than two other networks. In addition, we can conclude that Net_2 constructed from the microbe-drug heterogeneous network contributes more than Net_3 that is derived from the microbe-disease-drug heterogeneous network. The main reason is that Net_3 is extremely sparse with a limited number of microbes and drugs.

Table 3. Performance comparison among different network combinations under the setting CVS1.

Networks	AUC	AUPR
$Global\ Net$	0.8943±0.0114	0.8835±0.0153
Net_1	0.9527±0.0054	0.9189±0.0174
Net_2	0.9126±0.0140	0.9075±0.0196
Net_3	0.8677±0.0142	0.8473±0.0157
$Net_1 + Net_2$	0.9551±0.0054	0.9300±0.0145
$Net_1 + Net_3$	0.9542±0.0112	0.9170±0.0126
$Net_2 + Net_3$	0.9139±0.0127	0.8942±0.0197
$Net_1 + Net_2 + Net_3$	**0.9586±0.0083**	**0.9460±0.0112**

In particular, $Global\ Net$ represents the global network that is constructed by integrating Net_1, Net_2 with Net_3. As $Global\ Net$ is a single network, we can only run the node-level attention to learn node representations. As shown in Table 3, $Global\ Net$ achieves much lower performance than $Net_1 + Net_2 + Net_3$, indicating that our ensemble framework with graph-level attention indeed boosts the prediction performance. The results of the ablation study under CVS2 and CVS3 can be found in Supplementary Table S2, from which we can draw similar conclusions.

5.4 Analysis of hierarchical attention mechanism

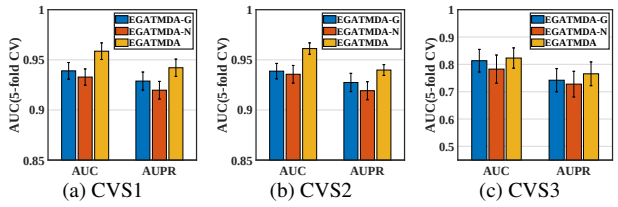
Our EGATMDA model consists of dual attention, including node-level attention and graph-level attention. The goal of these two attention is to learn the importance of graph-specific neighbors and graphs, respectively. Here, we conduct an ablation study to evaluate their impact on the performance. In particular, we derive the following model variants for ablation study:

- **EGATMDA-G:** it uses graph-level attention only, i.e., it uses a random matrix instead of the node-level attention matrix in Equation 5.
- **EGATMDA-N:** it uses node-level attention only, i.e., it uses equal weight instead of bias weight in Equation 8 for graph-level attention.

Fig.2 shows that both EGATMDA-G and EGATMDA-N achieve consistently worse performance than EGATMDA under three CV settings, indicating that both node-level and graph-level attention are effective in capturing different semantic information of nodes in different networks. In addition, we can observe that graph-attention plays a more crucial role than node-attention, as EGATMDA-N achieves lower performance than EGATMDA-G.

5.5 Parameter sensitivity analysis

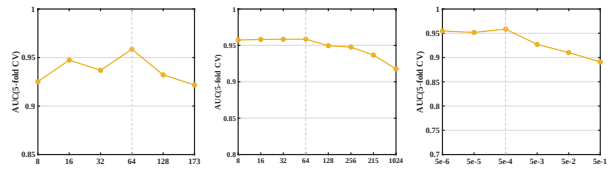
Several important parameters influence the model performance, such as the original feature dimension of microbes k, the size of latent factor l in GCN and weight factor γ. It should be noted that we perform the parameter sensitivity analysis using 5-fold CV for all parameters. Fig.3 shows the AUC results under CVS1.



(a) CVS1 (b) CVS2 (c) CVS3

Fig. 2. Comparative analysis between EGATMDA and its variants.

k determines the original feature information for microbes to be fed to the model. We select its value from $\{8, 16, 32, 64, 128, 173\}$ to evaluate its impact. Fig.3(a) indicates that a large or small value of k is not good for the model performance. The best performance is achieved when k is set as 64. To determine the influence of latent factor dimension l, we evaluate the performance of the model by varying l in the range of $\{8, 16, 32, 64, 128, 256, 512, 1024\}$. As shown in Fig.3(b), the performance first slightly increases and then decreases with l being increased. In particular, the best performance is achieved when l is set as 64. Lastly, the weight factor γ in our model is used to control the contribution of the regularization term in Equation 12 (i.e. the regularization for the weight matrices in the encoder and decoder). In our experiment, we vary γ from 0.000005 to 0.5 with a step value of 10. From Fig.3(c), we can observe that the best performance is achieved when γ is around 0.0005 and the performance decreases if we further increase the value of γ. In addition, Fig.S2 and Fig.S3 in our supplementary materials show the results under CVS2 and CVS3, respectively.



(a) k (b) l (c) γ

Fig. 3. Parameter sensitivity under CVS1 for (a) k, (b) l and (c) γ.

5.6 Case study

To further confirm the effectiveness of EGATMDA, we apply our model on two popular drugs, i.e., *Ciprofloxacin* and *Moxifloxaxin*, and two microbes, i.e., *Pseudomonas aeruginosa* and *Escherichia coli*, for our case studies. For each of them, we reset all known entries as unknown to simulate the prediction for new microbes and new drugs. Then, we prioritize candidate microbes (or drugs) according to their predicted scores. We evaluate the performance of the model by verifying the top 10, 20, and 50 predicted candidate microbes (or drugs) using a literature search.

Particularly, drug *Ciprofloxacin* is a fluoroquinolone antibacterial agent (Davis *et al.*, 1996), which mainly treats Gram-negative pathogens-causing infectious diseases. An increasing number of reports have indicated that it closely interacts with an extensive range of human microbes. For example, Gollapudi *et al.* (1998) demonstrated that *Ciprofloxacin* can inhibit Human immunodeficiency virus 1 (HIV-1), which is predicted by our model to be the best possible candidate microbe for *Ciprofloxacin*. Hacioglu *et al.* (2019) confirmed that *Ciprofloxacin* can generate activity against *Candida albicans*. Kim and Woo (2017) showed that *Enterococcus faecalis* is a high-level *Ciprofloxacin*-resistant microbe. Eventually, the results indicated that 10, 18 and 45 out of top 10, 20, and 50 predicted *Ciprofloxacin*-associated microbes can be validated by

Table 4. The top 20 predicted Ciprofloxacin-associated microbes. The first column records top 10 microbes, while the third column records top 11-20 microbes.

Microbe	Evidence	Microbe	Evidence
Human immunodeficiency virus 1	PMID:9566552	Plasmodium falciparum	PMID:17214980
Candida albicans	PMID:31471074	Streptococcus pneumoniae	PMID:26100702
Staphylococcus epidermis	PMID:10632381	Enteric bacteria	PMID:27436461
Staphylococcus epidermidis	PMID:28481197	Actinomyces oris	Unconfirmed
Enterococcus faecalis	PMID:27790716	Serratia marcescens	PMID:23751969
Streptococcus mutans	PMID:30468214	Streptococcus epidermidis	Unconfirmed
Vibrio harveyi	PMID:27247095	Listeria monocytogenes	PMID:28355096
Salmonella enterica	PMID:26933017	Vibrio vulnificus	PMID:28971862
Eikenella corrodens	PMID:16875802	Burkholderia cenocepacia	PMID:27799222
Burkholderia pseudomallei	PMID:24502667	Porphyromonas gingivalis	PMID: 15231772

previously published literature. The high prediction accuracy, i.e., 100%, 90%, and 90%, indicates that EGATMDA is a very promising tool to assist the screening of candidate compounds for drug development in real-life applications. Table 4 shows the top-20 candidate microbes for *Ciprofloxacin*. Top-50 *Ciprofloxacin*-related microbes could be found in Supplementary Table S3.

On the other hand, drug *Moxifloxacin* is an extended-spectrum fluroquinolone antibacterial agent (Balfour and Wiseman, 1999), which can treat patients with community-acquired pneumonia, acute exacerbations of chronic bronchitis or acute sinusitis (Balfour and Lamb, 2000), and skin structure infections (Tulkens *et al.*, 2012). For example, Grillon *et al.* (2016) demonstrated that the inferred top candidate microbe, *Pseudomonas aeruginosa*, was highly susceptibility to *Moxifloxacin*. Greimel *et al.* (2017) indicated that *Moxifloxacin* was an effective candidate treatment compound for the infection caused by *Staphylococcus aureus*. Alharbi *et al.* (2019) found that more than 50% of *Escherichia coli* isolates obtained from wound infections were resistant to *Moxifloxacin*. As a result, 8, 17, and 38 out of top 10, 20, and 50 predicted candidate microbes related to *Moxifloxacin* are verified by existing publications, demonstrating EGATMDA has powerful capability in identifying potential target microbes for drugs and thus is extremely helpful for drug repurposing. The top 20 and 50 predicted candidate microbes for *Moxifloxacin* are displayed in Table 5 and Supplementary Table S4.

With regards to microbes, *Pseudomonas aeruginosa* is a Gram-negative bacillus that is classified as an opportunistic pathogen (Colmer-Hamood *et al.*, 2016). It causes frequent disease in patients with underlying or immunocompromising conditions. Supplementary Table S5 indicated that among the top 10, 20, and 50 predicted *Pseudomonas aeruginosa*-associated candidate drugs, 7, 12, and 25 microbe-drug interactions are confirmed by published reports, respectively. Finally, *Escherichia coli* is a bacterium commonly found in the human intestine (Tenaillon *et al.*, 2010). Most *Escherichia coli* are harmless and benefit human health, but some strains can cause diseases of the gastrointestinal and urinary (Nataro and Kaper, 1998). The prediction results show that 8, 17, and 38 out of 10, 20, and 50 *Escherichia coli*-related candidate drugs are verified from existing evidences, as shown in Supplementary Table S6.

Overall, the above case study demonstrates our model's strong capabilities to accurately predict unknown microbes for existing drugs, as well as predict unknown drugs for existing microbes.

6 Discussion and conclusion

In this work, we propose a novel end-to-end deep learning model, named EGATMDA, based on graph neural network to predict new microbe-drug associations. In order to take full advantage of different semantic information of nodes from diverse networks, we design a hierarchical dual attention mechanism, i.e., node-level and graph-level attention, which can efficiently preserve the importance of graph-specific neighbors and graphs and remove irrelevant noise. Furthermore, we combine graph convolutional network with graph attention network to learn the importance of high-order neighbors in the node-level attention. In the graph-level attention, a knowledge-aware attention mechanism is developed, which assigns greater weight values to more useful graphs for preserving the importance of graphs, leading to more accurate node presentations. Comprehensive experiments demonstrate that the proposed EGATMDA model is reliable and promising in identifying potential target microbes for drugs, including both new drugs and new microbes.

However, there are still some limitations that influence the performance of our model. Currently, our model can make predictions for new drugs and new microbes using multiply types of biological data (e.g., drug structure similarity and microbe sequence similarity information). Due to the noises in the features extracted from such similarities, our model is still far away from perfect and there is room for us to further improve our prediction results. In the future, we aim to improve and enrich the features for drugs and microbes by incorporating more biological data, such as microbe functional similarity (Kamneva, 2017) and side-effect-based drug similarity (Kuhn *et al.*, 2010).

Acknowledgements

This work has been supported by the National Natural Science Foundation of China (Grant no.61873089) and the Chinese Scholarship Council (CSC) (201906130027).

References

Alharbi, N. S. *et al.* (2019). Prevalence of escherichia coli strains resistance to antibiotics in wound infections and raw milk. *Saudi journal of biological sciences*, **26**(7), 1557–1562.

Balfour, J. A. B. and Lamb, H. M. (2000). Moxifloxacin. *Drugs*, **59**(1), 115–139.

Balfour, J. A. B. and Wiseman, L. R. (1999). Moxifloxacin. *Drugs*, **57**(3), 363–373.

Cai, R. *et al.* (2020). Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers. *Bioinformatics*.

Chen, T. *et al.* (2002). Principle component analysis and its variants for biometrics. In *Proceedings of International Conference on Image Processing*, pages 61–64.

Chen, X. *et al.* (2018). Predicting mirna–disease association based on inductive matrix completion. *Bioinformatics*, **34**(24), 4256–4265.

Colmer-Hamood, J. *et al.* (2016). In vitro analysis of pseudomonas aeruginosa virulence using conditions that mimic the environment at specific infection sites. In *Progress in molecular biology and translational science*, pages 151–191.

Davis, A. P. *et al.* (2019). The comparative toxicogenomics database: update 2019. *Nucleic acids research*, **47**(D1), D948–D954.

Davis, R. *et al.* (1996). Ciprofloxacin. *Drugs*, **51**(6), 1019–1074.

Durand, G. A. *et al.* (2019). Antibiotic discovery: History, methods and perspectives. *International journal of antimicrobial agents*, **53**(4), 371–382.

Gollapudi, S. *et al.* (1998). 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.

Table 5. The top 20 predicted Moxiflocacin-associated microbes. The first column records top 10 microbes, while the third column records top 11-20 microbes.

Microbe	Evidence	Microbe	Evidence
Pseudomonas aeruginosa	PMID:31691651	Plasmodium falciparum	PMID:15125930
Staphylococcus aureus	PMID:31689174	Bacillus subtilis	PMID:30036828
Escherichia coli	PMID:31542319	Eikenella corrodens	PMID:14614671
Staphylococcus epidermis	PMID: 11249827	Streptococcus pneumoniae	PMID:31542319
Staphylococcus epidermidis	PMID:31516359	Burkholderia pseudomallei	PMID:15731198
Human immunodeficiency virus 1	Unconfirmed	Actinomyces oris	PMID: 26538502
Streptococcus mutans	PMID:29160117	Streptococcus sanguinis	PMID:10629010
Enterococcus faecalis	PMID:31763048	Burkholderia cenocepacia	Unconfirmed
Vibrio harveyi	Unconfirmed	Listeria monocytogenes	PMID:28739228
Salmonella enterica	PMID:22151215	Serratia marcescens	PMID:17592324

Greimel, F. *et al.* (2017). Efficacy of antibiotic treatment of implant-associated staphylococcus aureus infections with moxifloxacin, flucloxacillin, rifampin, and combination therapy: an animal study. *Drug design, development and therapy*, **11**, 1729.

Grillon, A. *et al.* (2016). Comparative activity of ciprofloxacin, levofloxacin and moxifloxacin against klebsiella pneumoniae, pseudomonas aeruginosa and stenotrophomonas maltophilia assessed by minimum inhibitory concentrations and time-kill studies. *PloS one*, **11**(6).

Guthrie, L. *et al.* (2017). Human microbiome signatures of differential colorectal cancer drug metabolism. *NPJ biofilms and microbiomes*, **3**(1), 27.

Hacioglu, M. *et al.* (2019). Effects of ceragenins and conventional antimicrobials on candida albicans and staphylococcus aureus mono and multispecies biofilms. *Diagnostic microbiology and infectious disease*, **95**(3), 114863.

Haiser, H. J. *et al.* (2013). Predicting and manipulating cardiac drug inactivation by the human gut bacterium eggerthella lenta. *Science*, **341**(6143), 295–298.

Han, P. *et al.* (2019). Gcn-mf: Disease-gene association identification by graph convolutional networks and matrix factorization. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 705–713.

Huang, Y.-a. *et al.* (2019). Graph convolution for predicting associations between mirna and drug resistance. *Bioinformatics*.

Huttenhower, C. *et al.* (2012). Structure, function and diversity of the healthy human microbiome. *nature*, **486**(7402), 207.

Jain, P. and Dhillon, I. S. (2013). Provable inductive matrix completion. *arXiv preprint arXiv:1306.0626*.

Kamneva, O. K. (2017). Genome composition and phylogeny of microbes predict their co-occurrence in the environment. *PLoS computational biology*, **13**(2), e1005366.

Kashyap, P. C. *et al.* (2017). Microbiome at the frontier of personalized medicine. In *Mayo Clinic Proceedings*, pages 1855–1864.

Kau, A. L. *et al.* (2011). Human nutrition, the gut microbiome and the immune system. *Nature*, **474**(7351), 327.

Kim, M.-C. and Woo, G.-J. (2017). 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.

Kingma, D. P. and Ba, J. A. (2019). A method for stochastic optimization. arxiv 2014. *arXiv preprint arXiv:1412.6980*.

Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Kuhn, M. *et al.* (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology*, **6**(1), 343.

Lewis, K. and Strandwitz, P. (2019). Microbes make metabolic mischief by targeting drugs. *Nature*, **570**, 453–454.

Linmei, H. *et al.* (2019). Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4823–4832.

Liu, Z. *et al.* (2020). Basconv: Aggregating heterogeneous interactions for basket recommendation with graph convolutional neural network. *arXiv preprint arXiv:2001.09900*.

Luo, J. and Long, Y. (2018). Ntshmda: Prediction of human microbe-disease association based on random walk by integrating network topological similarity. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–12.

Ma, W. *et al.* (2017). An analysis of human microbe–disease associations. *Briefings in bioinformatics*, **18**(1), 85–97.

Mei, J.-P. *et al.* (2012). Drug–target interaction prediction by learning from local information and neighbors. *Bioinformatics*, **29**(2), 238–245.

Mshvildadze, M. *et al.* (2010). Intestinal microbial ecology in premature infants assessed with non–culture-based techniques. *The Journal of pediatrics*, **156**(1), 20–25.

Nataro, J. P. and Kaper, J. B. (1998). Diarrheagenic escherichia coli. *Clinical microbiology reviews*, **11**(1), 142–201.

Qiu, J. *et al.* (2018). Deepinf: Social influence prediction with deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2110–2119.

Ravindra, N. G. *et al.* (2020). Disease state prediction from single-cell data using graph attention networks. *arXiv preprint arXiv:2002.07128*.

Schwabe, R. F. and Jobin, C. (2013). The microbiome and cancer. *Nature Reviews Cancer*, **13**(11), 800.

Sommer, F. and Bäckhed, F. (2013). The gut microbiota—masters of host development and physiology. *Nature Reviews Microbiology*, **11**(4), 227.

Sun, Y.-Z. *et al.* (2018). Mdad: a special resource for microbe-drug associations. *Frontiers in cellular and infection microbiology*, **8**.

Tenaillon, O. *et al.* (2010). The population genetics of commensal escherichia coli. *Nature Reviews Microbiology*, **8**(3), 207–217.

Tulkens, P. M. *et al.* (2012). Moxifloxacin safety. *Drugs in R&D*, **12**(2), 71–100.

Van Laarhoven, T. and Marchiori, E. (2013). Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PloS one*, **8**(6), e66952.

Veličković, P. *et al.* (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Ventura, M. *et al.* (2009). Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nature Reviews Microbiology*, **7**(1), 61.

Wang, X. *et al.* (2019). Heterogeneous graph attention network. In *The World Wide Web Conference*, pages 2022–2032.

Wen, L. *et al.* (2008). Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature*, **455**(7216), 1109.

Wishart, D. S. *et al.* (2018). Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research*, **46**(D1), D1074–D1082.

Wu, Q. *et al.* (2019). Dual graph attention networks for deep latent representation of multifaceted social effects in recommender systems. In *The World Wide Web Conference*, pages 2091–2102.

Xia, Z. *et al.* (2010). Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC systems biology*, **4**(2), S6.

Yao, L. *et al.* (2019). Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7370–7377.

Zhang, H. *et al.* (2009). Human gut microbiota in obesity and after gastric bypass. *Proceedings of the National Academy of Sciences*, **106**(7), 2365–2370.

Zhang, Y. *et al.* (2018). Graph convolution over pruned dependency trees improves relation extraction. *arXiv preprint arXiv:1809.10185*.

Zhao, J. *et al.* (2019). Intentgc: a scalable graph convolution framework fusing heterogeneous information for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2347–2357.

Zhu, L. *et al.* (2019). Prediction of microbe-drug associations based on katz measure. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 183–187.

Zimmermann, M. *et al.* (2019). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*, **570**(7762), 462–467.

Zitnik, M. *et al.* (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, **34**(13), i457–i466.