Supplementary Materials for

DrugRepSS: Textual Semantic and Graph Structural Progressive Representation Learning for Drug Repositioning

# Supplemental Materials and Methods

## 1.1 Baseline Methods for Comparison

In the comparison experiment, we adopted two types of baseline methods. One is the existing drug repositioning methods, and the other is prediction methods based on embedding feature similarity. A detailed description of these baseline methods is described below:

1. Drug repositioning methods

LAGCN[1] is a drug repositioning model based on Graph Convolutional Network (GCN), which constructs a heterogeneous drug-disease network and applies GCN to predict the drug-disease relationship.

NIMCGCN[2] is a drug repositioning model that combines GCN with a neural induction matrix. It uses GCN from miRNA and disease-like networks to learn the deep features of diseases and employs a new neural induction matrix completion module to predict drug indications.

HNet-DNN[3] is a deep neural network model for drug-disease relationship prediction, which utilizes Deep Neural Network (DNN) to extract deep features from the drug-disease heterogeneous network.

DRONet[4] is a drug repositioning model that combines network embedding and sequencing learning by using the effectiveness comparison relationship between drugs as prior information.

1. Network embedding methods
2. Matrix decomposition-related methods

SVD is a commonly used matrix decomposition method that can decompose a matrix into the product of three matrices to achieve purposes such as data dimension reduction, denoising, and compression, thereby mapping nodes to a low-dimensional space.

HOPE[5] models the higher-order proximity as a tensor, and then maps the lower-dimensional proximity through tensor decomposition techniques.

GF[6] learns the embeddings of nodes by minimizing the difference between the adjacency matrix and the node representations.

GraRep[7] encodes the network structure information as multiple adjacency matrices of different orders and then maps these adjacency matrices to a low-dimensional space using matrix decomposition techniques.\

1. Random walk-related methods

DeepWalk[8] generates node sequences by conducting random walks in the network and utilizes technologies such as Word2Vec to convert these node sequences into low-dimensional vector representations.

Node2vec[9] generates diverse node sequences in the network by controlling the random walk strategy and uses technologies such as Skip-gram to convert the node sequences into low-dimensional vector representations.

1. Neural network-related methods

LINE[10] uses two neural network models to learn the first-order and second-order proximities of nodes and concatenates the learned representations to form the final node embeddings.

SDNE[11] uses a deep autoencoder to learn the embedding representations of nodes, while considering the structural information of the network and the proximities of nodes.

# Supplemental Results

## 2.1 Literature Verification

To verify the potential relationship between diabetes and the predicted other candidate drugs, some possible relationships were found by referring to medical literature in PubMed. For instance, the study by Jill et al. discussed that patients with diabetes have a higher risk of depression, and Escitalopram (rank=3) has the effect of treating both diabetes and depression[12]; the research by Al-Sofiani et al. indicated that Aspirin (rank=4) helps reduce the risk of cardiovascular disease in patients with diabetes[13]; Hassanzadeh et al. studied that Clonidine (rank=8) and gabapentin could reduce neuropathic pain and the severity of neuropathic pain in diabetic patients[14].

To verify the potential relationship between coronary heart disease and the predicted other candidate drugs, the medical literature in PubMed was referenced. Several possible relationships were found. For example, Schernthaner et al. studied that Metformin (rank=8) had a beneficial effect on patients with coronary heart disease[15]; Fu et al. found that the clinical effectiveness of Amlodipine (rank=9) could significantly reduce blood pressure indicators and blood lipid indicators in patients with hypertension and coronary heart disease, and had good safety[16].

## 2.2 Network Medicine Analysis

For coronary heart disease, the dense connectivity within the PPI network (34 real links vs. 2.697 expected links, P=5.66E-26) suggests that these proteins tend to have a tighter interaction than expected.

## 2.3 Molecular docking verification

For coronary heart disease, the predicted drug Ketamine (rank=4) was selected for docking verification. The docking results shows that the drug has a strong binding energy (BE=-5.97) to the coronary heart disease-associated protein ANXA5, and Ketamine docks to the amino acid residues ASP-322 of CX3CR.

# References

1. Yu Z, Huang F, Zhao X, et al. Predicting drug-disease associations through layer attention graph convolutional network[J]. Briefings in bioinformatics, 2021, 22(4): bbaa243.
2. Li J, Zhang S, Liu T, et al. Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction[J]. Bioinformatics, 2020, 36(8): 2538-2546.
3. Liu H, Zhang W, Song Y, et al. HNet-DNN: inferring new drug-disease associations with deep neural network based on heterogeneous network features[J]. Journal of Chemical Information and Modeling, 2020, 60(4): 2367-2376.
4. Yang K, Yang Y, Fan S, et al. DRONet: effectiveness-driven drug repositioning framework using network embedding and ranking learning[J]. Briefings in Bioinformatics, 2023, 24(1): bbac518.
5. Ou M, Cui P, Pei J, et al. Asymmetric transitivity preserving graph embedding[C]. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016: 1105-1114.
6. Ahmed A, Shervashidze N, Narayanamurthy S, et al. Distributed large-scale natural graph factorization[C]. Proceedings of the 22nd international conference on World Wide Web, 2013: 37-48.
7. Cao S, Lu W, Xu Q. Grarep: Learning graph representations with global structural information[C]. Proceedings of the 24th ACM international on conference on information and knowledge management, 2015: 891-900.
8. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations[C]. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014: 701-710.
9. Grover A, Leskovec J. node2vec: Scalable feature learning for networks[C]. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016: 855-864.
10. Tang J, Qu M, Wang M, et al. Line: Large-scale information network embedding[C]. Proceedings of the 24th international conference on world wide web, 2015: 1067-1077.
11. Wang D, Cui P, Zhu W. Structural deep network embedding[C]. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016: 1225-1234.
12. Mathews J, Dange S V, Dutta A, et al. An observational study of the effect of escitalopram and etizolam in type 2 diabetes mellitus patients with depression[J]. International Journal of Basic & Clinical Pharmacology, 2016, 5(5): 1.
13. Al-Sofiani M E, Derenbecker R, Quartuccio M, et al. Aspirin for primary prevention of cardiovascular disease in diabetes: a review of the evidence[J]. Current diabetes reports, 2019, 19: 1-10.
14. Hassanzadeh S, Bagheri S, Majid Ahmadi S, et al. Effectiveness of oral clonidine and gabapentin on peripheral neuropathy in diabetic patients in southwestern Iran: a randomized clinical trial[J]. BMC Endocrine Disorders, 2023, 23(1): 224.
15. Schernthaner G, Schernthaner G H. The right place for metformin today[J]. diabetes research and clinical practice, 2020, 159: 107946.
16. Fu J, Wang T, Li B, et al. The efficacy of rosuvastatin, amlodipine, and aspirin in the treatment of hypertension with coronary heart disease and its effect on platelet aggregation[J]. Disease Markers, 2022, 2022:1111438.