Multi-view Contrastive Learning Hypergraph Neural Network for Drug-Microbe-Disease Association Prediction

Luotao Liu[†], **Feng Huang**[†], **Xuan Liu**, **Zhankun Xiong**, **Menglu Li**, **Congzhi Song** and **Wen Zhang**[*]

College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
{luotaoliu, fhuang233, lx666, xiongzk, mengluli, song_cz }@webmail.hzau.edu.cn, zhangwen@mail.hzau.edu.cn

Abstract

Identifying the potential associations among drugs, microbes, and diseases is of great significance in exploring the pathogenesis and improving precision medicine. There are plenty of computational methods for pair-wise association prediction, such as drug-microbe and microbe-disease associations, but few methods focus on the higher-order triple-wise drug-microbe-disease (DMD) associations. Driven by the advancement of hypergraph neural networks (HGNNs), we expect them to fully capture high-order interaction patterns behind the hypergraph formulated by DMD associations and realize sound prediction performance. However, the confirmed DMD associations are insufficient due to the high cost of *in vitro* screening, which forms a sparse DMD hypergraph and thus brings in suboptimal generalization ability. To mitigate the limitation, we propose a **M**ulti-view **C**ontrastive Learning **H**ypergraph **N**eural **N**etwork, named MCHNN, for DMD association prediction. We design a novel multi-view contrastive learning (CL) on the DMD hypergraph as an auxiliary task, which guides the HGNN to learn more discriminative representations and enhances the generalization ability. Extensive experiments show that MCHNN achieves satisfactory performance in DMD association prediction and, more importantly, demonstrate the effectiveness of our devised the multi-view CL on the sparse DMD hypergraph.

1 Introduction

Microbes are instrumental in modulating drug efficacy and toxicity during disease treatment, and drugs can also affect the diversity and function of microbial communities [Long *et al.*, 2020a]. Exploring potential associations among drugs, microbes, and diseases can help to understand the underlying disease mechanisms and facilitate personalized treatments. Currently, a large number of methods have been proposed to predict drug-microbe [Long *et al.*, 2020b], microbe-disease [Wang *et al.*, 2021], and drug-disease associations [Liu *et al.*, 2021]. Despite the apparent connections between these pair-wise association prediction tasks, most existing methods deal with them separately and fail to provide in-depth insights into intricate drug-microbe-disease (DMD) interaction patterns. Indeed, recent studies in human metabolic systems pointed out the importance of identifying triple-wise DMD associations [Wu *et al.*, 2022; Wang *et al.*, 2022a]. For that reason, it is necessary to develop effective methods for predicting DMD associations that have been few investigated before.

Predicting triple-wise associations is a fundamental issue in multiple domains, e.g., recommendation systems [Wang *et al.*, 2022b; Cheng *et al.*, 2022], knowledge graphs [Qian *et al.*, 2018; Guan *et al.*, 2018] and bioinformatics [Sidorov *et al.*, 2019; Chen and Li, 2019; Chen and Li, 2020; Liu *et al.*, 2022]. For example, [Chen and Li, 2019] developed a tensor decomposition model to predict which target a drug binds to when administered to a disease. [Balažević *et al.*, 2019] constructed a neural tensor factorization for knowledge graph completion. [Wang *et al.*, 2022b] built a Spatio-temporal convolutional attention network for personalized point-of-interest recommendation. Recently, hypergraph neural network (HGNN)-based methods have increasingly been proposed for triple-wise association prediction tasks, where HGNNs have shown powerful abilities for modeling high-order relations and learning expressive representations. For example, [Cheng *et al.*, 2022] devised an interactive hypergraph neural network for personalized product search. [Liu *et al.*, 2022] proposed a multi-way relation-enhanced hypergraph representation learning method to predict anti-cancer drug synergy. Driven by the benefits of HGNNs, we reason that using them to handle the hypergraph formulated by DMD associations is expected to capture the complex interaction patterns behind the data and provide a desirable prediction performance. However, the confirmed DMD associations are much insufficient due to the prohibitive cost of *in vitro* screening. This inevitably results in a sparse DMD hypergraph which limits the expressive power of HGNNs and further hinders the generalization ability for prediction.

To alleviate the above limitations, we propose a **M**ulti-view **C**ontrastive Learning **H**ypergraph **N**eural **N**etwork, named MCHNN, for DMD association prediction. We construct a DMD hypergraph where node attributes are composed of drug features learned from molecular graphs by a GIN [Xu *et al.*, 2019], microbe features extracted from pair-wise microbe similarities and disease features encoded from pair-wise disease similarities. Then, we simply deploy a hyper-

graph convolutional network (HGCN) [Bai *et al.*, 2021] over the DMD hypergraph, which absorbs these biological features, captures the high-order relations of DMD associations and generates high-level node representations. Moreover, to mitigate the sparsity issue of the DMD hypergraph, we bring in contrastive learning (CL) acting as an auxiliary task to enhance the learned node representations. CL on graph data is a promising technique for graph representation learning that enriches supervision signals by exploiting abundant pseudo-label data, but few studies deal with hypergraphs. Here, we fulfill a multi-view CL for the DMD hypergraph in a global-local mutual information maximization paradigm [Velickovic *et al.*, 2019], where four task-specific augmentation schemes are devised to form multiple views of hypergraph counterparts. To our knowledge, this is the first work aiming at triple-wise DMD association prediction. In summary, the main contributions of this paper are described as follows:

- We leverage the merits of hypergraphs modeling high-order relations and the powerful abilities of HGNNs in representation learning to develop a hypergraph-based framework for the DMD association prediction.
- We design a multi-view CL with four task-specific hypergraph augmentation schemes, which permits more expressive and discriminative representation learning on the sparse DMD hypergraph.
- We conduct extensive experiments across four testing scenarios. The results show that MCHNN achieves competitive performance in DMD association prediction and our designed multi-view CL is an effective remedy for the sparsity problem of the DMD hypergraph.

2 Related Work

2.1 Hypergraph Neural Networks

Hypergraph neural networks (HGNNs) have received tremendous attention caused of their ability to capture high-order interaction patterns in hypergraph data. HGNN [Feng *et al.*, 2019] and HyperGCN [Yadati *et al.*, 2019] extends convolution operation to hypergraphs from a spectral perspective. On that basis, [Jiang *et al.*, 2019] extended HyperGCN to a dynamic hypergraph, [Zhang *et al.*, 2020] devised a self-attention based HGNN, and [Yi and Park, 2020] developed a hypergraph convolutional recurrent neural network. Moreover, [Bai *et al.*, 2021] introduced two differentiable operators to the family of HGNNs: hypergraph convolution and attention, achieving state-of-the-art results in hypergraph representation learning. Recently, these powerful HGNNs have been widely applied to representation learning in numerous domains, including recommendation systems [Ji *et al.*, 2020; Li *et al.*, 2022; Xia *et al.*, 2022b], social networks [Gao *et al.*, 2022] and bioinformatics [Liu *et al.*, 2022]. Inspired by these works, we expect HGNNs to learn informative representations from the DMD hypergraph and to produce sound prediction performance.

2.2 Contrastive Learning on Graphs

Contrastive learning (CL) has become currently state-of-the-art in unsupervised graph representation learning, which allows models to learn structural semantics behind graph data without any explicit supervision. A line of studies related to this work has successfully introduced the mutual information maximization principle [Hjelm *et al.*, 2018] into graph learning, and what they essentially do is to pull positive pairs close to each other and keep negative pairs irrelevant to each other [Yang *et al.*, 2022]. As a result, different schemes (usually graph augmentation approaches) to yield the positive and negative pairs lead to diverse semantic biases that govern model performance. Deep Graph Infomax (DGI) [Velickovic *et al.*, 2019] fabricates a corrupted graph counterpart by shuffling node features or perturbing edges, which leads to 'fake' node representations contrasting against true representations learned from the original graph. [Hassani and Khasahmadi, 2020] extended DGI to multi-view learning by leveraging graph diffusion to generate a positive counterpart of the original graph. Although CL on hypergraphs has not been systematically discussed [Wei *et al.*, 2022], some domain-specific hypergraph contrastive strategies are designed in recommendation systems [Xia *et al.*, 2022a; Yu *et al.*, 2021]. Different from them, we design a novel multi-view CL on our constructed DMD hypergraph in a DGI paradigm, which acts as an auxiliary objective to boost our model performance.

3 Methodology

3.1 Problem Formulation

Given a drug set $\mathcal{D}$, a microbe set $\mathcal{M}$ and a disease set $\mathcal{N}$, their Cartesian product $\mathcal{S} = \mathcal{D} \times \mathcal{M} \times \mathcal{N}$ is a set of all possible DMD triplets. For each triplet $(d, m, n) \in \mathcal{S}$, we assign it a label $p \in \{0, 1\}$ where $p = 1$ if the association between the triplet is confirmed, otherwise $p = 0$. Note that $p = 0$ does not definitely mean that there is no relationship between the triplet, whereas it is an unknown association and also may be a potential association that was undiscovered before. Our goal is to learn a model that predicts potential associations from those unknown ones.

3.2 Model Architecture

MCHNN is an end-to-end deep learning model consisting of four phases (Figure 1): 1) DMD Hypergraph construction; 2) Hypergraph representation Learning; 3) Multi-View Contrastive Learning; 4) Model training. In the following, we review each phase and provide the necessary details on how we operated.

DMD Hypergraph Construction

The complex and high-order DMD associations can be modeled as a hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, in which drugs/microbes/diseases are represented as nodes $\mathcal{V} = \mathcal{D} \cup \mathcal{M} \cup \mathcal{N}$ and known DMD associations are represented as hyperedges $\mathcal{E}$. Note that these hyperedges correspond to the triplets labeled as 1, i.e. $\mathcal{E} \subset \mathcal{S}$. Technically, $\mathcal{G}$ is further formulated as an attributed hypergraph with an incidence matrix $\boldsymbol{Y} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$ and node attributes $\boldsymbol{X} \in \mathbb{R}^{|\mathcal{V}| \times F}$.

Incidence matrix. The hyperedges are stored in an incidence matrix $\boldsymbol{Y}$, with entries defined as:

$$\boldsymbol{Y}_{ve} = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \tag{1}$$

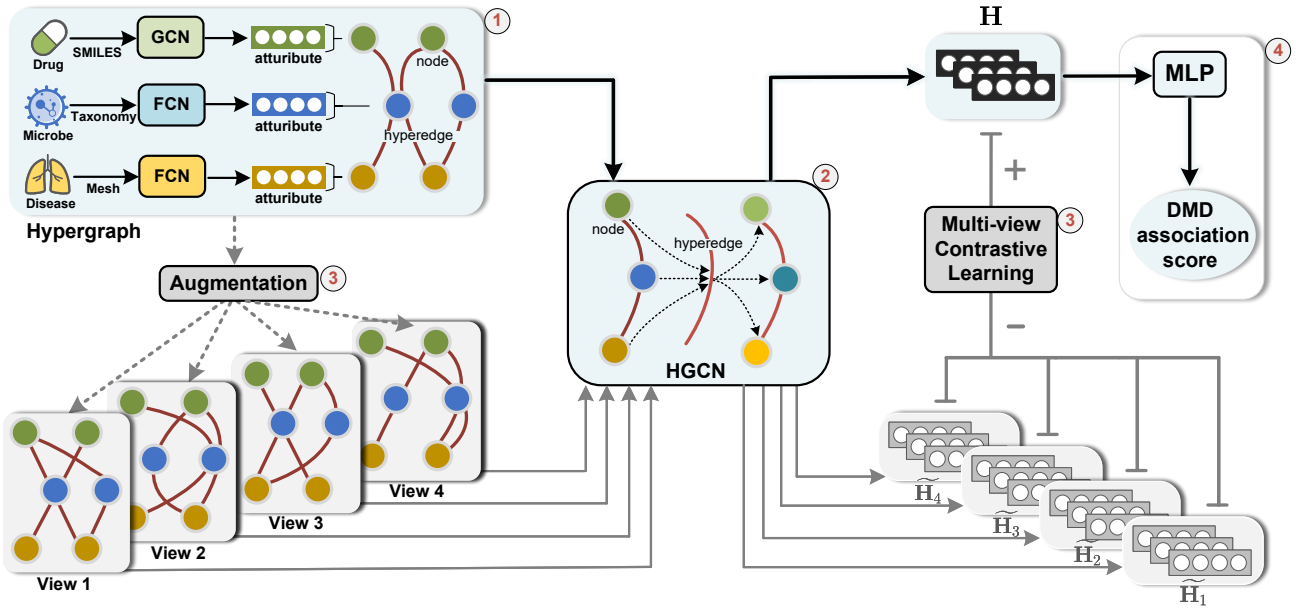

Figure 1: Workflow of MCHNN: ① DMD Hypergraph construction, ② Hypergraph representation Learning, ③ Multi-View Contrastive Learning, ④ Model training.

Node attributes. The node attributes $\boldsymbol{X}$ consists of drug features $\boldsymbol{X}_{\mathcal{D}}$, microbe features $\boldsymbol{X}_{\mathcal{M}}$ and disease features $\boldsymbol{X}_{\mathcal{N}}$. In our specific scenario, we take full advantage of the domain knowledge of bio-entities to initialize the node attributes. For each drug, its SMILES string can be converted into a molecular graph $\boldsymbol{G} = (\boldsymbol{Z}, \boldsymbol{A})$ via the DeepChem package, where $\boldsymbol{Z}$ is the attribute matrix of all nodes representing the atoms and $\boldsymbol{A}$ is the adjacency matrix encoding the bonds. We simply adopt a graph isomorphism network (GIN) [Xu *et al.*, 2019] on the molecular graph to learn atom representations which then are summarized into a drug feature vector through a global max pooling (GMP). Concretely, the k-th layer of the GIN encoder is defined as:

$$\boldsymbol{Z}^{(k)} = \mathrm{MLP}^{(k)}\left((\boldsymbol{A} + (1+\epsilon)\boldsymbol{I})\,\boldsymbol{Z}^{(k-1)}\right) \tag{2}$$

where MLP is a multi-layer perceptron, $\boldsymbol{I}$ is the identity matrix, ϵ is a fixed scalar, and $\boldsymbol{Z}^{(0)} = \boldsymbol{Z}$. Afterward, we apply the GMP over all molecular graphs, we obtain all drug features that can be compiled into $\boldsymbol{X}_{\mathcal{D}} \in^{|\mathcal{D}|\times F}$. For microbe nodes and disease nodes, we compile similarity matrices $\boldsymbol{S}_{\mathcal{M}} \in \{1,0\}^{|\mathcal{M}|\times|\mathcal{M}|}$ and $\boldsymbol{S}_{\mathcal{N}} \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ based on the methods provided by [Ma and Jiang, 2021; Wang *et al.*, 2007], which then are transformed into $\boldsymbol{X}_{\mathcal{M}} \in \mathbb{R}^{|\mathcal{M}|\times F}$ and $\boldsymbol{X}_{\mathcal{N}} \in \mathbb{R}^{|\mathcal{N}|\times F}$ by fully-connected networks.

Hypergraph Representation Learning

To encode high-order interaction information behind DMD associations and absorb biological domain knowledge stored in the node attributes, we adopt a simple yet effective hypergraph convolutional network (HGCN) [Bai *et al.*, 2021] deployed on our constructed DMD hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The HGCN defines a message-passing rule over hypergraphs, where node embeddings are updated by aggregating the information from the nodes linked through the hyperedges. Specifically, one step of hypergraph convolution is formulated as:

$$\boldsymbol{H}^{(l)} = \sigma\left(\boldsymbol{D}^{-1}\boldsymbol{Y}\boldsymbol{W}\boldsymbol{B}^{-1}\boldsymbol{Y}^{\top}\boldsymbol{H}^{(l-1)}\boldsymbol{\Theta}^{(l-1)}\right) \tag{3}$$

where $\sigma(\cdot)$ indicate a nonlinear activation function (ReLU), $\boldsymbol{\Theta}$ is a learnable weight matrix, $\boldsymbol{H}^{(l)}$ is the node embeddings at the l-th layer and $\boldsymbol{H}^{(0)}$ is initialized with $\boldsymbol{X}$; $\boldsymbol{D}$ and $\boldsymbol{B}$ are diagonal matrices respectively corresponding to the sums of rows and columns in $\boldsymbol{Y}$ called degrees of nodes and hyperedges; $\boldsymbol{W}$ is also a diagonal matrix that stores hyperedge weights and it is set equal to the identity matrix with appropriate dimensions because we assumed that the contribution of each hyperedge in the hypergraph is the same. Finally, we denote the learned node representations as $\boldsymbol{H}$ and given a node $v \in \mathcal{V}$, its representation $\boldsymbol{h}_v$ is available from $\boldsymbol{H}$.

Multi-View Contrastive Learning

The used HGCN may only learn limited knowledge from the sparse hypergraph induced by inadequate DMD associations, which results in insufficiently powerful node representations, especially for those nodes with few degrees. To mitigate this limitation, we introduce CL with a DGI style, as it permits more expressivity by maximizing the mutual information between node-level and graph-level representations. To implement a DGI-style CL on our DMD hypergraph, the first we should do is to corrupt the DMD hypergraph to generate negative-sampled hypergraphs. Different from DGI working on the underlying graphs where a single-view negative graph produced by randomly perturbing edges is useful enough, we focus on the DMD hypergraph where

different hyperedge perturbation schemes would bring about extremely diverse connectivity biases because the hyperedges represent higher-order associations among multiple types of nodes (bioentities) in the DMD hypergraph. These biases can pronouncedly govern model training and lead to unsteerable prediction. Thus, we consider multi-view negative hypergraphs corresponding to different perturbation schemes. The multi-view setting is speculated to be helpful for more rational and robust CL over the DMD hypergraph on the basis of two-fold well-known facts: 1) compared with the single-view setting, the multi-view setting involves more negative examples (pairs) in the DGI-style CL, while adequate supervisory signals are instrumental for modeling the complexity of the DMD hypergraph; 2) synthesizing multi-view information can reduce the single-view biases resulting from different perturbation schemes. In what follows, we detail our multi-view CL under several hyperedge perturbation schemes.

For each hyperedge (or triple-wise association) (d, m, n) in the DMD hypergraph, we can corrupt it to obtain one or more 'fake' triplets, and simultaneously we ensure the selected 'fake' triplets are not in the original hyperedge set $\mathcal{E}$. After that, we can maintain 'fake' hyperedge sets $\{\tilde{\mathcal{E}}_z\}_{z=1}^{Z}$ and construct corresponding negative hypergraphs $\{(\boldsymbol{X}, \tilde{\boldsymbol{Y}}_z)\}_{i=z}^{Z}$, where Z is the number of views we consider. To be more specific, we adopt four types of perturbation schemes to generate four views of negative hypergraphs: **View 1** (drug-mode perturbation): we break the joint between drug d and the microbe-disease pair (m, n) and then link (m, n) to another drug $d' \in \mathcal{D}$ to form the 'fake' hyperedge (d', m, n); **View 2** (microbe-mode perturbation): likewise, we draw the 'fake' hyperedge (d, m', n) by linking the drug-disease pair (d, n) to another microbe node $m' \in \mathcal{M}$; **View 3** (disease-mode perturbation): we link the drug-microbe pair (d, m) to another disease node $n' \in \mathcal{N}$ and produce the 'fake' hyperedge (d, m, n'); **View 4** (random perturbation): we break the chain of the entities d, m and n and chain other ones randomly selected from $\mathcal{D}$, $\mathcal{M}$ and $\mathcal{N}$ to generate the 'fake' hyperedge (d', m', n'). Then, the augmented multi-view node embeddings $\{\tilde{\boldsymbol{H}}_z\}_{z=1}^{Z}$ are further computed from these perturbed hypergraphs $\{(\boldsymbol{X}, \tilde{\boldsymbol{Y}}_z)\}_{z=1}^{Z}$ through the same encoder as HGCN deployed on the original hypergraph. $\tilde{\boldsymbol{h}}_v^z$ is the z-th counterpart of the node representation $\boldsymbol{h}_v$.

Similar to DGI, $(\boldsymbol{h}_v, \boldsymbol{s})$ is deemed as a positive example and $(\tilde{\boldsymbol{h}}_v^z, \boldsymbol{s})$ is a negative example, where $\boldsymbol{s}$ is the graph-level representation obtained through a global mean pooling layer: $\boldsymbol{H} \in \mathbb{R}^{|\mathcal{V}| \times F} \rightarrow \boldsymbol{s} \in \mathbb{R}^{F}$. Then, the objective of our multi-view CL task is formulated as:

$$\mathcal{L}_c = -\frac{1}{5|\mathcal{V}|}\left(\sum_{v\in\mathcal{V}} \log \Psi(\boldsymbol{h}_v, \boldsymbol{s}) + \sum_{z=1}^{4}\sum_{v\in\mathcal{V}} \log\left(1 - \Psi(\tilde{\boldsymbol{h}}_v^z, \boldsymbol{s})\right)\right) \tag{4}$$

where $\Psi(\cdot, \cdot)$ is the contrastive discriminator constructed by a simple bilinear function $\mathrm{Sigmoid}(\boldsymbol{h}^\top \boldsymbol{W} \boldsymbol{s})$ that estimates similarities between the node-level representations and the graph-level representation.

Model Training

For the DMD association prediction, we utilize the learned embeddings of drug $\boldsymbol{h}_d$, microbe $\boldsymbol{h}_m$, and disease $\boldsymbol{h}_n$ to output the probability of their association $\hat{p}$ through a scoring function:

$$\hat{p} = \mathrm{MLP}(\boldsymbol{h}_d \parallel \boldsymbol{h}_m \parallel \boldsymbol{h}_n) \tag{5}$$

After that, the loss of the supervised prediction task can be formulated as:

$$\mathcal{L}_p = -\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}} (p_i \log \hat{p}_i + (1 - p_i)\log(1 - \hat{p}_i)) \tag{6}$$

where $\mathcal{T}$ is the training sets and p represents the true label.

The supervised prediction task jointly optimizes the model with the aforementioned CL task during the training phase. To implement the prediction task and the CL task simultaneously, we optimize the following objective function that combines Eq.(4) and Eq.(6):

$$\mathcal{L} = \alpha\mathcal{L}_p + (1 - \alpha)\mathcal{L}_c \tag{7}$$

where α is a hyperparameter for the trade-off for different loss components.

4 Experiment

4.1 Datasets

We obtain data from several public datasets. The drug-microbe associations are collected from MDAD [Sun *et al.*, 2018], aBiofilm [Rajput *et al.*, 2018], and DrugVirus [Andersen *et al.*, 2020], and the microbe-disease associations are obtained from HMDAD [Ma *et al.*, 2017], Disbiome [Janssens *et al.*, 2018], gutMDisorder [Cheng *et al.*, 2020], and Peryton [Skoufos *et al.*, 2021]. Then we merge the drug-microbe and microbe-disease associations into data schema <drug, microbe, disease>, as such, we obtain 2,763 triplets of DMD, involving 270 drugs, 58 microbes, and 167 diseases. For drugs, their SMILES strings are downloaded from PubChem. For microbes and diseases, their taxonomic information is collected from NCBI's Taxonomy and MeSH databases, respectively. Compared with the number of conceivable DMD combinations (270×58×167), the confirmed associations we obtained are conspicuously insufficient, as when we use the collected DMD associations vs all conceivable DMD associations, the proportion is 0.11%.

4.2 Baselines

We compare MCHNN with baselines as follows:

- **Random Forest (RF) and multilayer perception (MLP)** are trained with the concatenation of the node attributes as sample features.
- **Graph convolutional network (GCN)** [Welling and Kipf, 2016] is deployed on binary association graphs, i.e. drug-microbe and microbe-disease association graphs. The predictions from GCNs are then combined to predict DMD triple-wise associations.
- **CP and Tucker** [Kolda and Bader, 2009] are two standard tensor factorization models, which reconstruct the known higher-order associations via tensor completion.

- **CoSTCo** [Liu *et al.*, 2019] designs a CNN-based tensor completion model for capturing complex interactions within the higher-order tensors.
- **NeurTN** [Chen and Li, 2020] proposes a neural tensor network combining the tensor algebra and deep neural network to capture the relationships among triplets.
- **HypergraphSynergy** [Liu *et al.*, 2022] is a hypergraph representation learning model, utilizing the HGCNs to encode the triple-wise associations.

The implemented details of the above baselines are described in Supplementary materials.

4.3 Experimental Settings

Evaluation protocols. We randomly split the dataset into a 90% cross-validation (CV) set and a 10% independent test set. On the CV set, the 5-fold CV is implemented. Moreover, we conduct independent testing, in which the model is trained on the CV set and tested on the independent test set. To evaluate the model's ability to discover associations, we focus on two top-n metrics: hit ratio (hit@n) and normalized discounted cumulative gain (ndcg@n), which have been widely used in recommendation tasks [Huang *et al.*, 2018; Huang *et al.*, 2019]. For each test triplet, we pair it with a number of sampled triplets (i.e., negative samples) that are unobserved in the dataset, and rank the test triplet among these sampled triplets according to their predicted scores. The hit@n can measure whether a test triplet is retrieved within the top-n ranked list and the ndcg@n accounts for the position of the hit by assigning higher scores to hits at top ranks. Following previous studies [Chen and Li, 2020], we employ four scenarios to produce negative samples. **Scenario 1** (drug-level negative sampling): For the triplet (d, m, n), we replace the drug d with a random new drug d' so that the triplet of (d', m, n) is unobserved in the dataset. **Scenario 2** (microbe-level negative sampling): we replace the triplet (d, m, n) with (d, m', n). **Scenario 3** (disease-level negative sampling): we replace the triplet (d, m, n) with (d, m, n'). **Scenario 4** (random negative sampling): we randomly sample a negative triplet (d', m', n'), which is unobserved in the dataset. In each scenario, we generate 29 negative samples for each test triplet and report the average metrics overall test triplets. These scenario settings can comprehensively evaluate the ability of the model to identify positive and negative samples under these stringent conditions.

Implementation details. The number of GIN layer k for drug node attributes and HGNN layer l for hypergraph encoding are both set to 3. In DMD association prediction, the scoring function is a 3-layer MLP with Dropout. We fix α to 0.8 as they produced the best performance in hyper-parameter optimization. Moreover, we employ Adam with a learning rate of 0.005 to optimize the model and adopt early stopping to control the training epochs based on validation loss. It is worth mentioning that in the training step, we use the negative samples generated from four negative sampling scenarios to train together (note that in each scenario, we generate 2 negative samples for each train triplet), while in the test process, the four scenarios are tested separately to evaluate the performance of the model. More detailed settings, source code, data set, and additional results of MCHNN are presented in Supplementary materials[1].

4.4 Performance Comparison

Figure 2 shows the performances of MCHNN and baselines on the CV set. According to the results, MCHNN surpasses other baselines on almost all scenarios by achieving the highest average hit@n/ndcg@n score of 0.9340/0.9280, suggesting its effectiveness in DMD association prediction. We also have the following observations: (1) Compared with RF, MLP, and NeurTN which only consider node feature information, and GCN which only considers information from pairwise associations, MCHNN significantly exceeds all these baselines in most scenarios, which indicates that the information of associations needs to be considered and only considering pair-wise associations can not extract sufficient information from known DMD associations. (2) CP, Tucker, and CoSTCo, which can model higher-order associations, achieve better performances than RF, MLP, GCN, and NeurTN, which implies that modeling high-order interactions among multiple entities can enhance the prediction for DMD associations. (3) HypergraphSynergy, which utilizes hypergraph to model the triple-wise associations, achieves a suboptimal performance in DMD association prediction, implying the advantage of utilizing hypergraph to model DMD triple-wise associations. Our model MCHNN still makes improvements over HypergraphSynergy. It may be attributed to the reason that the multi-view CL task with four task-specific augmentations can guide the HGCN to learn more expressive and discriminative representations and further improve the performances of DMD association prediction.

Further, we focus on the performances of MCHNN and baselines on the independent test (Supplementary Figure 3). As expected, MCHNN outperforms all the baselines in almost all scenarios and exceeded the two best baselines: HypergraphSynergy and CoSTco by 0.91% and 6.20% in average hit@5 values, and 1.19% and 6.37% in average ndcg@5 values, validating its high generalization ability.

To verify the effectiveness of MCHNN for learning accurate DMD representations, we conduct the t-SNE [Van der Maaten and Hinton, 2008] visualization on MCHNN and two deep learning-based models (HypergraphSynergy and NeurTN) using the embeddings learned from the independent test. As illustrated in Figure 3, MCHNN can better distinguish positive samples from negative samples than the compared methods and acquires the best silhouette score of 0.5124, which indicates that our model MCHNN can indeed learn more expressive and discriminative representations for DMD association prediction. The details and rest results are shown in Supplementary Figure 4.

4.5 Ablation Study

To investigate the necessity of each component in MCHNN, we conducted several comparisons between MCHNN and its variants on the independent test:

- **MCHNN without contrastive learning (w/o CL)** removes the multi-view CL task from MCHNN.

[1]https://github.com/Liuluotao/MCHNN

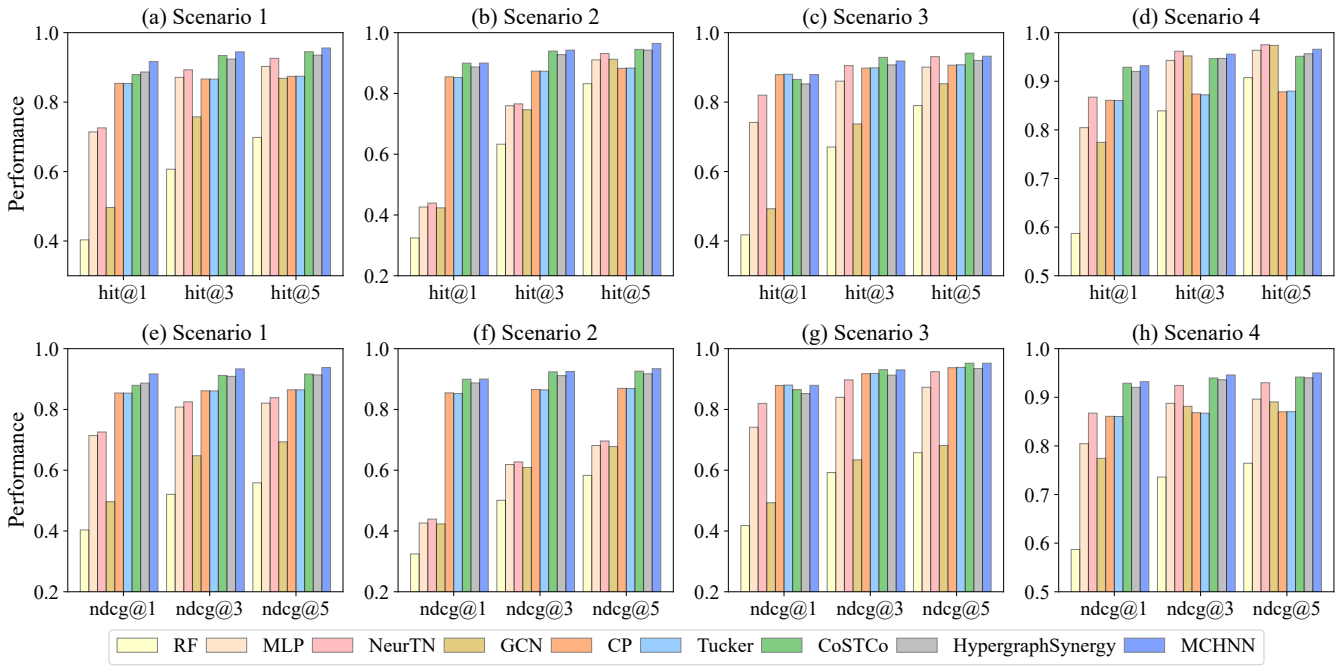

Figure 2: 5-CV performance of MCHNN and baselines in four scenarios in terms of hit@n and ndcg@n.

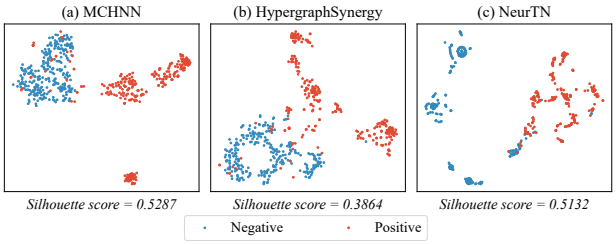

Figure 3: The t-SNE visualization of three models on scenario 4.

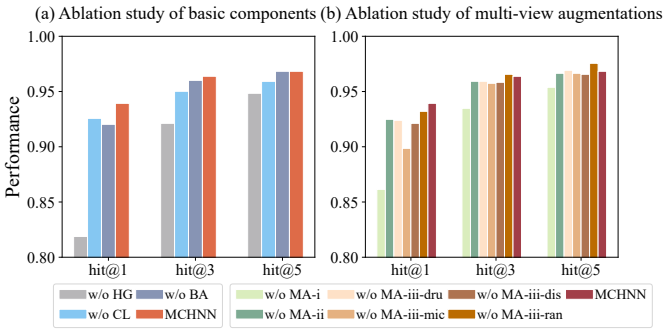

Figure 4: Average values of MCHNN and its variants on four scenarios in ablation study.

- **MCHNN without HGCN (w/o HG)** removes the hypergraph representation learning and directly uses biochemical attributes of drugs, microbes, and diseases for DMD association prediction.
- **MCHNN without biochemical attributes of nodes (w/o BA)** utilizes one-hot encoding instead of biochemical attributes of nodes in the hypergraph.
- **MCHNN without multi-view augmentations (w/o MA) (w/o MA-i)** generates four augmented hypergraphs under the ordinary scheme of View 4 (random perturbation). **(w/o MA-ii)** fuses the multi-view augmented hypergraphs into a single hypergraph in the CL task. **(w/o MA-iii-dru, mic, dis, ran)** removes the augmented hypergraph of the corresponding View.

We first investigate the importance of various basic components to our model. As we can see from Figure 4 (a), when basic components of MCHNN have been removed, the performances of corresponding variants significantly decline, indicating that these components all contribute to the prediction. Besides, we have the following observations: (1) MCHNN(w/o HG) gets the worst results, and all metrics have been significantly dropped, which demonstrates the prediction performance is boosted mostly by the HGCN and HGCN is the core component of the model architecture. (2) MCHNN(w/o CL) gets an obvious performance drop, which indicates that our designed CL can indeed learn more expressive and discriminative representations and enhance the performance of DMD association prediction. (3) MCHNN(w/o BA) also gets a performance drop on almost all metrics, which shows that the biochemical information from drugs, microbes, and diseases can provide useful domain knowledge for DMD association prediction.

We further investigate the importance of different augmented hypergraphs and multi-view augmentation strategies to our model. As we can see from Figure 4 (b), MCHNN still achieves the best performance in most scenarios, which demonstrates that our multi-view CL can effectively inte-

Methods	MCHNN			w/o CL		
	hits@1	hits@3	hits@5	hits@1	hits@3	hits@5
$[0, 50]$	0.6736	**0.8299**	**0.8785**	**0.6840**	0.7882	0.8160
$[51, 100]$	**0.9229**	**0.9840**	**0.9894**	0.9016	0.9335	0.9548
$[101, +\infty]$	0.9909	0.9955	**1.0000**	**0.9955**	**1.0000**	1.0000

Table 1: Average values of MCHNN and w/o CL on four scenarios in terms of average degrees of nodes in triplets.

grate different augmented hypergraphs to promote DMD association prediction. We also have the following observation: (1) MCHNN(w/o MA-i) achieves the worst performance, which demonstrates that the success of our model is due to the rational design of the multi-view augmentations from 4 views with different semantics. (2) The comparison of MCHNN and MCHNN(w/o MA-ii) reveals that simply and directly fusing all augmented hypergraphs is insufficient for DMD association prediction, and task-specific augmentations can learn more discriminative information from the drug-microbe-disease triple-wise associations. (3)MCHNN almost exceeds all variants of MCHNN(w/o MA-iii), especially MCHNN(w/o MA-iii-dru, mic, dis), which shows that every augmented hypergraph can enhance the prediction for DMD associations, especially the augmented hypergraph that correspondingly and independently generated for DMD triple-wise associations.

4.6 Effectiveness of Contrastive Learning on Sparse DMD Hypergraph

In this section, we aim to furtherly verify whether our designed CL is effective on sparse DMD hypergraph induced by insufficient DMD associations. Firstly, we partition the independent test set into three groups, [0,50], [51,100], and [101,+∞], according to the average node degree, which is calculated by the average of the three node degrees in each test triplet. Next, we compute the evaluation metrics (i.e., hit@n) of the test triplets for the MCHNN in the corresponding groups, and then calculate the average metrics on four scenarios. Moreover, we also compute the metrics for our model without contrastive learning (w/o CL) under the same scenarios. The results are shown in Table 1.

According to the results in Table 1, we can mainly draw the following observations: (1) The higher the average degree of the triplet, the better the prediction performance of the models, which may be owing that the triplets with low-degree will form sparse DMD hypergraph, and the models may only learn limited knowledge from it, thus results in insufficiently powerful node representations for DMD association prediction. (2) Compared with w/o CL, the performance of MCHNN is significantly increased. Specifically, in the groups of [0,50] and [51,100], MCHNN exceeds w/o CL by 6.25% and 3.46% in average hit@5 values. This indicates that the use of CL greatly improves the performance of the model in the spare DMD hypergraph induced by low-degree triplets, which further proves that CL can alleviate the issue caused by insufficient confirmed associations that occurs in our DMD association prediction scenarios.

5 Conclusion

In this paper, we propose a hypergraph-based model with multi-view CL for DMD association prediction, namely MCHNN. MCHNN constructs a hypergraph to express DMD associations in a natural way and utilizes HGCN to encode it. To tackle the trammel of HGCN's expressive power caused by the sparse hypergraph, we implement a multi-view CL on the DMD hypergraph with four task-specific augmentations schemes, guiding the HGCN to capture more expressive and discriminative representations for sparse associations. The experimental results show the superior performances of MCHNN over baselines, and the multi-view CL can indeed tackle the trammel of expressive power caused by the sparse hypergraph.

In the future, we have several directions to improve DMD association prediction, such as incorporating more diverse biological association information (e.g., drug-target interactions, gene-disease interactions) into the hypergraph, considering more diverse entity features (e.g., microbe genomic sequence feature, disease phenotypes feature), and further predicting specific types of drug-microbe associations and microbe-disease associations(e.g, increase/decrease relationships between microbes and diseases).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No.62072206, 62102158, 61772381), the Fundamental Research Funds for the Central Universities (Grant No.2662022JC004, 2662021JC008). Huazhong Agricultural University Scientific & Technological Self-innovation Foundation. The funders have no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Contribution Statement

W.Z., the corresponding author of this paper, supervised this project and advised on all parts of this paper. L.L. and F.H. took the lead in designing and writing the manuscript, which should be considered to equally contribute to this work. L.L. carried out the computational experiments with support from X.L., M.L. and C.S.. Z.X. provided help in analysis and discussion on results. All authors reviewed and approved the final version of the manuscript.

References

[Andersen *et al.*, 2020] Petter I Andersen, Aleksandr Ianevski, Hilde Lysvand, Astra Vitkauskiene, Valentyn Oksenych, Magnar Bjørås, Kaidi Telling, Irja Lutsar, Uga Dumpis, Yasuhiko Irie, et al. Discovery and development of safe-in-man broad-spectrum antiviral agents. *International Journal of Infectious Diseases*, 93:268–276, 2020.

[Bai *et al.*, 2021] Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.

[Balažević *et al.*, 2019] Ivana Balažević, Carl Allen, and Timothy M Hospedales. Tucker: Tensor factorization for knowledge graph completion. *arXiv preprint arXiv:1901.09590*, 2019.

[Chen and Li, 2019] Huiyuan Chen and Jing Li. Modeling relational drug-target-disease interactions via tensor factorization with multiple web sources. In *The World Wide Web Conference*, pages 218–227, 2019.

[Chen and Li, 2020] Huiyuan Chen and Jing Li. Learning data-driven drug-target-disease interaction via neural tensor network. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2020.

[Cheng *et al.*, 2020] Liang Cheng, Changlu Qi, He Zhuang, Tongze Fu, and Xue Zhang. gutmdisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Research*, 48(D1):D554–D560, 2020.

[Cheng *et al.*, 2022] Dian Cheng, Jiawei Chen, Wenjun Peng, Wenqin Ye, Fuyu Lv, Tao Zhuang, Xiaoyi Zeng, and Xiangnan He. Ihgnn: Interactive hypergraph neural network for personalized product search. In *Proceedings of the ACM Web Conference 2022*, pages 256–265, 2022.

[Feng *et al.*, 2019] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3558–3565, 2019.

[Gao *et al.*, 2022] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn$\hat{+}$: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[Guan *et al.*, 2018] Saiping Guan, Xiaolong Jin, Yuanzhuo Wang, and Xueqi Cheng. Shared embedding based neural networks for knowledge graph completion. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 247–256, 2018.

[Hassani and Khasahmadi, 2020] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In *International Conference on Machine Learning*, pages 4116–4126. PMLR, 2020.

[Hjelm *et al.*, 2018] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. In *International Conference on Learning Representations*, 2018.

[Huang *et al.*, 2018] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y Chang. Improving sequential recommendation with knowledge-enhanced memory networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 505–514, 2018.

[Huang *et al.*, 2019] Jin Huang, Zhaochun Ren, Wayne Xin Zhao, Gaole He, Ji-Rong Wen, and Daxiang Dong. Taxonomy-aware multi-hop reasoning networks for sequential recommendation. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 573–581, 2019.

[Janssens *et al.*, 2018] Yorick Janssens, Joachim Nielandt, Antoon Bronselaer, Nathan Debunne, Frederick Verbeke, Evelien Wynendaele, Filip Van Immerseel, Yves-Paul Vandewynckel, Guy De Tré, and Bart De Spiegeleer. Disbiome database: linking the microbiome to disease. *BMC microbiology*, 18(1):1–6, 2018.

[Ji *et al.*, 2020] Shuyi Ji, Yifan Feng, Rongrong Ji, Xibin Zhao, Wanwan Tang, and Yue Gao. Dual channel hypergraph collaborative filtering. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2020–2029, 2020.

[Jiang *et al.*, 2019] Jianwen Jiang, Yuxuan Wei, Yifan Feng, Jingxuan Cao, and Yue Gao. Dynamic hypergraph neural networks. In *IJCAI*, pages 2635–2641, 2019.

[Kolda and Bader, 2009] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[Li *et al.*, 2022] Anchen Li, Bo Yang, Huan Huo, and Farookh Hussain. Hypercomplex graph collaborative filtering. In *Proceedings of the ACM Web Conference 2022*, pages 1914–1922, 2022.

[Liu *et al.*, 2019] Hanpeng Liu, Yaguang Li, Michael Tsang, and Yan Liu. Costco: A neural tensor completion model for sparse tensors. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 324–334, 2019.

[Liu *et al.*, 2021] Ruoqi Liu, Lai Wei, and Ping Zhang. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nature Machine Intelligence*, 3(1):68–75, 2021.

[Liu *et al.*, 2022] Xuan Liu, Congzhi Song, Shichao Liu, Menglu Li, Xionghui Zhou, and Wen Zhang. Multi-way relation-enhanced hypergraph representation learning for anti-cancer drug synergy prediction. *Bioinformatics*, 38(20):4782–4789, 2022.

[Long *et al.*, 2020a] Yahui Long, Min Wu, Yong Liu, Chee Keong Kwoh, Jiawei Luo, and Xiaoli Li. Ensembling graph attention networks for human microbe–drug association prediction. *Bioinformatics*, 36(Supplement_2):i779–i786, 2020.

[Long *et al.*, 2020b] Yahui Long, Min Wu, Yong Liu, Chee Keong Kwoh, Jiawei Luo, and Xiaoli Li. Ensembling graph attention networks for human microbe–drug association prediction. *Bioinformatics*, 36(Supplement_2):i779–i786, 2020.

[Ma and Jiang, 2021] Yuanjing Ma and Hongmei Jiang. Ninimhmda: neural integration of neighborhood information on a multiplex heterogeneous network for multiple types of human microbe–disease association. *Bioinformatics*, 36(24):5665–5671, 2021.

[Ma *et al.*, 2017] Wei Ma, Lu Zhang, Pan Zeng, Chuanbo Huang, Jianwei Li, Bin Geng, Jichun Yang, Wei Kong, Xuezhong Zhou, and Qinghua Cui. An analysis of human

microbe–disease associations. *Briefings in bioinformatics*, 18(1):85–97, 2017.

[Qian *et al.*, 2018] Wei Qian, Cong Fu, Yu Zhu, Deng Cai, and Xiaofei He. Translating embeddings for knowledge graph completion with relation attention mechanism. In *IJCAI*, pages 4286–4292, 2018.

[Rajput *et al.*, 2018] Akanksha Rajput, Anamika Thakur, Shivangi Sharma, and Manoj Kumar. abiofilm: a resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance. *Nucleic acids research*, 46(D1):D894–D900, 2018.

[Sidorov *et al.*, 2019] Pavel Sidorov, Stefan Naulaerts, Jérémy Ariey-Bonnet, Eddy Pasquier, and Pedro J Ballester. Predicting synergism of cancer drug combinations using nci-almanac data. *Frontiers in chemistry*, 7:509, 2019.

[Skoufos *et al.*, 2021] Giorgos Skoufos, Filippos S Kardaras, Athanasios Alexiou, Ioannis Kavakiotis, Anastasia Lambropoulou, Vasiliki Kotsira, Spyros Tastsoglou, and Artemis G Hatzigeorgiou. Peryton: a manual collection of experimentally supported microbe-disease associations. *Nucleic acids research*, 49(D1):D1328–D1333, 2021.

[Sun *et al.*, 2018] Ya-Zhou Sun, De-Hong Zhang, Shu-Bin Cai, Zhong Ming, Jian-Qiang Li, and Xing Chen. Mdad: a special resource for microbe-drug associations. *Frontiers in cellular and infection microbiology*, page 424, 2018.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[Velickovic *et al.*, 2019] Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. *ICLR (Poster)*, 2(3):4, 2019.

[Wang *et al.*, 2007] James Z Wang, Zhidian Du, Rapeeporn Payattakool, Philip S Yu, and Chin-Fu Chen. A new method to measure the semantic similarity of go terms. *Bioinformatics*, 23(10):1274–1281, 2007.

[Wang *et al.*, 2021] Yueyue Wang, Xiujuan Lei, Cheng Lu, and Yi Pan. Predicting microbe-disease association based on multiple similarities and line algorithm. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021.

[Wang *et al.*, 2022a] Lei Wang, Yaqin Tan, Xiaoyu Yang, Linai Kuang, and Pengyao Ping. Review on predicting pairwise relationships between human microbes, drugs and diseases: from biological data to computational models. *Briefings in Bioinformatics*, 2022.

[Wang *et al.*, 2022b] Xiaolin Wang, Guohao Sun, Xiu Fang, Jian Yang, and Shoujin Wang. Modeling spatio-temporal neighbourhood for personalized point-of-interest recommendation. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2022.

[Wei *et al.*, 2022] Tianxin Wei, Yuning You, Tianlong Chen, Yang Shen, Jingrui He, and Zhangyang Wang. Augmentations in hypergraph contrastive learning: Fabricated and generative. *arXiv preprint arXiv:2210.03801*, 2022.

[Welling and Kipf, 2016] Max Welling and Thomas N Kipf. Semi-supervised classification with graph convolutional networks. In *J. International Conference on Learning Representations (ICLR 2017)*, 2016.

[Wu *et al.*, 2022] Shengbo Wu, Shujuan Yang, Manman Wang, Nan Song, Jie Feng, Hao Wu, Aidong Yang, Chunjiang Liu, Yanni Li, Fei Guo, et al. Quorum sensing-based interactions among drugs, microbes, and diseases. *Science China Life Sciences*, pages 1–15, 2022.

[Xia *et al.*, 2022a] Lianghao Xia, Chao Huang, Yong Xu, Jiashu Zhao, Dawei Yin, and Jimmy Huang. Hypergraph contrastive collaborative filtering. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*, pages 70–79, 2022.

[Xia *et al.*, 2022b] Lianghao Xia, Chao Huang, and Chuxu Zhang. Self-supervised hypergraph transformer for recommender systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2100–2109, 2022.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

[Yadati *et al.*, 2019] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. *Advances in neural information processing systems*, 32, 2019.

[Yang *et al.*, 2022] Haoran Yang, Hongxu Chen, Shirui Pan, Lin Li, Philip S Yu, and Guandong Xu. Dual space graph contrastive learning. In *Proceedings of the ACM Web Conference 2022*, pages 1238–1247, 2022.

[Yi and Park, 2020] Jaehyuk Yi and Jinkyoo Park. Hypergraph convolutional recurrent neural network. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3366–3376, 2020.

[Yu *et al.*, 2021] Junliang Yu, Hongzhi Yin, Jundong Li, Qinyong Wang, Nguyen Quoc Viet Hung, and Xiangliang Zhang. Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *Proceedings of the Web Conference 2021*, pages 413–424, 2021.

[Zhang *et al.*, 2020] Ruochi Zhang, Yuesong Zou, and Jian Ma. Hyper-sagnn: a self-attention based graph neural network for hypergraphs. In *International Conference on Learning Representations (ICLR)*, 2020.