Appendix A FPSC-DTI supplement

A.1 Mean percentile ranking MPR

The mean percentile ranking (MPR) [1–3], a recall-based statistical metric, is adopted to evaluate the method's performance. This is a good evaluation criterion for one-class datasets [4, 5]. Specifically, for each drug d_i in the test set, we generate a ranked list of the targets sorted in descending order by the predicted scores between the current drug with all targets in the dataset. Let $rank_{ji}$ denote the percentile ranking (PR) of target t_j with drug d_i. The smaller the $rank$ value is, the better the prediction performance of the algorithm. For example, $rank_{ji} = 0\%$ indicates that drug d_i is predicted to interact with target t_j with the highest probability. Similarly, $rank_{ji} = 100\%$ signifies that drug d_i is predicted to interact with target t_j with the lowest probability. Herein, the definition of MPR is as follows:

$$MPR = \frac{\sum_{i=1}^{N_D^t} R_i}{N_D^t} \tag{1}$$

where N_D^t denotes the number of drugs in the test set, and R_i can be computed as follows:

$$R_i = \frac{\sum_{j=1}^{N_T^t} rank_{ji}}{N_T^t} \tag{2}$$

where N_T^t denotes the number of targets in the test set for the current drug d_i.

A.2 Cluster analysis of the four benchmark datasets

Since there is no class label information in the drug and target data of the four benchmark datasets, we firstly determine them the number of clusters. In this paper, we use the decision graph of DPCSA [6] to determine the number of clusters, which requires none predefined parameter. Then, we use INCK [7], an improved K-medoids algorithm, to cluster drug and target data, respectively. All the clustering results are given in Table A1.

Table A1: Clustering results of four benchmark datasets

Dataset	Number of objects in each cluster	
	Drug	Target
Enzyme	59,19,17,101,15,7,46,4,13,3 10,10,19,19,33,12,12,20,11,15	123,40,19,2,2,43,20,25,6,9 9,15,21,3,10,6,289,6,10,6
GPCR	48,27,11,19,16,6,14,11,6,22 9,10,10,14	30,23,10,4,5,16,7
IC	72,10,12,14,12,10 15,22,11,14,18	39,17,6,41,9,4,7,20 6,5,14,5,11,8,12
NR	18,15,7,9,5	12,5,5,4

References

[1] Christopher C Johnson. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems*, 27:1–9, 2014.

[2] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 263–272. Ieee, 2008.

[3] Ming Hao, Stephen H Bryant, and Yanli Wang. Open-source chemogenomic data-driven algorithms for predicting drug–target interactions. *Briefings in Bioinformatics*, 2018.

[4] Ming Hao, Stephen H Bryant, and Yanli Wang. A new chemoinformatics approach with improved strategies for effective predictions of potential drugs. *Journal of Cheminformatics*, 10:50, 2018.

[5] Xu Zhou, Enyu Dai, Qian Song, Xueyan Ma, Qianqian Meng, Yongshuai Jiang, and Wei Jiang. In silico drug repositioning based on drug-mirna associations. *Briefings in Bioinformatics*, 2019.

[6] Donghua Yu, Guojun Liu, Maozu Guo, Xiaoyan Liu, and Shuang Yao. Density peaks clustering based on weighted local density sequence and nearest neighbor assignment. *IEEE Access*, 7:34301–34317, 2019.

[7] Donghua Yu, Guojun Liu, Maozu Guo, and Xiaoyan Liu. An improved k-medoids algorithm based on step increasing and optimizing medoids. *Expert Systems with Applications*, 92:464–473, 2018.