Dear Prof. Jonathan Wren and Prof. Alison Hutchins,

**RE**: BIOINF-2024-1571, “Dynamic category-sensitive hypergraph inferring and homo-heterogeneous neighbor feature learning for drug-related microbe prediction”

Thank you very much for the constructive comments and the opportunity to resubmit the revised paper to Bioinformatics.

Enclosed is the revised version of the paper and point-by-point responses to the comments and suggestions.

Major revisions include:

* We added the description about the calculation of Gaussian kernel similarity between two drugs in the manuscript.
* We added the experiments to estimate the effect of the hyperedge number on the prediction performance. The experimental result and the analysis were added in the supplementary file.
* We conducted the ablation experiments on the node feature propagation strategy and the category features. The experimental results and the analysis were added in the manuscript.
* The time complexity analysis of DHDMP was listed in the supplementary file.

We look forward to hearing from you. We greatly appreciate for your effort and help.

Yours sincerely,

Ping Xuan on behalf of all the authors,

Department of Computer Science and Technology, Shantou University, China.

Email: [pxuan@stu.edu.cn](mailto:pxuan@stu.edu.cn)

Reviewer: 1  
  
Comments to the Author  
1) General comments  
The authors presented a drug-microbe association prediction model based on category-sensitive dynamic hypergraph learning, homogenous neighbor feature integration, and heterogeneous neighbor feature integration. The ideas that construction of hypergraph with dynamic topology, category-sensitive hypergraph learning, feature propagation across the homogeneous and heterogeneous graphs, and spatial cross-attention to capture long-distance correlations among multiple pairwise attribute patches are novel. The comparison experiments and the ablation studies showed the presented method achieved superior prediction performance and the effectiveness of its major innovations. The case studies also demonstrated that it could retrieve the potential candidate microbes associated with the drugs.  
**Response** – We thank the reviewer for the valuable comments. According to the suggestions, we carefully revised the manuscript and provided the following responses. In the revised manuscript, the revisions are highlighted in blue.

2) Specific comments for revision: a) major;  
1. Section 2.2, 2nd paragraph. The Gaussian kernel similarity between two drugs was calculated. Please give the calculation formula.  
**Response –** We added the calculation formula about Gaussian kernel similarity between two drugs in 2nd paragraph of Section 2.2 “Construction of homogeneous and heterogeneous graph for drugs and microbes”. The description is also listed as follows.

“The Gaussian kernel similarity between the *i-*th drug and the *j-*th one is computed based on the two groups of microbes interacted with them,

(3)

where is the *i-*th row of the drug-microbe association matrix. represents the normalized kernel bandwidth,

(4)

where is the original bandwidth and it was set to 1.”

2. Section 2.2, 2nd paragraph. The similarities and associations between a drug (microbe) node and another drugs or microbes are utilized as the attributes of the node. Why do you choose them as the node attributes?  
**Response** – For a pair of drug node and microbe node , contains the similarity between and all the drugs, and records the similarities between and all the microbes. is composed of the associations between and all the microbes, and records the associations between and all the drugs. If a pair of microbes and drugs are associated or similar with more common drugs and microbes, they are more likely to be associated. On the basis of the biological premise, we concatenate and to form the attribute of , and concatenate and to get the attribute of . We added the description in the second paragraph of section 2.2 “Construction of homogeneous and heterogeneous graph for drugs and microbes”. It is also listed as follows.

“If a pair of microbes and drugs are associated or similar with more common drugs and microbes, they are more likely to be associated. The *i*-th row of matrix Z, denoted as , contains the similarities and associations between node and all the drugs and microbes.”

3. Section 2.3. The dynamic hypergraph topology was constructed in the proposed method. What is the benefit of dynamic topology construction?  
**Response** – Thank you very much for your constructive suggestion. During the training process, the hyperedge embedding matrix E is continuously updated, forming a dynamically changed hypergraph topology. Its advantage is the hypergraph topology can be continuously adjusted based on the model's loss, which is helpful for better reflecting the complex relationships among multiple drugs and microbes. We added the description in the first paragraph of section 2.3 “Multiple drugs and microbes association learning based on NHCN”. The description is also listed as follows,

“The hypergraph topology is continuously updated during the training process, which is helpful for better reflecting the complex relationships among multiple drugs and microbes.”

4. Figure 2. There is an annotation “Iteration” in Figure 2. Please explain its meaning.  
**Response** – We added the description of annotation “Iteration” in the 2nd paragraph of Section 2.3 “Multiple drugs and microbes association learning based on NHCN”. It is also listed as follows,

“ is an attribute matrix embedded with the category features. We perform hypergraph convolution on all the nodes and then update the hyperedge embedding matrix. The updated hyperedge embedding matrix is multiplied with the node attributes to form a new hypergraph topology. We perform hypergraph convolution on the new hypergraph topology and repeat the iteration process (as shown in Figure 2).”

5. Section 2.4, 1st paragraph. It was mentioned that multiple types of neighbors reflect the characteristics of the target node from different perspectives. Please give the detailed explanation.  
**Response** – We added a description in the first paragraph of section 2.4 “Homogeneous and heterogeneous neighbor representation learning based on GCNFP”. The description is also listed as follows.

“Thus, the homogeneous neighbors reflect the characteristics of the target node from the perspective of similarity, while the heterogeneous neighbors reveal the characteristics of the target node from the perspective of association.”

6. Section 2.4, 1st paragraph. It was mentioned that “αg is a vector containing the learnable attention coefficients for each node”. How did you initialize αg?

**Response** –We added the explanation below formula (13). It is also listed as follows.

“As each drug (microbe) node has its own unique information, we set a learnable attention coefficient for each node to capture the unique information. The attention coefficient vector of all the nodes is and it is randomly initialized.”

7. Section 2.5. The same weight matrices including *Wq, Wk, Wv* were utilized to construct the query matrix, key matrix, and value matrix for both drug node and microbe node. What is the advantage of utilized the shared weight matrices?

**Response** – For all the patches of a pair of drug and microbe nodes, , , are the shared weight matrices. It is beneficial for jointly capturing the spatial correlations among the features of the drug-microbe pair. We added the description in section 2.5 “Long-distance spatial representation learning based on spatial cross-attention”, paragraph 2nd.

8. Section 3.1, 1st paragraph. The designed NHCN and GCNFP contained two encoding layers. How did you choose the layer numbers?

**Response** – The numbers of encoding layers in NHCN and GCNFP are selected from {1, 2, 3}. We conducted the experiments on all the combinations of layer number of NHCN and that of GCNFP (Supplementary Table ST1). ST1 and the corresponding analysis are also listed as follows.

Supplementary Table ST1: Prediction results for the encoding layer numbers of NHCN and GCNFP.

|  |  |  |  |
| --- | --- | --- | --- |
| Layer number of NHCN | Layer number of GCNFP | AUC | AUPR |
| 1 | 1 | 0.947 | 0.774 |
| 1 | 2 | 0.942 | 0.815 |
| 1 | 3 | 0.943 | 0.788 |
| 2 | 1 | 0.948 | 0.811 |
| 2 | 2 | 0.959 | 0.823 |
| 2 | 3 | 0.946 | 0.806 |
| 3 | 1 | 0.945 | 0.796 |
| 3 | 2 | 0.945 | 0.812 |
| 3 | 3 | 0.947 | 0.783 |

“The numbers of the encoding layers in NHCN and GCNFP are selected from {1, 2, 3}. We conducted the experiments on all the combinations of layer number of NHCN and that of GCNFP (Supplementary Table ST1). When NHCN and GCNFP contain two encoding layers, the model achieves the best prediction performance.”

Reviewer: 2  
  
Comments to the Author  
Authors propose one noel approach to predict the candidate microbes for drugs by learning the correlations among multiple drugs and microbes and encoding the features from the homogeneous and heterogeneous neighbors. The propose model mainly includes the following novelties: (1) the designed hypergraph with dynamic topology, (2) the proposed node category-sensitive hypergraph convolution strategy, (3) the feature transferring from the homogeneous graph to the heterogeneous graph, and (4) the spatial cross-attention for the pairwise feature learning. The authors perform the comparison experiments and the results demonstrated that their proposed method achieve the best prediction performance. Meanwhile, I have the following concerns.

**Response** – We thank the reviewer for the valuable comments. According to the suggestions, we carefully revised the manuscript and provided the following responses. In the revised manuscript, the revisions are highlighted in blue.

a) Section “Multiple drugs and microbes association learning based on NHCN”. The category features were designed for discriminating multiple categories of nodes. I suggest to add an ablation experiment to verify the effectiveness of category feature encoding.

**Response** – Thank you very much for your precious suggestion. We added the ablation experiment about the category features. The experimental results and the analysis are added to Table 1 and the first paragraph of Section 3.3 “Ablation experiments”.

“After the category features were removed from the prediction model, its AUC and AUPR decreased by 1.7% and 5.5%, respectively. It indicated that the category features may reveal the heterogeneity of multiple kinds of nodes.”

Table 1: Results of ablation studies of DHDMP

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Category  Feature | NHCN | GCNFP | SCA | Average AUC | Average AUPR |
| √ | × | √ | √ | 0.932 | 0.691 |
| √ | √ | × | √ | 0.944 | 0.693 |
| √ | √ | √ | × | 0.945 | 0.759 |
| × | √ | √ | √ | 0.942 | 0.768 |
| √ | √ | √ | √ | 0.959 | 0.823 |

b) Section “Multiple drugs and microbes association learning based on NHCN”. There is a sentence “Db and De are the degree matrices of nodes and hyperedges”. DB and DE formula (6) are also seem to be the degree matrices. Please check whether they are same matrices.

**Response** – We are very sorry for making the mistake. We changed to and changed to .

c) Section “Homogeneous and heterogeneous neighbor representation learning based on GCNFP”. Each drug (microbe) node has its own learnable attention coefficient. Is there any advance about it?

**Response** – As each drug (microbe) node has its own unique information, we set a learnable attention coefficient for each node. It is helpful for capturing the unique information of each node. We added the description in the first paragraph of section 2.4 “Homogeneous and heterogeneous neighbor representation learning based on GCNFP”. The description is also listed as follows.

“As each drug (microbe) node has its own unique information, we set a learnable attention coefficient for each node to capture the unique information. The attention coefficient vector of all the nodes is and it is randomly initialized.”

d) Section “Long-distance spatial representation learning based on spatial cross-attention”. The attributes of a pair of drug and microbe nodes are divided into multiple patches. What is the advantage of the attribute patch division?

**Response** – We divided the pairwise attributes into multiple patches and then designed the spatial cross-attention to capture the spatial correlations among these patches. We added the description in the first paragraph of section 2.5 “Long-distance spatial representation learning based on spatial cross-attention”. The description is also listed as follows.

“The partitioning of attribute patches is helpful for designing an SCA (spatial cross-attention) mechanism to capture the spatial correlations among the patches (as shown in Figure 3).”

e) Section “Long-distance spatial representation learning based on spatial cross-attention”. Both self-attention and cross-attention were proposed when the long-distance spatial relationships were encoded. It would be better to describe what each attention focuses on learning.

**Response** – We describe their differences in section 2.5 “Long-distance spatial representation learning based on spatial cross-attention”, paragraph 3rd. It is also listed as follows.

“The self-attention focuses on capturing the spatial correlations among the attributes of a drug (microbe) itself. The cross-attention concentrates on reveal the spatial correlations among the attributes of drug node and those of the microbe node.”

f) Section “Ablation experiments”. The full names of NHCN, GCNFP, and SCA were given in the introduction. I suggest to also list their full names in the corresponding strategy description.

**Response** – We listed the full name of NHCN in the second paragraph of section 2.3 “Multiple drugs and microbes association learning based on NHCN”, and gave the full name of GCNFP in the first paragraph of section 2.4 “Homogeneous and heterogeneous neighbor representation learning based on GCNFP”. We also listed the full name of SCA in section 2.5 “Long-distance spatial representation learning based on spatial cross-attention”. In addition, we listed the full names of NHCN, GCNFP, and SCA in the section 3.3 “Ablation experiments”.

g) Two symbols “*k*” appear in different positions. The first one is in “For each drug, we calculated the recall rates of candidate microbes at various top *k* values (Figure 4)”. The second one is in formula (23) where *k* represents the number of the convolutional-pooling layers. I suggest to use different symbols to denote the different meanings.

**Response** – We replaced the symbol “*k*” with “*l*” in equation (23) and updated the relevant description.

i) The description about DHDMP seems complex. Authors had better add a time complex analysis for DHDMP.

**Response** – We added the time complexity analysis in Supplementary file SF1. It is also listed as follows.

DHDMP consists of NHCN, GCNFP, and SCA. NHCN contains 2 hypergraph convolutional encoding layers, GCNFP contains 2 graph convolutional encoding layers, and SCA calculates the long-distance spatial correlations of the pairwise attributes for all the node pairs. Thus, the time complexity is,

where *Nd* is the number of drug nodes, *Nm* is the number of microbe nodes, and *Nf* is the attribute dimension with the category features, *Ns* is the attribute dimension without the category features, *Ne* is the number of hyperedges, *P* is the patch size, and *Ndim* is the feature dimension.

Reviewer: 3  
  
Comments to the Author  
The task of the manuscript is to infer the reliable microbe-drug association candidates, and a prediction model was presented based on node category-sensitive hypergraph convolution and graph convolutional network with cross-graph feature propagation. Several novel components were presented in the manuscript: construction of hypergraph with dynamic topology, encoding the heterogeneity of node attributes from multiple categories, cross-graph feature propagation, and the enhanced pairwise feature learning. The constructed prediction model was compared with 6 baselines and there is a significant improvement in microbe-drug association prediction. The manuscript is easy to follow and well-organized. I have several comments for the authors.  
**Response** – We thank the reviewer for the valuable comments. According to the suggestions, we carefully revised the manuscript and provided the following responses. In the revised manuscript, the revisions are highlighted in blue.

\*\*\*Major concerns\*\*\*:  
-- Page 4. The node feature propagation mechanism was presented to supplement the features from the homogeneous neighbors to the feature learning for the heterogeneous neighbors. The ablation experiment may be added to show the contribution of the feature propagation.

**Response –** Thank you very much for your valuable suggestion. We added the ablation experiment to verify the effectiveness of the node feature propagation strategy. The experimental results and analysis were added to Table 1 and the first paragraph of the section 3.3 “Ablation experiments”.

“After the node feature propagation was eliminated from the model, its AUC and AUPR decreased by 1.4% and 4.2%. It demonstrates that propagating node information from the homogeneous graphs to the heterogeneous graph facilitates the heterogeneous neighbor feature learning.”

Table 1: Results of ablation studies of DHDMP

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NFPS | NHCN | GCNFP | SCA | Average AUC | Average AUPR |
| √ | × | √ | √ | 0.932 | 0.691 |
| √ | √ | × | √ | 0.944 | 0.693 |
| √ | √ | √ | × | 0.945 | 0.759 |
| × | √ | √ | √ | 0.945 | 0.781 |
| √ | √ | √ | √ | 0.959 | 0.823 |

-- In the “Parameter settings” part, I wondered how the number of hyper-edges was selected. Will its number have an impact on the prediction performance of the model?

**Response –** Thank you very much for your precious suggestion**.** The number of hyperedges was selected from {16, 32, 64, 128}, and we added the experiments to demonstrate the effect of the hyperedge number on the prediction performance (ST2). It is also listed as follows.

“The number of hyperedges is Ne. The value of Ne is selected from {16, 32, 64, 128}, and the model achieves the best performance (AUC=0.959 and AUPR=0.823) when Ne is 32. A smaller value of Ne might make it difficult for the prediction model to fully encode the biological characteristic of the associations among multiple drug and microbe nodes. On the other hand, a larger number might introduce the noisy data into the hypergraph learning.”

Supplementary Table ST2: Prediction performance by adjusting the hyperedge numbers.

|  |  |  |
| --- | --- | --- |
| number of hyperedges | AUC | AUPR |
| 16 | 0.941 | 0.786 |
| 32 | 0.959 | 0.823 |
| 64 | 0.950 | 0.809 |
| 128 | 0.935 | 0.799 |

\*\*\*Minor concerns\*\*\*:  
-- Page 4, near equation (13). What does diag() in αG = diag(αg) mean?  
**Response –** We added the description in the first paragraph of section 2.4 “Homogeneous and heterogeneous neighbor representation learning based on GCNFP”. The description is also listed as follows.

“The operation may convert the vector into a diagonal matrix ”

-- Figure 2 and Figure 3 utilize “⮾” and “☉” to represent matrix multiplication. The symbols should be unified.  
**Response –** We use ⮾ to represent matrix multiplication for both Figure 2 and Figure 3.

-- The manuscript and the several compared methods used the same number of negative samples with that of positive samples to train their prediction models. What will happen when the more negative samples are used?  
**Response –** Thank you very much for your thoughtful suggestion.We conducted the experiments with more negative samples and added the results to the supplementary file SF1. It is also listed as follows.

The ratio of the positive samples (the known drug-microbe associations) to the negative samples (the unobserved drug-microbe associations) was nearly 1:95. When the ratio of the positive samples to the negative samples is 1:1, the prediction result was given in Table 2. We constructed the prediction models DHDMP1:2, DHDMP1:5, DHDMP1:10 by utilizing the datasets with the positive and negative example ratios of 1:2, 1:5, and 1:10, respectively. The experimental results are shown in Supplementary Table ST3, and DHDMP1:1 achieved the best prediction performance. The AUC and AUPR of DHDMP1:2 decreased by 1.1% and 1.9% respectively compared to DHDMP1:1. The AUC and AUPR of DHDMP1:5 also decreased by 1.3% and 5.4% respectively. It indicated that utilizing more negative samples to train the prediction model caused the worse prediction performance.

Supplementary Table ST3: Effects of different ratios of positive samples to negative samples on the performance of the proposed model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ratios | 1:1 | 1:2 | 1:5 | 1:10 |
| AUC | 0.959 | 0.948 | 0.946 | 0.931 |
| AUPR | 0.823 | 0.804 | 0.769 | 0.756 |

-- In case studies, how did you obtain the top 20 microbe candidates for each drug?

**Response –** We added the description in the first paragraph of the section 3.5 “Case studies”. It is also listed as follows.

“For each drug, the candidate microbes were ranked in descending order according to their association scores. The top 20 candidate microbes for each drug are regarded as its potential candidate microbes.”