Hindawi
BioMed Research International
Volume 2021, Article ID 5599263, 10 pages
https://doi.org/10.1155/2021/5599263

Research Article

Drug-Target Interaction Prediction via Dual Laplacian Graph Regularized Logistic Matrix Factorization

Aizhen Wang [1] **and Minhui Wang** [2]

[1] *Department of Pharmacy, The Affiliated Huai'an Hospital of Xuzhou Medical University and The Second People's Hospital of Huai'an, Huai'an 223002, China*
[2] *Department of Pharmacy, Lianshui People's Hospital Affiliated to Kangda College, Nanjing Medical University, Huai'an 223300, China*

Correspondence should be addressed to Minhui Wang; 12679@whut.edu.cn

Received 15 January 2021; Revised 6 March 2021; Accepted 13 March 2021; Published 28 March 2021

Academic Editor: Chang Tang

Copyright © 2021 Aizhen Wang and Minhui Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drug-target interactions provide useful information for biomedical drug discovery as well as drug development. However, it is costly and time consuming to find drug-target interactions by experimental methods. As a result, developing computational approaches for this task is necessary and has practical significance. In this study, we establish a novel dual Laplacian graph regularized logistic matrix factorization model for drug-target interaction prediction, referred to as DLGrLMF briefly. Specifically, DLGrLMF regards the task of drug-target interaction prediction as a weighted logistic matrix factorization problem, in which the experimentally validated interactions are allocated with larger weights. Meanwhile, by considering that drugs with similar chemical structure should have interactions with similar targets and targets with similar genomic sequence similarity should in turn have interactions with similar drugs, the drug pairwise chemical structure similarities as well as the target pairwise genomic sequence similarities are fully exploited to serve the matrix factorization problem by using a dual Laplacian graph regularization term. In addition, we design a gradient descent algorithm to solve the resultant optimization problem. Finally, the efficacy of DLGrLMF is validated on various benchmark datasets and the experimental results demonstrate that DLGrLMF performs better than other state-of-the-art methods. Case studies are also conducted to validate that DLGrLMF can successfully predict most of the experimental validated drug-target interactions.

1. Introduction

It is well known that drug discovery is a difficult and expensive process, and identifying potential drug-target interactions (DTIs) plays an important role in yielding successful candidate compounds for drug development. Predicting interactions between different drugs and targets can provide critical information by discovering off-target effects. Accurate prediction of DTIs can also substantially accelerate lead generation. Drug-target interaction prediction can be also regarded as a useful step in biomedical research and precision medicine [1–9]. However, it is still time consuming for traditional experimental approaches to identify potential DTIs, and the success rates are also very low. In addition, only a very limited number of DTIs have been experimentally validated. Therefore, it is necessary to develop computational methods for DTIs, which can significantly reduce both the time and labor costs, as well as improve the efficiency of drug discovery. Furthermore, there are various datasets which contain experimentally validated interactions of drugs and targets, such as KEGG [10], DrugBank [11], and GenBank [12], which also benefit the prediction of potential DTIs by using computational techniques.

In recent years, a large number of computational methods for DTI prediction have been proposed, and these methods are often based on some machine learning and data mining models, e.g., logistic regression [13, 14], support vector machine (SVM) [15–17], Bayesian classifiers [18], matrix completion [9], matrix factorization [19, 20], kernel learning [21, 22], and network inference [23–25]. For classification-

Table 1: The statistics of drugs, targets, and interactions in each dataset.

Datasets	NRs	GPCRs	ICs	Es
No. of drugs	54	223	210	445
No. of targets	26	95	204	664
No. of interactions	90	635	1476	2926
Average no. of drugs/target	3.46	6.68	7.24	4.41
Average no. of targets/drug	1.67	2.85	7.03	6.58
Sparsity of the interaction matrix (%)	93.59	97	96.55	99.01

based methods, they treat drug-target interaction pairs and noninteraction pairs as positive instances or negative instances and convert the DTI prediction problem into a label classification task [14, 17]. In [15], a genetic algorithm is used to screen related compounds; the drug-target pairs with strong binding capacity were found with SVM and particle swarm optimization. Garcasosa et al. [13, 18] used logistic regression and naive Bayesian classifiers for the classification of compounds. In [26], the experimentally validated targets are employed to train a SVM model and find potential proteins with similar structure. Network-based methods [23, 27, 28] utilize the theory of network and graph [29] and incorporate the drug and target similarities with experimentally validated interactions to infer the potential unknown drug-target interactions. Due to the strong learning capability of data representation, matrix factorization has also been used for drug-target interaction prediction, such as Bayesian matrix factorization [30], collaborative matrix factorization [31], and robust graph regularized matrix factorization [20]. The principle of these matrix factorization-based methods lies in that a high-dimensional drug-target interaction matrix can be decomposed into a multiplication of low-dimensional matrices, and the intrinsic property of original data can be well captured by these low-dimensional matrices. Due to the powerful feature representation and linear relation learning capability, deep neural network-based methods are also proposed to learn the relationship between drug and target [32, 33].

Although achieving great success by previous computational model learning-based methods, there are still some limitations and much room for improvement. Firstly, in previous methods, the experimentally validated and unknown interactions are treated equally during the learning process, which could deduce some noisy information. Secondly, in some learning models, drug features and target features are difficult to select. In order to solve the above issues, we propose a novel drug-target interaction prediction model based on logistic matrix factorization with dual Laplacian graph regularization term by using experimentally validated interactions, referred to as DLGrLMF briefly. In our DLGrLMF model, the chemical structure similarities between drug pairs, the genomic sequence similarities between target pairs, and the experimentally validated interactions are integrated together. The similarities between drug neighbors and target neighbors are exploited to represent the latent factor vector of the factorized matrices, and the potential interactions are determined by a probability score through the logic function.

The efficacy of the proposed DLGrLMF was evaluated on five benchmark datasets, and we compared DLGrLMF with several other state-of-the-art drug-target interaction prediction approaches in terms of 10-fold cross-validation, and the results demonstrate that the proposed DLGrLMF clearly outperforms other methods. In addition, in order to validate the ability to predict potential drug-target interactions, case studies are also performed and the results also demonstrate that DLGrLMF can accurately predict most of the experimental validated drug-target interactions.

2. Materials and Methods

2.1. Datasets Used in Experiments. In this work, four small-scale benchmark datasets and a large-scale dataset are used in the experiments to evaluate the DTI prediction performance of the proposed DLGrLMF model. The four small-scale datasets include nuclear receptors (NRs), G protein-coupled receptors (GPCRs), ion channels (ICs), and enzymes (Es) [34], there are four different types of target protein, and they are publicly available at http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/. As to the large-scale dataset DrugBank (DB) [35], it is a unique bioinformatics and cheminformatics resource which combines detailed drug data with comprehensive drug-target information. We use the data released on Jul. 03, 2018 (version 5.1.1), in our experiments. The drug and target data were extracted from the DrugBank database website at http://www.drugbank.ca/. We only use the approved drug-target interactions for experiments. To this end, there are totally 1936 drugs, 1609 targets, and 7019 approved drug-target interactions, respectively. We download the approved drug structures and approved target sequences from https://www.drugbank.ca/releases/latest#structures and https://www.drugbank.ca/releases/latest#target-sequences, respectively.

In Table 1, we present the detailed statistics of the five datasets. Three types of information for each dataset are summarized, including the similarities between drug pairs, the similarities between target pairs, and the experimental validated DTIs. Specifically, the validated DTIs are obtained from public datasets including KEGG BRITE [36], BRENDA [37], DrugBank [38], and SuperTarget [39].

2.2. Problem Formulation of DTI Prediction. In order to make the subsequent expression clearer, we first give a brief problem formulation of DTI prediction. Throughout this paper, we use two sets $\mathcal{T} = \{T_i\}_{i=1}^{t}$ and $\mathcal{D} = \{D_i\}_{i=1}^{d}$ to represent t targets and d drugs, respectively. The experimentally

validated DTIs are denoted as a binary matrix $\mathbf{A} \in \{0, 1\}^{d \times t}$. If a drug D_i has been experimentally validated to interact with a target T_j, then $A_{ij} = 1$; otherwise, $A_{ij} = 0$. The elements with a value of "1" in $\mathbf{A}$ represent the "known interactions" and can be regarded as positive observations, while the zero elements in $\mathbf{A}$ are set as "unknown interactions" and can be regarded as negative observations. In addition, the drug similarities are denoted as $\mathbf{DS} \in \mathbb{R}^{d \times d}$, and the target similarities are represented as $\mathbf{TS} \in \mathbb{R}^{t \times t}$. DTI prediction is aimed at discovering the potential interactions from the negative observations by using certain prior information of drugs and targets. The candidate drug-target interactions will be chosen as predicted interactions according to their predicted probabilities in descending order.

For each dataset, three matrices including $\mathbf{A}$, $\mathbf{DS}$, and $\mathbf{TS}$ are provided, which represent the drug-target interactions, drug similarities, and target similarities, respectively. Each entry of $\mathbf{DS}$ represents the similarity between a drug pair, which is measured by using SIMCOMP [40] that describes the chemical structure similarity between drugs. In SIMCOMP, the similarity between two compounds $c1$ and $c2$ can be computed as $S(c1, c2) = |c1 \cap c2| / |c1 \cup c2|$. As to $\mathbf{TS}$, genomic sequence similarity is used to denoting the similarity score between two proteins, which are obtained from the KEGG GENES dataset [36]. The sequence similarities between two proteins $p1$ and $p2$ are computed via a normalized version of the Smith–Waterman scores [41], which is defined as $S(p1, p2) = SW(p1, p2)/\sqrt{SW(p1, p1)SW(p2, p2)}$, where $SW(\cdot, \cdot)$ represents the original unnormalized Smith-Waterman score.

2.3. Proposed DLGrLMF Model. Matrix factorization has been developed for recommendation systems in the very beginning, which decomposes the observation data matrix $\mathbf{X} \in \mathscr{R}^{m \times n}$ into two low-dimensional matrices $\mathbf{U} \in \mathscr{R}^{m \times k}$ and $\mathbf{V} \in \mathscr{R}^{n \times k}$, where k is the so-called number of hidden factors. Then, $\mathbf{U}$ and $\mathbf{V}$ can be regarded as the latent representation of drugs and targets in the hidden space. In recent years, it has also been used for predicting the incRNA-miRNA relationship [42] and drug-target interaction [20]. In this work, we propose a drug-target interaction prediction model via logistic matrix factorization with dual Laplacian graph regularization. Here, the occurrence probability p_{ij} of a certain drug-target interaction is calculated based on the inner product of the latent factor vectors from drug and target. Specifically, p_{ij} can be formulated as follows:

$$p_{ij} = \frac{e^{\left(u_i v_j^T\right)}}{\left(1 + e^{\left(u_i v_j^T\right)}\right)}, \tag{1}$$

where u_i is a latent row vector to represent drug d_i, and v_j is a latent row vector to represent target t_j. Then, U and V can be used to represent the potential characteristics of all drugs and targets, respectively. In this work, $\mathbf{U}$ and $\mathbf{V}$ are initialized by zero-mean spherical Gaussian priors as follows:

$$\begin{aligned} p\left(\mathbf{U} \mid {\sigma_u}^2\right) &= \prod_{i=1}^{d} N\left(u_i \mid 0, {\sigma_u}^2 \mathbf{I}\right), \\ p\left(\mathbf{V} \mid {\sigma_v}^2\right) &= \prod_{j=1}^{t} N\left(v_j \mid 0, {\sigma_v}^2 \mathbf{I}\right), \end{aligned} \tag{2}$$

where $N(x_i \mid z, \mathbf{Z})$ represents a Gaussian probability density function with a mean of z, a variance of $\mathbf{Z}$, and an independent variable x_i.In drug-target interaction prediction studies, known DTIs are experimentally verified and they should be more reliable than unknown relationships. Therefore, we should allocate higher weights to those known DTIs [43, 44]. Specifically, the known DTI pairs and $n(n > 1)$ negative samples are used for training, where the constant n determines the significance of the interaction pairs. Then, the posterior distribution in logarithm can be written as follows:

$$\begin{aligned} &\log p\left(\mathbf{U}, \mathbf{V} \mid \mathbf{A}, {\sigma_u}^2, {\sigma_v}^2\right) \\ &\quad = \sum_{i=1}^{d} \sum_{j=1}^{t} n a_{ij} u_i v_j^T - \left(1 + n a_{ij} - a_{ij}\right) \log \left[1 + e^{\left(u_i v_j^T\right)}\right] \\ &\qquad - 12 {\sigma_u}^2 \sum_{i=1}^{d} \left\|u_i\right\|_2^2 - 12 {\sigma_v}^2 \sum_{j=1}^{t} \left\|v_j\right\|_2^2 + C, \end{aligned} \tag{3}$$

where C is a constant.

Although minimizing Equation (3) can exploit latent vectors globally to predict potential DTIs, the local similarity information implied among drugs and targets is not taken into consideration. Therefore, we use the drug similarities and the target similarities to boost the prediction performance. Instead of using all of the similarities in $\mathbf{DS}$ and $\mathbf{TS}$, we extract the local neighbor information of drugs and targets. As to different drugs, the local neighbor similarity matrix $\overline{\mathbf{DS}}$ can be obtained from $\mathbf{DS}$ by

$$\overline{\mathbf{DS}}(ij) = \begin{cases} \mathbf{DS}(ij), & \text{if} \quad d_i \in \mathcal{N}\left(d_j\right), \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where $\mathcal{N}(d_j)$ denoted the neighbors of drug d_j.In a similar way, the local neighbor similarity matrix $\overline{\mathbf{TS}}$ can be obtained from $\mathbf{TS}$ by

$$\overline{\mathbf{TS}}(ij) = \begin{cases} \mathbf{TS}(ij), & \text{if} \quad t_i \in \mathcal{N}\left(t_j\right), \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where $\mathcal{N}(t_j)$ denoted the neighbors of target t_j.Since the iterations between drugs and targets are complex, different to previous graph regularized methods that only use the first-order connections to reflect the local pairwise proximity between vertices in a graph [45–48], we use the second-order connection to constrain that similar drugs should be connected with similar targets. Therefore, we have the following similarity affinity matrices calculating form:

$$\overline{\mathbf{DS}}(ij) = \overline{\mathbf{DS}}(i)\overline{\mathbf{DS}}(j)^T, \quad \overline{\mathbf{TS}}(ij) = \overline{\mathbf{TS}}(i)\overline{\mathbf{TS}}(j)^T, \tag{6}$$

where $\overline{\mathbf{DS}}(i)$ and $\overline{\mathbf{DS}}(j)$ represent the ith and jth column of original $\overline{\mathbf{DS}}$, respectively, and $\overline{\mathbf{TS}}(i)$ and $\overline{\mathbf{TS}}(j)$ represent the ith and jth column of original $\overline{\mathbf{TS}}$, respectively.

The main idea of our proposed DLGrLMF model is under the assumption that if the chemical structure of two drugs from a drug pair is similar to each other, their latent representation should also be closed to each other. Similarly, the latent representation of two targets should also be similar to each other if their genomic sequence similarities are closed to each other. For drugs, we can minimize the following problem:

$$\min_{\mathbf{U}} \sum_{i=1}^{d}\sum_{j=1}^{d} \overline{\mathbf{DS}}(ij)\left\|u_i - u_j\right\|_2^2. \tag{7}$$

As to different targets, we have the following similar minimization problem:

$$\min_{\mathbf{V}} \sum_{i=1}^{t}\sum_{j=1}^{t} \overline{\mathbf{TS}}(ij)\left\|v_i - v_j\right\|_2^2. \tag{8}$$

By some simple algebra, Equations (7) and (8) can be transformed into the following form:

$$\min_{\mathbf{U}} Tr\left(\mathbf{U}^T\mathbf{L_D}\mathbf{U}\right), \tag{9}$$

$$\min_{\mathbf{V}} Tr\left(\mathbf{V}^T\mathbf{L_T}\mathbf{V}\right), \tag{10}$$

where $\mathbf{L_D} \in \mathbb{R}^{d\times d}$ is the corresponding Laplacian matrix of drugs with $\mathbf{L_D} = \mathbf{D_D} - \overline{\mathbf{DS}}$, and $\mathbf{D_D}$ is the diagonal matrix with $\mathbf{D_D}(i,i) = \sum_j \overline{\mathbf{DS}}(i,j)$. $\mathbf{L_T} \in \mathbb{R}^{t\times t}$ is the corresponding target Laplacian matrix of targets with $\mathbf{L_T} = \mathbf{D_T} - \overline{\mathbf{TS}}$, and $\mathbf{D_T}$ is the diagonal matrix with $\mathbf{D_T}(p,p) = \sum_q \overline{\mathbf{TS}}(p,q)$.By combining Equations (9) and (10) and the maximization of Equation (3) together, we have our final DLGrLMF model as follows:

$$\begin{aligned}&\min_{\mathbf{U},\mathbf{V}} \left(1 + na_{ij} - a_{ij}\right) \log\left[1 + e^{\left(u_i v_j{}^T\right)}\right] \\ &- \sum_{i=1}^{d}\sum_{j=1}^{t} na_{ij}u_i v_j{}^T + 12\sigma_u{}^2 \sum_{i=1}^{d}\left\|u_i\right\|_2^2 \\ &+ 12\sigma_v{}^2 \sum_{j=1}^{t}\left\|v_j\right\|_2^2 + Tr\left(\mathbf{U}^T\mathbf{L_D}\mathbf{U}\right) + Tr\left(\mathbf{V}^T\mathbf{L_T}\mathbf{V}\right),\end{aligned} \tag{11}$$

which is equal to the following problem:

$$\begin{aligned}&\min_{\mathbf{U},\mathbf{V}} \left(1 + na_{ij} - a_{ij}\right) \ln\left[1 + e^{\left(u_i v_j{}^T\right)}\right] \\ &- \sum_{i=1}^{d}\sum_{j=1}^{t} na_{ij}u_i v_j{}^T + \alpha Tr\left(\mathbf{U}^T(\mathbf{L_D} + \mathbf{I}_m)U\right) \\ &+ \beta Tr\left(\mathbf{V}^T(\mathbf{L_T} + \mathbf{I}_n)\mathbf{V}\right),\end{aligned} \tag{12}$$

where $\mathbf{I}_m$ and $\mathbf{I}_n$ represent two identity matrices with size $m \times m$ and $n \times n$, respectively. α and β are two nonnegative constants to balance the regularization terms. In Equation (12), the first and second terms constitute the logistic matrix factorization model to formulate the drug-target interaction probability. The third term is the Laplacian regularization term to capture the local relationship between drug pairs, and the fourth term is the Laplacian regularization term to capture the local relationship between target pairs.

As can be seen from Equation (12), DLGrLMF models the interaction probability between a drug-target pair by a logistic function and decomposes the probability matrix into drug-specific and target-specific latent vectors. In DLGrLMF, a biologically validated drug-target pair is treated as positive examples, while an unknown pair is treated as a negative example. In such a manner, DLGrLMF assigns higher weights to positive observations than negatives. Since the positive pairs are biologically validated and thus usually more trustworthy while the negative pairs could contain potential DTIs and are thus unreliable, our method can fully exploit the useful information in validated interaction pairs.

In this work, we use gradient descent to optimize Equation (12). Supposing the objective function is denoted as $\mathscr{F}$, then the partial derivatives of $\mathscr{F}$ with respect to $\mathbf{U}$ and $\mathbf{V}$ can be obtained as follows:

$$\begin{aligned}\frac{\partial\mathscr{F}}{\partial\mathbf{U}} &= \mathbf{P}\mathbf{V} + (n-1)(\mathbf{A}\odot\mathbf{P})\mathbf{V} - n\mathbf{A}\mathbf{V} + \alpha(\mathbf{I}_m + \mathbf{L_D})\mathbf{U}, \\ \frac{\partial\mathscr{F}}{\partial\mathbf{V}} &= \mathbf{P}^T\mathbf{U} + (n-1)\left(\mathbf{A}^T\odot\mathbf{P}^T\right)\mathbf{U} - n\mathbf{A}^T\mathbf{U} + \beta(\mathbf{I}_n + \mathbf{L}_T)\mathbf{V},\end{aligned} \tag{13}$$

where $\odot$ denotes the Hadamard product of two matrices. Each element of $\mathbf{P}$ (i.e., $p_i j$) is formulated by Equation (1), which denotes the probability of interaction between drug i and target i. $\mathbf{U}$ and $\mathbf{V}$ are randomly initialized. During the optimization process, $\mathbf{U}$ and $\mathbf{V}$ are updated until to be stable. After we get the final solution $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$, the final probability of interaction between drug i and target j can be calculated as follows:

$$\widehat{p}_{ij} = \frac{e^{\left(\widehat{u}_i\widehat{v}_j{}^T\right)}}{\left(1 + e^{\left(\widehat{u}_i\widehat{v}_j{}^T\right)}\right)}. \tag{14}$$

3. Experimental Results

3.1. Evaluation Metrics. In order to validate the efficacy of our proposed method, experiments on five datasets mentioned in

Table 2: Average AUPR values of different methods on the four datasets under CV_1 setting (the values following the symbol "±" are the standard deviations of 5 repetition results).

Methods	NRs	GPCRs	Ics	Es
BLMNII	0.642 ± 0.039	0.482 ± 0.017	0.647 ± 0.012	0.622 ± 0.014
WNN	0.568 ± 0.023	0.589 ± 0.021	0.586 ± 0.016	0.593 ± 0.015
CMF	0.578 ± 0.037	0.676 ± 0.012	0.855 ± 0.006	0.804 ± 0.004
GRMF	0.595 ± 0.024	0.677 ± 0.013	0.368 ± 0.017	0.325 ± 0.013
NRLMF	0.677 ± 0.035	0.679 ± 0.014	0.887 ± 0.012	0.846 ± 0.005
DLGRMC	0.697 ± 0.025	0.705 ± 0.016	0.897 ± 0.016	0.877 ± 0.006
DLGrLMF	0.713 ± 0.020	0.724 ± 0.012	0.908 ± 0.015	0.896 ± 0.006

Table 3: Average AUPR values of different methods on the four datasets under CV_2 setting (the values following the symbol "±" are the standard deviations of 5 repetition results).

Methods	NRs	GPCRs	Ics	Es
BLMNII	0.429 ± 0.046	0.309 ± 0.022	0.288 ± 0.027	0.287 ± 0.027
WNN	0.505 ± 0.503	0.287 ± 0.017	0.238 ± 0.035	0.236 ± 0.035
CMF	0.467 ± 0.056	0.357 ± 0.015	0.266 ± 0.033	0.269 ± 0.030
GRMF	0.483 ± 0.052	0.355 ± 0.019	0.287 ± 0.025	0.286 ± 0.024
NRLMF	0.543 ± 0.052	0.364 ± 0.017	0.347 ± 0.033	0.346 ± 0.034
DLGRMC	0.576 ± 0.053	0.375 ± 0.015	0.367 ± 0.023	0.367 ± 0.023
DLGrLMF	0.587 ± 0.018	0.395 ± 0.011	0.379 ± 0.014	0.382 ± 0.019

Table 4: Average AUPR values of different methods on the four datasets under CV_3 setting (the values following the symbol "±" are the standard deviations of 5 repetition results).

Methods	NRs	GPCRs	Ics	Es
BLMNII	0.414 ± 0.041	0.335 ± 0.013	0.203 ± 0.015	0.164 ± 0.014
WNN	0.515 ± 0.023	0.365 ± 0.006	0.320 ± 0.013	0.386 ± 0.015
CMF	0.486 ± 0.037	0.404 ± 0.008	0.357 ± 0.012	0.378 ± 0.005
GRMF	0.518 ± 0.028	0.364 ± 0.013	0.345 ± 0.019	0.348 ± 0.012
NRLMF	0.494 ± 0.046	0.412 ± 0.045	0.356 ± 0.014	0.397 ± 0.015
DLGRMC	0.528 ± 0.023	0.415 ± 0.016	0.364 ± 0.018	0.414 ± 0.013
DLGrLMF	0.538 ± 0.022	0.432 ± 0.013	0.388 ± 0.014	0.425 ± 0.011

Section 2.1 are conducted. Similar to several previous works [49–51], two evaluation metrics including precision-recall (PR) curves and the Area Under the Precision-Recall curves (AUPR) [52] are utilized for performance evaluation. Since we intend to avoid incorrect predictions being recommended by the prediction algorithms [52], AUPR is desirable for evaluation because it can penalize the false positives more.

3.2. Experiment Settings. In our experiments, we use other six drug-target interaction prediction techniques for performance comparison, they are bipartite local model using neighbor-based interaction-profile inferring (BLMNII) [49], weighted nearest neighbor profile (WNN) [50], collaborative matrix factorization (CMF) [31], graph regularized matrix factorization (GRMF) [51], neighborhood regularized logistic matrix factorization (NRLMF) [45] and label propagation with linear neighborhood information (LPLNI) [53], and dual Laplacian graph regularized matrix completion for drug-target interaction prediction (DLGRMC). For each method, we perform 5 repetitions of 10-fold cross-validation (CV) on different datasets. In each repetition, the observed DTI indicator matrix $\mathbf{A}$ was divided into 10 folds. Then, each fold was selected for testing while the remaining 9 folds were used for training; the final AUPR score was the average results over 5 repetitions.

Similar to previous works [31, 54, 55], we conduct CV under the following three different settings:

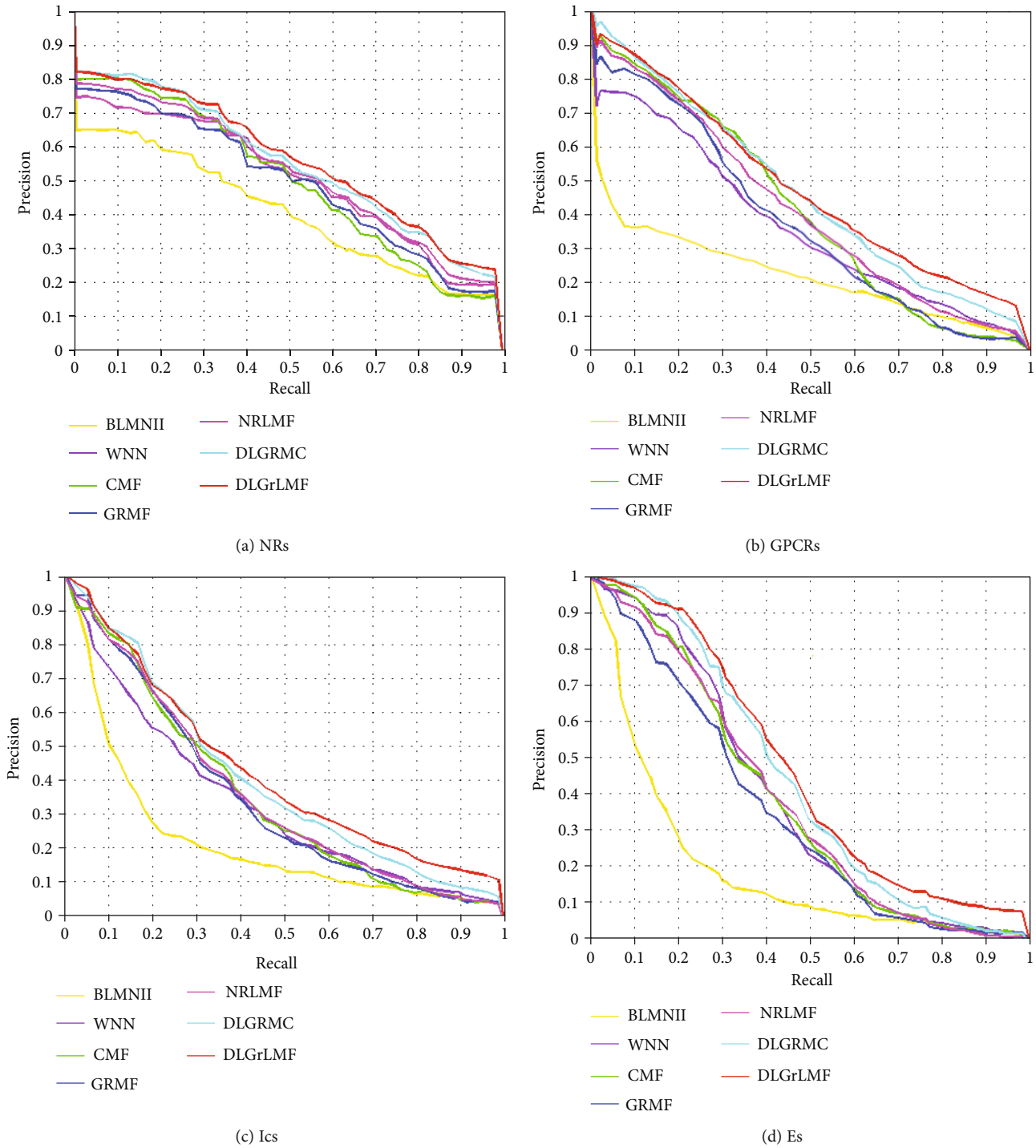

Figure 1: The PR curves of different methods on different datasets.

(i) CV_1: CV on drug-target pairs—we randomly select some entries from **A** (i.e., drug-target pairs) for testing, which refers to test the efficacy of the DTI prediction method for new (unknown) drug-target pairs

(ii) CV_2: CV on drugs—we randomly select several rows in **A** (i.e., drugs) for testing, which refers to the DTI prediction for new drugs

(iii) CV_3: CV on targets—we randomly select a portion of columns in **A** (i.e., targets) for testing; this setting refers to the DTI prediction for new targets

As to CV_1, CV_2, and CV_3, we use 90% of entries in **A**, 90% of rows in **A**, and 90% of columns in **A** as training data and the remaining data as testing data in each round, respectively.

Table 5: The top 10 interacted targets of drug "D00094" in dataset NRs predicted by different methods ("√" denotes experimentally validated targets, and "×" denotes nonvalidated targets).

Rank	Targets predicted by different methods						
	BLMNII	WNN	CMF	GRMF	NRLMF	DLGRMC	DLGrLMF
1	hsa5914 (√)	hsa190 (√)	hsa6096 (√)	hsa6257 (√)	hsa5915 (√)	hsa5914 (√)	hsa5914 (√)
2	hsa5915 (√)	hsa6257 (√)	hsa6257 (√)	hsa5915 (√)	hsa190 (√)	hsa5915 (√)	hsa5915 (√)
3	hsa6257 (√)	hsa5915 (√)	hsa5915 (√)	hsa6256 (√)	hsa6096 (√)	hsa190 (√)	hsa6096 (√)
4	hsa190 (√)	hsa6256 (√)	hsa190 (√)	hsa190 (√)	hsa5914 (√)	hsa6096 (√)	hsa190 (√)
5	hsa6258 (√)	hsa190 (√)	hsa6256 (√)	hsa6258 (√)	hsa6097 (√)	hsa6257 (√)	hsa5256 (√)
6	hsa6097 (√)	hsa6097 (√)	hsa5916 (√)	hsa5916 (√)	hsa6258 (√)	hsa6256 (√)	hsa6257 (√)
7	hsa2099 (×)	hsa5916 (√)	hsa2104 (×)	hsa5915 (√)	hsa5916 (√)	hsa6258 (√)	hsa6258 (√)
8	hsa4306 (×)	hsa2908 (×)	hsa2421 (×)	hsa2101 (×)	hsa6257 (√)	hsa5916 (√)	hsa5916 (√)
9	hsa5465 (×)	hsa2104 (×)	hsa4306 (×)	hsa2104 (×)	hsa367 (×)	hsa2099 (×)	hsa6097 (√)
10	hsa2104 (×)	hsa2421 (×)	hsa9970 (×)	hsa5465 (×)	hsa4306 (×)	hsa2908 (×)	hsa9970 (×)

Table 6: The top 10 interacted targets of drug "D00255" in dataset GPCRs predicted by different methods ("√ "denotes experimentally validated targets, and "×" denotes nonvalidated targets).

Rank	Targets predicted by different methods						
	BLMNII	WNN	CMF	GRMF	NRLMF	DLGRMC	DLGrLMF
1	hsa147 (√)	hsa150 (√)	hsa151 (√)	hsa155 (√)	hsa155 (√)	hsa147 (√)	hsa155 (√)
2	hsa148 (√)	hsa146 (√)	hsa146 (√)	hsa150 (√)	hsa147 (√)	hsa155 (√)	hsa151 (√)
3	hsa146 (√)	hsa155 (√)	hsa147 (√)	hsa151 (√)	hsa146 (√)	hsa151 (√)	hsa150 (√)
4	hsa150 (√)	hsa153 (√)	hsa148 (√)	hsa147 (√)	hsa150 (√)	hsa150 (√)	hsa147 (√)
5	hsa1812 (×)	hsa154 (√)	hsa155 (√)	hsa154 (√)	hsa148 (√)	hsa146 (√)	hsa148 (√)
6	hsa2550 (×)	hsa1234 (×)	hsa154 (√)	hsa1268 (×)	hsa2550 (×)	hsa154 (√)	hsa154 (√)
7	hsa2913 (×)	hsa1241 (×)	hsa2911 (×)	hsa135 (×)	hsa3361 (×)	hsa1128 (×)	hsa153 (√)
8	hsa5739 (×)	hsa3354 (×)	hsa1241 (×)	hsa2911 (×)	hsa5729 (×)	hsa2911 (×)	hsa2911 (√)
9	hsa7201 (×)	hsa7201 (×)	hsa3354 (×)	hsa57105 (×)	hsa9052 (×)	hsa3269 (×)	hsa3354 (×)
10	hsa552 (×)	hsa6751 (×)	hsa6751 (×)	hsa886 (×)	hsa2911 (×)	hsa3352 (×)	hsa1268 (×)

Table 7: The top 10 interacted targets of drug "D00110" in dataset ICs predicted by different methods ("√" denotes experimentally validated targets, and "× "denotes nonvalidated targets).

Rank	Targets predicted by different methods						
	BLMNII	WNN	CMF	GRMF	NRLMF	DLGRMC	DLGrLMF
1	hsa6336 (√)	hsa11280 (√)	hsa6530 (√)	hsa6532 (√)	hsa6529 (√)	hsa6331 (√)	hsa6530 (√)
2	hsa6532 (√)	hsa6530 (√)	hsa6532 (√)	hsa11280 (√)	hsa6532 (√)	hsa6336 (√)	hsa6532 (√)
3	hsa6530 (√)	hsa6529 (√)	hsa11280 (√)	hsa6336 (√)	hsa6336 (√)	hsa6530 (√)	hsa6336 (√)
4	hsa11280 (√)	hsa6331 (√)	hsa6529 (√)	hsa6336 (√)	hsa6331 (√)	hsa6532 (√)	hsa6331 (√)
5	hsa6529 (√)	hsa6532 (√)	hsa6331 (√)	hsa6530 (√)	hsa11280 (√)	hsa11280 (√)	hsa11280 (√)
6	hsa2554 (×)	hsa2554 (×)	hsa6336 (√)	hsa6529 (√)	hsa9312 (×)	hsa6529 (√)	hsa6529 (√)
7	hsa2901 (×)	hsa9177 (×)	hsa2901 (×)	hsa1137 (×)	hsa93589 (×)	hsa1141 (×)	hsa11254 (√)
8	hsa3748 (×)	hsa773 (×)	hsa27012 (×)	hsa9312 (×)	hsa23704 (×)	hsa1137 (×)	hsahsa10060 (×)
9	hsa1134 (×)	hsa8514 (×)	hsa8973 (×)	hsa3762 (×)	hsa2892 (×)	hsa9312 (×)	hsa9132 (×)
10	hsa9177 (×)	hsa9311 (×)	hsa2560 (×)	hsa1139 (×)	hsa3756 (×)	hsa93589 (×)	hsa9311 (×)

3.3. DTI Prediction Results. In Tables 2–4, we show the predicted AUPR values of different methods on different datasets under varying CV settings. As can be seen, our proposed DLGrLMF consistently outperforms other methods on all of the datasets. Considering that the drug discovery and development aim to serve the treatment of disease, in order to predict new targets which the drugs react, we plot the precision-recall (PR) curves of the results under CV_3 for all of the datasets. The PR curves are shown in Figure 1; the results also demonstrate the superiority of our proposed DLGrLMF.

Table 8: The top 10 interacted targets of drug "D00002" in dataset Es predicted by different methods ("√" denotes experimentally validated targets, and "×" denotes nonvalidated targets).

Rank	Targets predicted by different methods						
	BLMNII	WNN	CMF	GRMF	NRLMF	DLGRMC	DLGrLMF
1	hsa216 (√)	hsa108 (√)	hsa1725 (×)	hsa196883 (√)	hsa191 (√)	hsa191 (√)	hsa191 (√)
2	hsa108 (√)	hsa1725 (×)	hsa108 (√)	hsa191 (√)	hsa196883 (√)	hsa1725 (×)	hsa107 (√)
3	hsa1725 (×)	hsa191 (√)	hsa2936 (√)	hsa7498 (√)	hsa108 (√)	hsa196883 (√)	hsa108 (√)
4	hsa2746 (√)	hsa3939 (√)	hsa2639 (√)	hsa3033 (√)	hsa3292 (√)	hsa108 (√)	hsa6652 (√)
5	hsa196883 (√)	hsa3292 (√)	hsa115 (√)	hsa108 (√)	hsa3615 (√)	hsa2936 (√)	hsa64802 (√)
6	hsa7015 (×)	hsa349565 (√)	hsa2597 (√)	hsa7299 (×)	hsa3939 (√)	hsa3033 (√)	hsa4190 (√)
7	hsa4594 (×)	hsa34 (×)	hsa3156 (×)	hsa84152 (×)	hsa3818 (×)	hsa349565 (√)	hsa196883 (√)
8	hsa3035 (×)	hsa8435 (×)	hsa51095 (×)	hsa590 (×)	hsa5536 (×)	hsa339221 (×)	hsa3376 (×)
9	hsa306 (×)	hsa51095 (×)	hsa90 (×)	hsa3156 (×)	hsa34 (×)	hsa3156 (×)	hsa34 (×)
10	hsa8435 (×)	hsa306 (×)	hsa761 (×)	hsa34 (×)	hsa90 (×)	hsa3991 (×)	hsa3156 (×)

3.4. Case Study. In order to validate the capacity of DLGrLMF for potential DTI prediction, we randomly choose a drug from each dataset and report the top 10 predicted interactions of different methods under CV_3. The predicted results are reported in Tables 5–8. As can be seen from the results, our proposed DLGrLMF can successfully predict a larger amount of the experimental validated DTIs when compared with other methods, which also indicates that DLGrLMF is capable of predicting novel DTIs for drug development.

4. Discussion and Conclusions

In this paper, we propose a novel dual Laplacian graph regularized logistic matrix factorization model for drug-target interaction prediction, i.e., DLGrLMF. Specifically, DLGrLMF regards the task of drug-target interaction prediction as a weighted logistic matrix factorization problem, in which the experimentally validated interactions are allocated with larger weights. Meanwhile, by considering that drugs with similar chemical structure should have interactions with similar targets and targets with similar genomic sequence similarity should in turn have interactions with similar drugs, the drug pairwise chemical structure similarities as well as the target pairwise genomic sequence similarities are fully exploited to serve the matrix factorization problem by using a dual Laplacian graph regularization term. By performing extensive experiments, the efficacy of the proposed method can be well validated, and case studies demonstrate that the proposed method is powerful to predict potential novel drug-target interactions.

In addition, experimental results also demonstrate that there is still much room for improvement since there also exists missed interactions in case studies. In this work, only one type of representation for drugs or targets is used. In practical, each drug/target is often with multiple representations. For example, a drug can be represented by its chemical structure or by its chemical response in different cells. A protein target can be represented by its sequence or by its gene expression values in different cells. In our future work, we will try to integrate multiple representations for drug-target interaction prediction and we believe that the prediction results can be improved with a large margin.

Data Availability

The datasets used in this work are publicly available at http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Dr. Yong Liu for providing their code for implementing the logistic matrix factorization model.

References

[1] C. Dive and J. A. Hickman, "Drug-target interactions: only the first step in the commitment to a programmed cell death?," *Carcinogenesis: Integrative Cancer Research*, vol. 64, no. 1, pp. 192–196, 1991.

[2] M. A. Yıldırım, K. I. Goh, M. E. Cusick, A. L. Barabási, and M. Vidal, "Drug–target network," *Nature Biotechnology*, vol. 25, no. 10, pp. 1119–1126, 2007.

[3] N. A. Meanwell, "Drug-target interactions that involve the replacement or displacement of magnesium ions," *Bioorganic and Medicinal Chemistry Letters*, vol. 27, no. 24, pp. 5355–5372, 2017.

[4] R. A. Petrov, S. Y. Maklakova, Y. A. Ivanenkov et al., "Synthesis and biological evaluation of novel mono- and bivalent ASGP-R- targeted drug-conjugates," *Bioorganic and Medicinal Chemistry Letters*, vol. 28, no. 3, pp. 382–387, 2018.

[5] S. A. Ohnmacht and S. Neidle, "Small-molecule quadruplex-targeted drug discovery," *Bioorganic and Medicinal Chemistry Letters*, vol. 24, no. 12, pp. 2602–2612, 2014.

[6] J. Sun, M. Zhao, A. H. Fanous, and Z. Zhao, "Characterization of schizophrenia adverse drug interactions through a network approach and drug classification," *BioMed Research International*, vol. 2013, Article ID 458989, 10 pages, 2013.

[7] X. Li, G. Qin, Q. Yang, L. Chen, and L. Xie, "Biomolecular network-based synergistic drug combination discovery," *BioMed Research International*, vol. 2016, Article ID 8518945, 11 pages, 2016.

[8] H. Cui, M. Zhang, Q. Yang et al., "The prediction of drug-disease correlation based on gene expression data," *BioMed Research International*, vol. 2018, Article ID 4028473, 6 pages, 2018.

[9] M. Wang, C. Tang, and J. Chen, "Drug-target interaction prediction via dual Laplacian graph regularized matrix completion," *BioMed Research International*, vol. 2018, Article ID 1425608, 12 pages, 2018.

[10] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[11] C. Knox, V. Law, T. Jewison et al., "Drugbank 3.0: a comprehensive resource for omics research on drugs," *Nucleic Acids Research*, vol. 39, pp. D1035–D1041, 2010.

[12] D. L. Wheeler, T. Barrett, D. A. Benson et al., "Database resources of the national center for biotechnology information," *Nucleic Acids Research*, vol. 39, no. 1, pp. 13–16, 2003.

[13] A. T. García-Sosa, M. Oja, C. Hetényi, and U. Maran, "Druglogit: logistic discrimination between drugs and nondrugs including disease-specificity by assigning probabilities based on molecular properties," *Journal of Chemical Information & Modeling*, vol. 52, no. 8, pp. 2165–2180, 2012.

[14] Y. Tabei, E. Pauwels, V. Stoven, K. Takemoto, and Y. Yamanishi, "Identification of chemogenomic features from drugtarget interaction networks using interpretable classifiers," *Bioinformatics*, vol. 28, no. 18, pp. i487–i494, 2012.

[15] R. Bai, Q. Z. Teng, X. M. Yang, and S. Y. Yang, "Prediction of combinative activity of drugs and human serum albumin by using svm and ga," *Computer Engineering and Applications*, vol. 45, no. 12, pp. 226–229, 2009.

[16] N. Nagamine and Y. Sakakibara, "Statistical prediction of protein–chemical interactions based on chemical structure and mass spectrometry data," *Bioinformatics*, vol. 23, no. 15, pp. 2004–2012, 2007.

[17] Y. C. Wang, Z. X. Yang, Y. Wang, and N. Y. Deng, "Computationally probing drug-protein interactions via support vector machine," *Letters in Drug Design & Discovery*, vol. 7, no. 5, pp. 370–378, 2010.

[18] A. T. García-Sosa and U. Maran, "Drugs, non-drugs, and disease category specificity: organ effects by ligand pharmacology," *Sar & Qsar in Environmental Research*, vol. 24, no. 4, pp. 319–331, 2013.

[19] M. Gonen, "Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization," *Bioinformatics*, vol. 28, no. 18, pp. 2304–2310, 2012.

[20] Z. Cui, Y.-L. Gao, J.-X. Liu, L.-Y. Dai, and S.-S. Yuan, "L 2, 1-grmf: an improved graph regularized matrix factorization method to predict drugtarget interactions," *BMC Bioinformatics*, vol. 20, no. S8, p. 287, 2019.

[21] A. C. Nascimento, R. B. C. Prudêncio, and I. G. Costa, "A multiple kernel learning algorithm for drug-target interaction prediction," *BMC Bioinformatics*, vol. 17, no. 1, p. 46, 2016.

[22] X.-Y. Yan, S.-W. Zhang, and C.-R. He, "Prediction of drug-target interaction by integrating diverse heterogeneous information source with multiple kernel learning and clustering methods," *Computational Biology and Chemistry*, vol. 78, pp. 460–467, 2019.

[23] F. Cheng, C. Liu, J. Jiang et al., "Prediction of drug-target interactions and drug repositioning via network-based inference," *PLoS Computational Biology*, vol. 8, no. 5, article e1002503, 2012.

[24] J. You, R. D. McLeod, and P. Hu, "Predicting drug-target interaction network using deep learning model," *Computational Biology and Chemistry*, vol. 80, pp. 90–101, 2019.

[25] Y.-B. Wang, Z.-H. You, S. Yang, H.-C. Yi, Z.-H. Chen, and K. Zheng, "A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network," *BMC Medical Informatics and Decision Making*, vol. 20, no. S2, pp. 49–49, 2020.

[26] Z. W. Shang, L. I. Jin, Y. S. Jiang, M. M. Zhang, L. V. Hong-Chao, and R. J. Zhang, "A method of drug target prediction based on svm and its application," *Progress in Modern Biomedicine*, vol. 20, 2012.

[27] X. Chen, M. X. Liu, and G. Y. Yan, "Drugtarget interaction prediction by random walk on the heterogeneous network," *Molecular BioSystems*, vol. 8, no. 7, pp. 1970–1978, 2012.

[28] X.-Y. Yan, S.-W. Zhang, and S.-Y. Zhang, "Prediction of drug-target interaction by label propagation with mutual interaction information derived from heterogeneous network," *Molecular BioSystems*, vol. 12, no. 2, pp. 520–531, 2016.

[29] M. A. Thafar, S. Albaradie, R. S. Olayan, H. Ashoor, M. Essack, and V. B. Bajic, "Computational drug-target interaction prediction based on graph embedding and graph mining," in *Proceedings of the 2020 10th International Conference on Bioscience, Biochemistry and Bioinformatics*, pp. 14–21, Kyoto, Japan, January 2020.

[30] M. Gonen, "Predicting drugtarget interactions from chemical and genomic kernels using Bayesian matrix factorization," *Bioinformatics*, vol. 28, no. 18, pp. 2304–2310, 2012.

[31] X. Zheng, H. Ding, H. Mamitsuka, and S. Zhu, "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1025–1033, Chicago, Illinois, USA, August 2013.

[32] J. Peng, J. Li, and X. Shang, "A learning-based method for drug-target interaction prediction based on feature representation learning and deep neural network," *BMC Bioinformatics*, vol. 21, no. S13, pp. 1–13, 2020.

[33] K. Abbasi, P. Razzaghi, A. Poso, S. Ghanbari-Ara, and A. Masoudi-Nejad, "Deep learning in drug target interaction prediction: current and future perspective," *Current Medicinal Chemistry*, vol. 27, 2020.

[34] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drugtarget interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.

[35] D. S. Wishart, Y. D. Feunang, A. C. Guo et al., "Drugbank 5.0: a major update to the drugbank database for 2018," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1074–D1082, 2018.

[36] M. Kanehisa, S. Goto, M. Hattori et al., "From genomics to chemical genomics: new developments in kegg," *Nucleic Acids Research*, vol. 34, no. 90001, pp. D354–D357, 2006.

[37] I. Schomburg, A. Chang, C. Ebeling et al., "Brenda, the enzyme database: updates and major new developments," *Nucleic Acids Research*, vol. 32, no. 90001, pp. 431D–4433, 2004.

[38] D. S. Wishart, C. Knox, A. C. Guo et al., "Drugbank: a knowledgebase for drugs, drug actions and drug targets," *Nucleic Acids Research*, vol. 36, suppl_1, pp. D901–D906, 2008.

[39] S. Günther, M. Kuhn, M. Dunkel et al., "Supertarget and matador: resources for exploring drug-target relationships," *Nucleic Acids Research*, vol. 36, Supplement 1, pp. D919–D922, 2007.

[40] M. Hattori, N. Tanaka, M. Kanehisa, and S. Goto, "Simcomp/subcomp: chemical structure search servers for network analyses," *Nucleic Acids Research*, vol. 38, no. Web Server, pp. W652–W656, 2010.

[41] T. Smith and M. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, vol. 147, no. 1, pp. 195–197, 1981.

[42] H. Liu, G. Ren, H. Chen, Q. Liu, Y. Yang, and Q. Zhao, "Predicting lncRNA-miRNA interactions based on logistic matrix factorization with neighborhood regularized," *Knowledge-Based Systems*, vol. 191, p. 105261, 2020.

[43] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 739–748, Shanghai, China, November 2014.

[44] C. C. Johnson, "Logistic matrix factorization for implicit feedback data," *Advances in Neural Information Processing Systems*, vol. 27, p. 78, 2014.

[45] Y. Liu, M. Wu, C. Miao, P. Zhao, and X. L. Li, "Neighborhood regularized logistic matrix factorization for drug-target interaction prediction," *PLoS Computational Biology*, vol. 12, no. 2, article e1004760, 2016.

[46] Q. Zhao, Y. Zhang, H. Hu, G. Ren, W. Zhang, and H. Liu, "Irwnrlpi: integrating random walk and neighborhood regularized logistic matrix factorization for lncrna-protein interaction prediction," *Frontiers in Genetics*, vol. 9, p. 239, 2018.

[47] C. Tang, X. Zhu, X. Liu et al., "Learning a joint a_nity graph for multiview subspace clustering," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1724–1736, 2018.

[48] C. Tang, X. Liu, X. Zhu et al., "Feature selective projection with low-rank embedding and dual Laplacian regularization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 9, pp. 1747–1760, 2019.

[49] J. P. Mei, C. K. Kwoh, P. Yang, X. L. Li, and J. Zheng, "Drug-target interaction prediction by learning from local information and neighbors," *Bioinformatics*, vol. 29, no. 2, pp. 238–245, 2013.

[50] T. van Laarhoven and E. Marchiori, "Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile," *Plos One*, vol. 8, no. 6, article e66952, 2013.

[51] A. Ezzat, P. Zhao, M. Wu, X. Li, and C. K. Kwoh, "Drug-target interaction prediction with graph regularized matrix factorization," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 3, pp. 646–656, 2017.

[52] J. Davis and M. Goadrich, "The relationship between precision-recall and roccurves," in *International Conference on Machine Learning*, pp. 233–240, Pittsburgh, Pennsylvania, USA, June 2006.

[53] W. Zhang, Y. Chen, and D. Li, "Drug-target interaction prediction through label propagation with linear neighborhood information," *Molecules*, vol. 22, no. 12, p. 2056, 2017.

[54] H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu, "Similarity-based machine learning methods for predicting drugtarget interactions: a brief review," *Briefings in Bioinformatics*, vol. 15, no. 5, pp. 734–747, 2014.

[55] T. Pahikkala, A. Airola, S. Pietila et al., "Toward more realistic drugtarget interaction predictions," *Briefings in Bioinformatics*, vol. 16, no. 2, pp. 325–337, 2015.