Received: 20 July 2021 | Revised: 6 October 2021 | Accepted: 4 November 2021

DOI: 10.1049/syb2.12037

IET Systems Biology

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

PPDTS: Predicting potential drug–target interactions based on network similarity

Wei Wang[1,2,3] | **Yongqing Wang**[1] | **Yu Zhang**[1] | **Dong Liu**[1,2,3] | **Hongjun Zhang**[4] | **Xianfang Wang**[5]

[1]College of Computer and Information Engineering, Henan Normal University, Xinxiang, China

[2]Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province, Henan Normal University, Xinxiang, China

[3]Big Data Engineering Laboratory for Teaching Resources and Assessment of Education Quality of Henan Province, Henan Normal University, Xinxiang, China

[4]Computer Science and Technology, Anyang University, Anyang, China

[5]Computer Science and Technology, Henan Institute of Technology, Xinxiang, China

Correspondence

Wei Wang, College of computer and Information Engineering, Henan Normal University, Xinxiang, China.

Email: weiwang@htu.edu.cn

Funding information

Teaching Reform Research and Practice Project of Henan Normal University, Grant/Award Number: 201936; Natural Science Foundation of Henan Province, Grant/Award Numbers: 212300410367, 202300410102; Educational Science Research Foundation of Henan Normal University, Grant/Award Number: 2018JK19; Ph.D. Research Startup Foundation of Henan Normal University, Grant/Award Number: qd15130; National Natural Science Foundation of China, Grant/Award Numbers: 62072160, 62076089, 62072157; National Key Research and Development Program of China, Grant/Award Number: 2018YFB1701402; Science and Technology Research Key Project of Educational Department of Henan Province, Grant/Award Number: 21A520023; National Project Cultivation Fund Project of Henan Normal University, Grant/Award Number: 2020PL12; Key Project of Science and Technology Department of Henan Province, Grant/Award Numbers: 212102310381, 182102210208, 172102210338

Abstract

Identification of drug–target interactions (DTIs) has great practical importance in the drug discovery process for known diseases. However, only a small proportion of DTIs in these databases has been verified experimentally, and the computational methods for predicting the interactions remain challenging. As a result, some effective computational models have become increasingly popular for predicting DTIs. In this work, the authors predict potential DTIs from the local structure of drug–target associations' network, which is different from the traditional global network similarity methods based on structure and ligand. A novel method called PPDTS is proposed to predict DTIs. First, according to the DTIs' network local structure, the known DTIs are converted into a binary network. Second, the Resource Allocation algorithm is used to obtain a drug–drug similarity network and a target–target similarity network. Third, a Collaborative Filtering algorithm is used with the known drug–target topology information to obtain similarity scores. Fourth, the linear combination of drug–target similarity model and the target–drug similarity model are innovatively proposed to obtain the final prediction results. Finally, the experimental performance of PPDTS has proved to be higher than that of the previously mentioned four popular network-based similarity methods, which is validated in different experimental datasets. Some of the predicted results can be supported in UniProt and DrugBank databases.

KEYWORDS

Big Data, biocomputing, bioinformatics, biology computing, drugs

1 | INTRODUCTION

The identification of drug–target interactions (DTIs) plays a key role in the early stage of drug discovery. Identifying drug–target associations can provide important information for drug discovery and drug repositioning [1–3]. However, biological experiments are affected by a large amount of experimental data, high-accuracy requirements as well as excessive experimental expenses, so it is difficult to quickly identify a lot of potential DTIs. Nowadays, the potential drug–target interactions can be predicted based on computer algorithms, thereby greatly improving the efficiency of drug exploitation [4–6].

In the early stage, some researchers identified drug–target association from literature through text mining technique

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *IET Systems Biology* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

[7, 8], and explored drug–target association from common biological elements of drug and target [9, 10]. Some methods based on text mining collect the known drug–target association from literature, but they cannot predict the new associations. In fact, a large number of drugs and targets have no common elements [11], which also reduces the ability of text mining methods to recognise DTIs.

Recently, many machine-learning methods have been introduced into the drug–target association prediction, because they have the capability of handling complicated data. For example, Yamanishi et al. [12] integrated chemical information and genomic characteristics with pharmacological data into a focussed framework to predict potential DTIs on a large scale. Wang et al. [13] inputs the sequence information of the target into the stacked autoencoder and then inputs the characteristic and the fingerprint information into the rotating forest classifier to predict the potential DTIs. Palma et al. [14] proposed a method named semEP, assuming that similar drugs are more likely to interact with common targets, combined semantic similarity and edge division methods with the machine learning framework of link prediction to identify new drug–target. Based on the relationship between biomolecules, Lee et al. [15] constructed a directed network of protein interactions and gene data, consequently inferred the shortest path between targets and genes. Li et al. [16] integrated heterogeneous data networks, and then used DeepWalk to obtain node topology information, and hence inferred potential gene–disease interactions. Chen et al. [17] creatively used unsupervised pre-training and supervised fine-tuning to predict associations of miRNA-disease. Based on the known small molecule-miRNA association and miRNA-disease association, Qu et al. [18] constructed a three-layer heterogeneous network of small molecule similarity, miRNA similarity and disease similarity to predict unknown interactions. Zhao et al. [19] integrated the similarity matrix for matrix decomposition and calculated the Kronecker product of the newly integrated similarity matrix to obtain the prediction scores. Wang et al. [20] constructs multi-layer heterogeneous networks based on known small molecule-miRNA interactions, and then infers the potential representation of all layers by using the intra layer topology and known cross layer association for small molecule-miRNA interactions prediction. Zhang et al. [21] proposed a method based on linear domain similarity network to predict miRNA-disease associations. The author first transformed the known miRNA-disease associations into binary networks, and then scored the potential correlations based on label propagation algorithm. Luo et al. [22] integrated a variety of drug–target related data to build heterogeneous networks, and then predicted potential drug–target interactions by vector projection method. Wang et al. [23] constructed a drug–protein interaction network and combined link prediction and binary network to predict drug–protein interaction.

Although machine learning methods had been proposed to drug–target association prediction, the predictive performance of many methods needs to be improved. First, a large number of methods use the characteristics of drugs and targets with the known drug–target correlations to predict DTIs. However, not all drugs and targets have complete characteristics. If the information is incomplete, the prediction method cannot be effectively predicted. Second, some researchers found that the traditional similarity-based methods are effective for specific protein classes but not for other classes. Third, the traditional machine learning model will lose part of the information when building features to make them readable, and it is difficult to recover the lost information, which is rarely concerned [24].

Since it is more likely that neighbouring nodes in the network have the same characteristic information, these nodes in a certain area (a row or a column on binary network) may have more characteristics in common [25]. When the drug molecular structural information is incomplete, the calculation method based on the similarity of binary network is more efficient than that of the heterogeneous network of the molecular structural characteristics. Therefore, we focus on the local structural similarity of nodes in DTIs networks.

In this study, we propose a novel method called PPDTS, which transforms the known DTI into the bipartite network, and use the local structural information to predict potential DTIs. First, PPDTS builds a binary network based on the known DTIs. Second, the drug–drug similarity graph and target–target similarity graph can be obtained on the basis of the Resource Allocation algorithm [26] and the nodes' topology in the binary network. Third, the Collaborative Filtering method is applied to the drug–drug similarity map and the target–target similarity map to predict potential drug–target associations. Finally, PPDTS linearly combines the drug–target similarity network with the target–drug similarity network to produce the prediction result. In our experiments, we use fivefold cross validation in two different datasets and compared the performance of PPDTS with the experimental baseline methods (ProCF, RWR, DTINet and LP), which is significantly proved to be better than the other methods, and successfully predicted 37 potential DTIs. Some of the prediction drug–target interactions have been verified in UniProt and DrugBank.

2 | MATERIALS AND METHODS

2.1 | Materials

The PPDTS mainly uses the following databases for experiments. DrugBank is a comprehensive online database with experimental support [27], covering information related to drugs (chemistry, pharmacology and pharmacy) and drug–targets (sequence information, structural information and pathways). MATADOR [28] is a free online resource database containing protein-drug interactions. The MATADOR database has collected 801 compounds and 2901 target proteins, of which 15,843 compounds are connected to the targets, of which 8936 direct interactions account for 56.40%.

First, we download the drug–target interaction information from the DrugBank database. The dataset is called the golden dataset in drug–target prediction, and it is noticed that Wen et al. [29] also uses the same dataset. The original dataset has

6262 drug–target interactions; we deleted duplicate drug–target items, and in order to obtain more neighbour node information in the process of similarity calculation, we removed single drug–target interaction; finally, we got 786 drugs and 527 targets, which include a total of 4547 drug–target interactions. Second, we downloaded 8936 drug–target datasets with direct interactions from the latest MATADOR database. After the dataset was preprocessed in the same way, 8008 drug–target interactions were obtained including 575 drugs and 981 targets. In the experiment, two datasets are used to evaluate the performance of PPDTS, respectively.

Both of the datasets have 4547 and 8008 drug–target associations as positive samples. At the same time, because the number of non-interactions is in bigger quantities than that of direct interactions, in order to avoid a high false positive rate, we randomly selected equal amounts (4547 and 8008) as negative samples from the remaining non-interacting drug–target. We also analysed the ratio of drugs and targets in the negative samples, which is shown in Figure S1. The results showed that more than 82% of drugs and targets are selected, and the negative samples generally cover the main types of drugs and targets in the experimental data. Table 1 shows the information of two datasets in the experiment. Dataset 1 is used for experiments and verification, and Dataset 2 is used as an independent test set. The Datasets are available in https://github.com/HNUBioinformatics/PPDTS.git.

2.2 | Predicting potential drug–target interactions based on network similarity (PPDTS)

The specific implementation process of PPDTS is shown in Figure 1, which can be subdivided into the following steps. First, PPDTS converts the known DTIs into a binary network and takes the drugs set D and the targets set T as the starting node set. The drug–drug similarity and the target–target similarity are calculated by using the forward and reverse secondary propagation of the Resource Allocation algorithm. Second, based on similar neighbour nodes of D and T, drug–target similarity score network and target–drug similarity score network are computed by the Collaborative Filtering algorithm. Finally, the two similar networks are linearly integrated to infer the final prediction results; then the scores of the predicting results are ranked from high to low, and the predicting results of the higher score are more likely to be the potential DTIs. Here, we only consider the interaction with the highest score ($d[i, :]$ or $t[:, j]$) calculated by PPDTS as the final experimental prediction result.

The drug–target dataset is described as a binary network $V = (D, T, E)$. $D = \{d_1, d_2, d_3 \ldots d_m\}$ is a collection of drug nodes, m represents the total number of drugs in the dataset. $T = \{t_1, t_2, t_3 \ldots t_n\}$ is the set of target nodes, n represents the total number of targets in the dataset. $E = \{e_{i1}, \ldots, e_{ij} \ldots e_{mn}\}$ is the set of edges between interconnected nodes in the network, where D and T, respectively, represent two independent sets. If there is a known interaction between the drug d_i and the target t_j , then set $e_{ij} = 1$, otherwise set $e_{ij} = 0$ [30].

TABLE 1 PPDTS experimental datasets

Datasets	Drugs	Targets	Associations
Dataset1	786	527	4547
Dataset2	575	981	8008

In the actual drug–target interactions' network, a drug can bind on multiple targets, and the same target can be bound by multiple drugs. When the degree of nodes in the DTI network is larger, Zhou [26] shows that the accuracy of similarity calculation methods such as Adamic-Adar index [31], cosine similarity [32] and Jaccard similarity [33] coefficient are lower. In addition, it takes a long time and a large amount of memory to calculate the similarity based on the global nodes of the network. Therefore, this study aimed at the local structure of the network, and makes full use of the connection between adjacent nodes to calculate the similarity of network nodes.

Resource allocation algorithm (RA) is used to calculate the similarity between drug and target, and we focus on the similarity of local structure in the network. Potential drug–target interactions are not directly connected in known DTI networks. According to the principle of the Resource Allocation algorithm [26], the drug (or target) node can send some resources to the target (or drug) node, and their public neighbours act as transmitters. We assume that the initial size of the transmitter is a resource unit, and the initial resources are allocated to all adjacent transmitters on average according to the topological relationship in the DTI network. The calculation process of drug–drug similarity and target–target similarity is divided into two steps. First, D and T are regarded as the starting node set in turn, and each transmitter is assigned an initial resource unit of size 1, that is, $t_j = 1$, $d_i = 1$, transmitters evenly allot the initial resources to neighbour nodes. The calculating process of the first transmission from the target t_i to the drug d_j is defined below:

$$S_{t_i d_j} = \sum\nolimits_1^n \frac{1}{k_i} \tag{1}$$

k_i represents the number of interconnections between t_i and the drug node set D, and n are the number of connections between d_j and the target node set T. $S_{t_i d_j}$ are the first transmission scores of the target t_i to the drug d_j.

In the second transmission, the similarity between targets can be defined as STT_{mn}:

$$STT_{mn} = \sum\nolimits_{j \in (\Gamma_m \cap \Gamma_n)} \left(S_{t_m d_j} + S_{d_j t_n} \right) \tag{2}$$

The Γ_m represents the number of neighbour nodes of node m and Γ_n represents the number of neighbour nodes of node n, and the value of STT_{mn} is 0 to 1.

Similarly, $S_{d_i t_j}$ is calculated by Formula (1), and SDD_{mn} is calculated by Formula (2). Figure 2 shows a detailed flowchart for calculating the similarity of the target–target network.

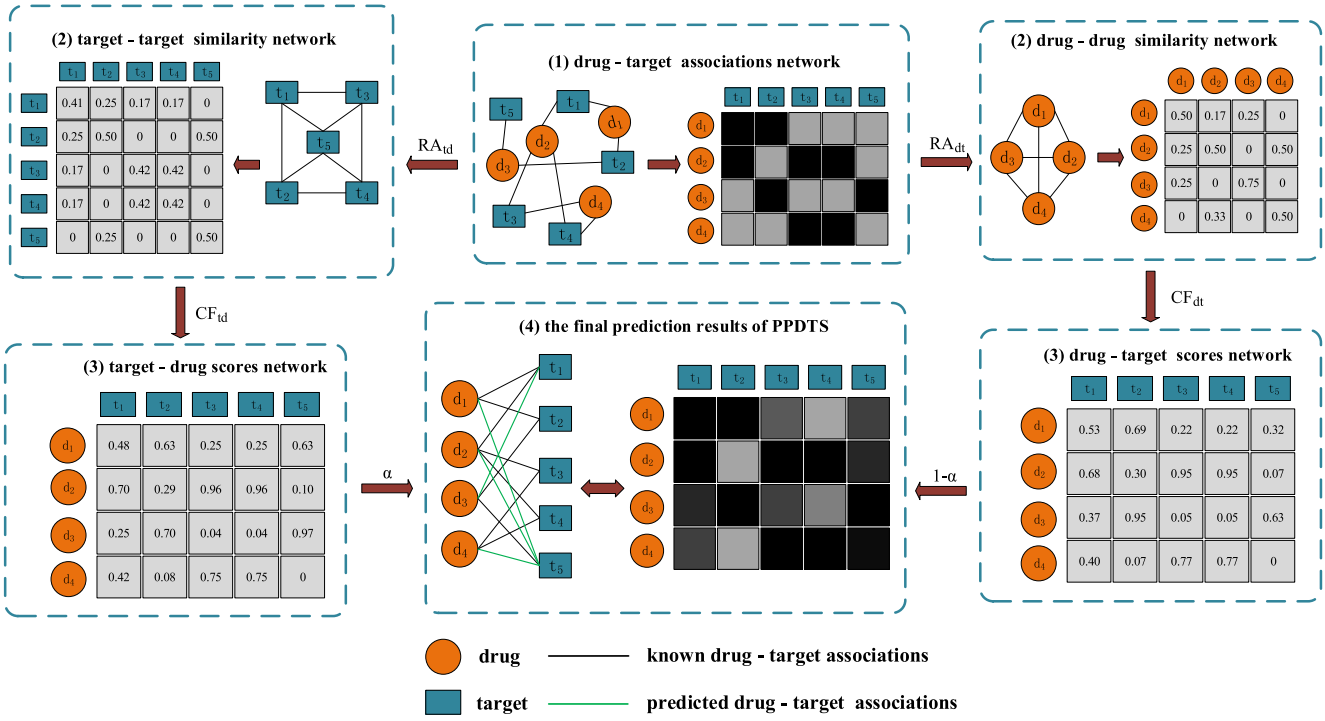

FIGURE 1 The flowchart of the PPDTS pipeline. RA_{td} represents that the target–target similarity network is calculated using the Resource Allocation algorithm; RA_{dt} represents that the drug–drug similarity network is calculated using the Resource Allocation algorithm; CF_{td} represents that the Collaborative Filtering algorithm is used to calculate the target–drug score network; CF_{dt} represents that the Collaborative Filtering algorithm is used to calculate the drug–target score network; α is the combination weight of drug–target and target–drug models ($0 \le \alpha \le 1$). Black solid lines represent the known DTIs, and the green solid lines represent the potential DTIs predicted by PPDTS. The different shades of grey indicate binding weights of drug-target in step (4), while the darker colors represent the greater probability of binding

The drug–target prediction score is calculated by using the Collaborative Filtering algorithm:

$$Score_{d_m t_n} = \frac{\sum_{k=1,k\neq n}^{x} STT(t_n, t_k) \cdot \gamma_{km}}{\sum_{k=1,k\neq n}^{x} STT|(t_n, t_k)|} \cdot E_{mn} \quad (3)$$

$Score_{d_m t_n}$ is the similarity score of drug d_m and target t_n from the similarity between drugs. $STT(t_n, t_k)$ is the similarity between target t_n and target t_k. x is the neighbour set of target t_n, and γ_{km} is the interaction matrix of drug d_m evaluated by its nearest neighbour k of target t_n. E is the drug–target interaction matrix, and the elements of E_{mn} is 1 if drug m is known to interact with target n, and the elements of E_{mn} is 0 otherwise.

In order to better explore the potential interaction in known DTI networks, we combine the Collaborative Filtering algorithm with E. E is not included in the original Collaborative Filtering algorithm. The subsequent experimental results show that the model is effective for predicting DTIs.

Likewise, the target–drug prediction score $Score_{t_m d_n}$ can be obtained in the same way.

Finally, the linear combination of $Score_{d_m t_n}$ and $Score_{t_m d_n}$ can get the final prediction result:

$$PPDTS = \alpha Score_{d_m t_n} + (1 - \alpha) Score_{t_n d_m} \quad (4)$$

The α is the weight value between the $Score_{d_m t_n}$ model and the $Score_{t_n d_m}$ model ($0 \le \alpha \le 1$).

3 | RESULTS AND DISCUSSION

3.1 | Baseline methods

In order to evaluate the prediction performance of PPDTS, this study compares Collaborative Filtering (ProCF) and Random Walk with Restart (RWR) to predict DTIs. The Collaborative Filtering (CF) method was initially used in recommendation systems to provide users with personalised recommendations [34]. Subsequently, especially in the case of a large amount of sparse data, ProCF was developed and achieved better results [35, 36]. The ProCF method is mainly based on inferring similar information from similar neighbour nodes. By calculating the similarity between different targets, the predicted score is calculated [37]. Compared with the calculating similarity by relying on neighbouring nodes, RWR can capture the global structural information of the network more comprehensively [38, 39]. Starting from a certain node in the target set, RWR selects adjacent nodes with probability c ($0 < c < 1$) at random, or return to the previous node with probability ($1-c$). The probability value obtained after multiple iterations can be regarded as the score of predicted similarity between different targets [40].

Our method is compared with the new machine learning method DTINet [22] and the typical network local structure similarity algorithm LP (Label Propagation). DTINet integrated a variety of drug–target related data to build

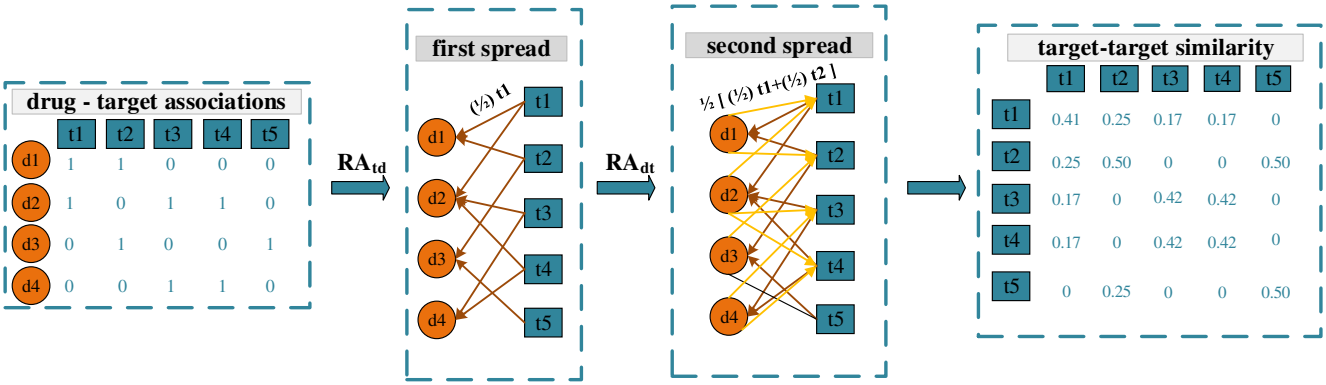

FIGURE 2 Target–target similarity calculation process. (1) The binary matrix represents the known drug–target associations, RA_{td} indicates that the target node set T evenly distributes the initial resources to the adjacent drug node set D; (2) The brown arrows are the process of the starting node targets to the drugs resource propagation, $(1/2)t1$ represents the propagation proportion of $t1 \rightarrow d1$, RA_{dt} indicates that the drug node set D evenly distributes the resources (the resources are obtained in RA_{td}) to the adjacent target node set T; (3) The yellow arrows are the drugs to the targets' resource propagation process, $(1/2)[(1/2)t1+(1/2)t2]$ represents the propagation proportion of $d1 \rightarrow t1$; (4) This represents the similarity between targets and targets, which be calculated by the Resource Allocation algorithm

heterogeneous networks, and then predicted potential drug–target interactions by the vector projection method. LP is widely used in semi-supervised learning of graph networks [41]. LP is mainly divided into two steps: First, LP constructs the target label matrix, and propagates it to the adjacent nodes according to the probability value τ. Second, LP updates the target similarity matrix to make it convergent, so as to obtain the target–target similarity score matrix [42, 43].

In order to systematically evaluate the performance of the method, fivefold cross validation is used to evaluate the generalisation ability of PPDTS. The experimental dataset is divided into five parts, and one sample set is randomly selected for testing, and the remaining four samples datasets are used for training. This process is performed five times in total, and the average of five times is taken as the final result. In addition, six indicators are used to evaluate PPDTS performance: area under the precision-recall (AUPR), area under ROC curve (AUC), ability to identify positive samples (RECALL), ability to identify negative samples (PRECISION), average Recall and Precision (F1_SCORE) and prediction accuracy (ACC). In the end, PPDTS obtains drug–target prediction scores. For drugs (or targets), the highest score in the corresponding target (or drug) is recorded as the final prediction results.

3.2 | Experimental model parameters

We set the restart probability of RWR (Random Walk With Restart) to $c = \{0.1, 0.2, 0.3, \ldots, 0.9\}$. When c takes different values, the AUC and AUPR of RWR are shown in the dotted lines in Figure 3. To judge from Figure 3, when $c = 0.1$, the highest performance of RWR is AUPR = 0.335 and AUC = 0.903. In the same way, the propagation probability of the LP algorithm (Label Propagation) is set to $\tau = \{0.1, 0.2, 0.3, \ldots, 0.9\}$. When τ represents different values, the changes AUC and AUPR of LP are shown the solid lines in Figure 3. The best performance of LP method are AUPR = 0.334 and AUC = 0.905 with $\tau = 0.1$.

The weights α of the drug–target network and the target–drug network of the PPDTS algorithm is assigned as = { 0, 0.1, 0.2, 0.3, …, 0.9, 1}. The weight α is to find the influence of drug–target similarity model and the combination degree of target–drug similarity model on DTIs prediction. When α is fluctuating among different values, the changes of AUC and AUPR are shown in Figure S2. The AUC values of PPDTS with different binding weights α are more than 0.91, and the highest value of AUPR is 0.494, which is significantly higher than other methods.

3.3 | Comparison of experimental evaluation indicators

PPDTS is compared with baseline methods, and the comparison of experimental results is shown in Table 2. The AUC and AUPR of PPDTS are 0.922 and 0.494, which are significantly higher than those of baseline methods (ProCF, DTINet, RWR, LP), and the AUC plots are shown in Figure S3. In addition, we also compare F1_score, Recall and Precision evaluation index. It is worth noting that the AUC of RWR, LP and DTINet is significantly lower than that of ProCF, which may be the result of less known drug–target associations, or RWR, LP and DTINet cannot make full use of the adjacent nodes to predict associations of drug–target.

We also analyse the robustness of PPDTS, we randomly removed 10%, 20%, 30%, 40% and 50% of the known drug–target interactions. After fivefold cross validation, the final experimental results are shown in Table 3, and we find that the performance of PPDTS and comparison methods decreases with the reduction of known drug–target interactions, but AUC and AUPR of PPDTS are better than those of comparison methods. The possible reason is that PPDTS is different from traditional methods based on the network similarity. In the experiment, PPDTS only uses the known drug–target interaction and uses a binary network to model the inter

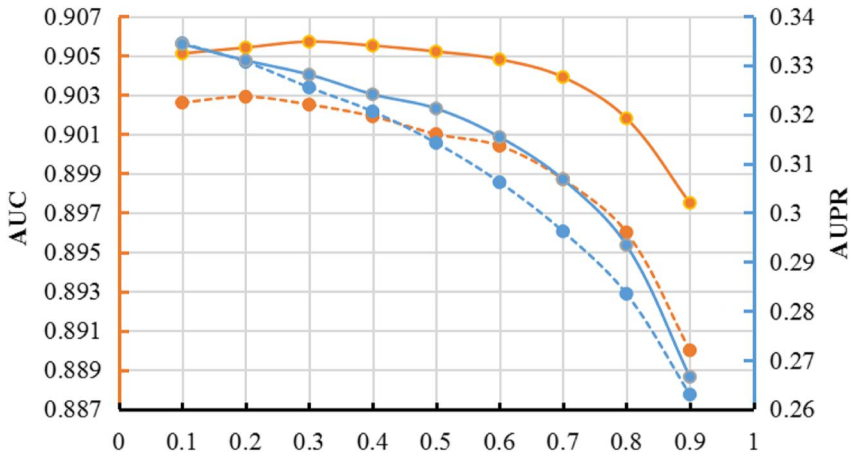

FIGURE 3 The RWR random walk probability c ($0 < c < 1$) and the LP propagation probability τ ($0 < \tau < 1$) are tested with different values, and the two evaluation indicators AUC and AUPR are shown in the figure. The dotted line shows the results of RWR, and the solid line shows the results of LP, and the orange line represents the AUC value with the change of with the abscissa (c or τ), and the blue line represents the AUPR value with the change of with the abscissa (c or τ)

TABLE 2 The experimental results of PPDTS and baseline methods are compared as shown in the table

Methods	AUPR	AUC	F1_SCORE	RECALL	PRECISION	ACC
ProCF	0.319	0.918	0.370	0.376	0.366	0.997
RWR(0.1)	0.335	0.903	0.376	0.371	0.390	0.997
DTINet	0.272	0.908	0.331	0.282	0.401	0.997
LP(0.1)	0.334	0.905	0.364	0.361	0.374	0.997
PPDTS	**0.494**	**0.922**	**0.509**	**0.435**	**0.618**	**0.998**

Note: The evaluation metrics is the area under ROC (receiver operating characteristic curve) curve (AUC), the area under the precision-recall curve (AUPR), how much of positive samples are predicted to be correct (RECALL), how many of the predicted positive sample are actually positive (PRECISION), the harmonic mean of Recall and Precision (F1_SCORE) and accuracy of prediction (ACC), respectively.

TABLE 3 PPDTS and baseline methods robustness test under different proportion data

	AUC					AUPR				
Methods	**50%**	**60%**	**70%**	**80%**	**90%**	**50%**	**60%**	**70%**	**80%**	**90%**
ProCF	0.812	0.839	0.864	0.890	0.905	0.059	0.085	0.116	0.170	0.233
RWR(0.1)	0.815	0.830	0.847	0.869	0.890	0.125	0.131	0.140	0.207	0.283
DTINet	0.825	0.845	0.863	0.882	0.898	0.069	0.094	0.120	0.165	0.214
LP(0.1)	0.817	0.831	0.849	0.871	0.891	0.062	0.085	0.126	0.159	0.245
PPDTS	**0.812**	**0.840**	**0.865**	**0.892**	**0.907**	**0.077**	**0.116**	**0.149**	**0.215**	**0.308**

Note: The evaluation metrics is the area under ROC (receiver operating characteristic curve) curve (AUC), the area under the precision-recall curve (AUPR), respectively.

domain information fusion. Compared with baseline methods, it shows better experimental performance.

3.4 | Independent test set

To further test the performance of PPDTS, we also studied the ability of PPDTS to predict the association between drugs and targets on different datasets. We use Dataset2 as an independent test set. Dataset2 is 8008 drug–target interactions collected from MATADOR database, which contains 575 drugs and 981 targets. The details of Dataset2 are in Section 2.1.

Through fivefold cross validation, the performance evaluation parameters of PPDTS and the baseline methods are reported in Table S1. It can be seen from the evaluation results that AUC of PPDTS is 0.975, which is significantly higher than the results deduced from other methods. The details of AUC are shown in Figure 4. In addition, PPDTS is also superior to other methods in AUPR, F1_SOCRE, RECALL, and PRECISION. All the results show that PPDTS has good performance on the independent test set, which is consistent with the performance of Dataset1.

3.5 | Validation of predicted results

In order to further prove the actual prediction ability of PPDTS, 100 kinds of DTIs were, respectively, selected as test samples from the prediction results of PPDTS and four baseline methods (LP, RWR, DTINet, ProCF), and there are 5 selections in total. After verifying the selected data, the final results are shown in Figure 5, some of which are supported by

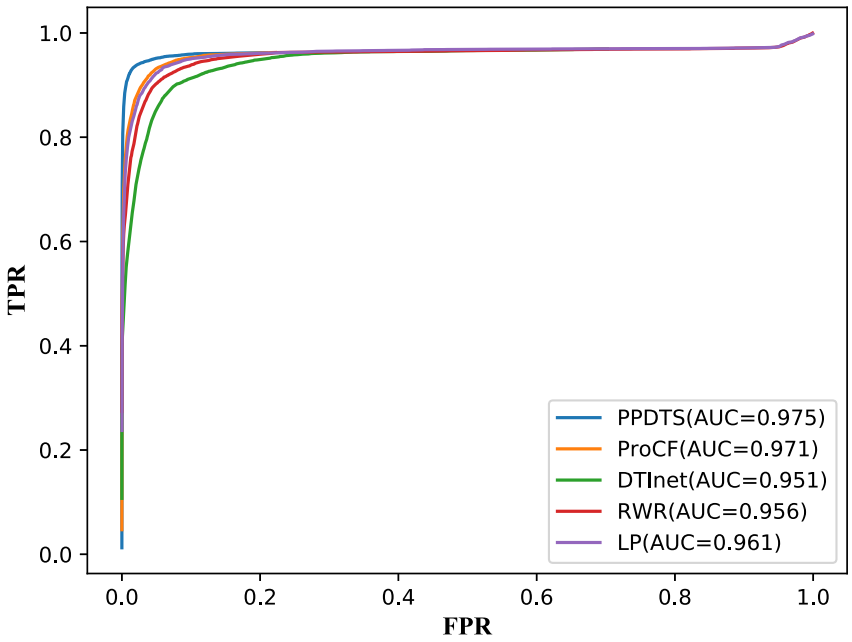

FIGURE 4 AUC curve of PPDTS and baseline methods on independent test set

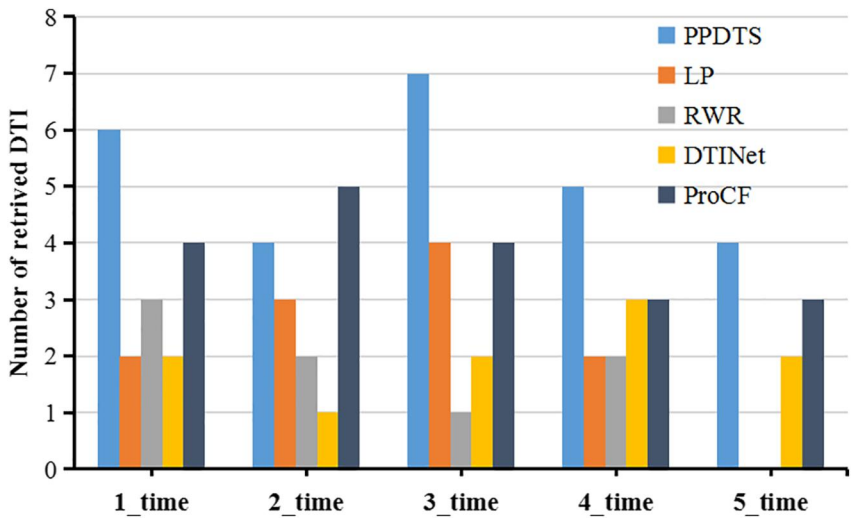

FIGURE 5 The comparison results of random sampling between PPDTS and baseline methods. Each bar represents the number of verified DTIs with the real DTIs, plotted by the sampling frequency on the horizontal axis and the number of verified DTIs on the vertical. At the fifth time random sampling, the number of predicted DTIs for LP and RWR was 0

database literature. Among the five selections, there are four selections of the predictions' results showing that PPDTS was more efficient than baseline methods. This also shows that our proposed PPDTS has an important auxiliary role in drug discovery. The experimental data and results are available from https://github.com/HNUBioinformatics/PPDTS.git.

In order to clarify the reliability of the experimental results, we further verify the prediction results of PPDTS. To dataset1, PPDTS has predicted 786 potential DTIs for each drug, we take the maximum score of targets as the prediction result for each drug in the prediction score matrix. The results are retrieved in DrugBank and UniProt databases, and 37 drug–target interactions are successfully confirmed. There are seven kinds of DTIs that have been proven obvious pharmacological effects in literature, as shown in Figure 6. For example, 49 drugs are known to be associated with the target P07550 (Beta-2 adrenergic receptor). According to the topology information of neighbour nodes of P07550, the prediction result after PPDTS calculation is DB01118 (Amiodarone), and the prediction result is verified by the literature [44]. The Q14500 (ATP-sensitive inward rectifier potassium channel-12) has two known neighbour nodes DB01392 (Yohimbine) and DB00204 (Dofetilide). The prediction result after PPDTS calculation is DB04855 (Dronedarone), which is verified by the literature [45].

In addition, we found that among the seven kinds of drug–target interaction mechanisms in Figure 6, four drugs (DB00201, DB00929, DB01021 and DB01118) produce efficacy by affecting the activity of receptors, and three drugs (DB00907, DB01043 and DB04855) produce efficacy by affecting the activity of sodium and potassium channels, which are consistent with the actual mechanism of efficacy [46]. For example, DB00201 (Caffeine)-P29275 (Adenosine receptor A2b) interaction, whose general function is to affect the G

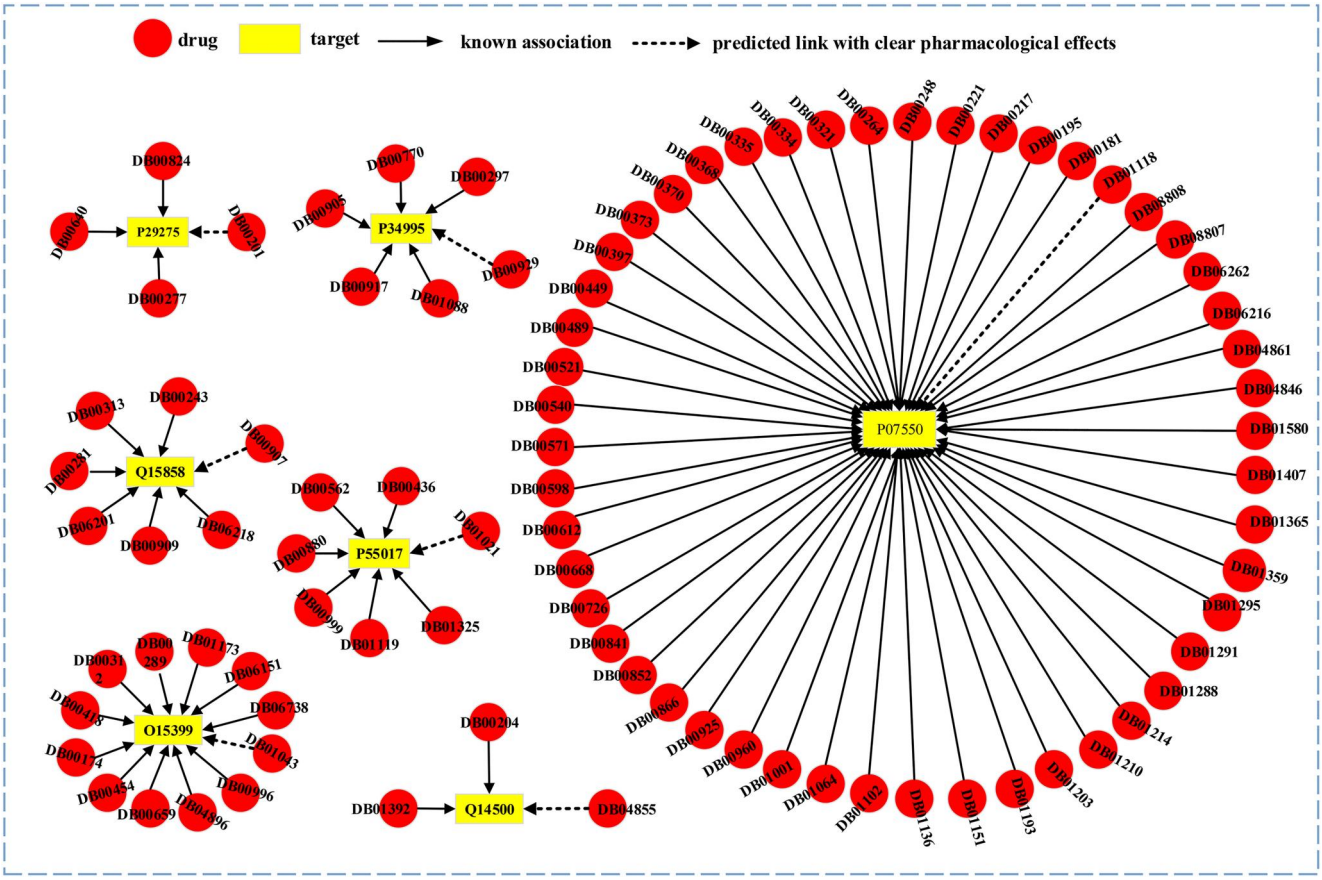

FIGURE 6 PPDTS successfully predicted seven kinds of DTIs with definite pharmacological effects. Red circles represent drugs, and yellow rectangles represent targets. Solid arrows indicate these known drug–target associations, and dotted lines indicate that PPDTS predicts DTI with pharmacological effects

protein-coupled adenosine receptor activity, can be manifested as using caffeine to treat retinopathy of prematurity [47]. DB00929 (Misoprostol)-P34995 (Prostaglandin E2 receptor EP1 subtype) interaction, which is helpful to intervene the activity of prostaglandin *e* receptor, can be manifested as the use of prostaglandin to promote cervical maturation and induce labour [48]. Details of the seven drug–target interaction mechanisms, supporting literature and prediction scores are shown in Table S3. We also validated the remaining 30 DTIs in DrugBank and UniProt databases, among which 29 DTIs are approved to launch studies, but some pharmacological effects are still unclear. In addition, DB01199 (Tubocurarine) and P06276 (Cholinesterase) proved to have no pharmacological effects [49]. There is more information about the 37 DTIs and the Supplement information in the link https://github.com/HNUBioinformatics/PPDTS.git. Therefore, the above results show that PPDTS has a practical reference value for predicting potential drug–target interactions.

4 | CONCLUSION

In this work, we propose a new method called PPDTS. The main contribution of PPDTS is to effectively identify potential drug–target interactions by the innovative integrated Resource Allocation algorithm and Collaborative Filtering algorithm, and the linear combination of drug–target similarity network and target–drug similarity network. After fivefold cross validation, the AUC and AUPR evaluation parameters of PPDTS are proved to be significantly higher than those of other methods, and some potential DTIs are successfully predicted.

PPDTS is an effective prediction method, which can efficiently predict the potential targets of different types of drugs. Since not all the drug–target data are available, PPDTS only used drugs and targets' associations to predict DTIs. In addition, PPDTS starts from the local structure of the network, combining the drug–target similarity network with the target–drug similarity network to predict the potential DTIs. The experimental results have been verified in DrugBank and UniProt databases. However, PPDTS still has some limitations. When it comes to isolated drug–target associations in the network, PPDTS cannot predict such DTIs due to the lack of interactive information. In the future, we will further integrate some other biological data including drugs' side-effect network and target sequence information together [50], so as to make up for the deficiency of PPDTS.

ACKNOWLEDGEMENTS

This research was funded by the National Key Research and Development Program of China (2018YFB1701402), the

Natural Science Foundation of Henan province (No.212300410367, 202300410102), the Science and Technology Research Key Project of Educational Department of Henan Province (No.21A520023), National Natural Science Foundation of China (No.62072160, 62076089, 62072157), Key Project of Science and Technology Department of Henan Province (No.212102310381, 182102210208, 172102210338), Educational Science Research Foundation of Henan Normal University (No.2018JK19), Teaching Reform Research and Practice Project of Henan Normal University (No.201936), Ph.D. Research Startup Foundation of Henan Normal University (No. qd15130), and the National Project Cultivation Fund Project of Henan Normal University (No.2020PL12).

CONFLICT OF INTEREST

The authors declare that they have no commercial or associative conflicts of interest in connection with the work submitted.

DATA AVAILABILITY STATEMENT

The code, dataset, experimental results and supporting materials information used in PPDTS experiment can be found online at https://github.com/HNUBioinformatics/PPDTS.git.

ORCID

Wei Wang https://orcid.org/0000-0002-9616-1145

REFERENCES

1. Lin, C., et al.: Learning to predict drug target interaction from missing not at random labels. IEEE Trans. Nanobiosci. 18(3), 353–359 (2019)
2. Anweiler, C., et al.: Counter-regulatory renin–angiotensin system-based candidate drugs to treat COVID-19 diseases in SARS-COV-2-infected patients. Infect. Disorders Drug Targets 20(4), 407–408 (2020)
3. Shahin, M.H., Giacomini, K.M.: Oral anticoagulants and precision medicine: something old, something new. Clin. Pharmacol. Ther. 107(6), 1273–1277 (2020)
4. Ashburn, T.T., Thor, K.B.: Drug repositioning: identifying and developing new uses for existing drugs. Nat. Rev. Drug Discov. 3(8), 673–683 (2004)
5. Chen, X., et al.: Drug–target interaction prediction: databases, web servers and computational models. Brief. Bioinform. 17(4), 696–712 (2016)
6. Chen, X., et al.: MicroRNA-small molecule association identification: from experimental results to computational models. Brief. Bioinform. 21(1), 47–61 (2020)
7. Xu, R., Wang, Q.: Large-scale extraction of accurate drug–disease treatment pairs from biomedical literature for drug repurposing. BMC Bioinform. 14(1), 181 (2013)
8. Alam, F., et al.: A knowledge-poor approach to chemical-disease relation extraction. Database, 2016, baw071 (2016)
9. Yang, J., et al.: Drug–disease association and drug-repositioning predictions in complex diseases using causal inference-probabilistic matrix factorization. J. Chem. Inf. Model. 54(9), 2562–2569 (2014)
10. Zhao, S., Li, S.: A co-module approach for elucidating drug–disease associations and revealing their molecular basis. Bioinformatics 28(7), 955–961 (2012)
11. Wang, L., et al.: Systematic analysis of new drug indications by drug–gene–disease coherent subnetworks. CPT Pharmacometrics Syst. Pharmacol. 3(11), 1–9 (2014)
12. Yamanishi, Y., et al.: Drug–target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. Bioinformatics. 26(12), i246–i254 (2010)
13. Wang, L., et al.: Computational methods for the prediction of drug–target interactions from drug fingerprints and protein sequences by stacked auto-encoder deep neural network. In: International Symposium on Bioinformatics Research and Applications, vol. 10330, pp. 46–58. (2017)
14. Palma, G., Vidal, M.E., Raschid, L.: Drug–target interaction prediction using semantic similarity and edge partitioning. In: International Semantic Web Conference, vol. 8796, pp. 131–146. (2014)
15. Lee, T., Yoon, Y.: Drug repositioning using drug–disease vectors based on an integrated network. BMC Bioinform. 19(1), 1236–1286 (2018)
16. Li, G., et al.: Predicting microRNA-disease associations using network topological similarity based on deepwalk. IEEE Access 5, 24032–24039 (2017)
17. Chen, X., et al.: Deep-belief network for predicting potential miRNA-disease associations. Brief. Bioinform. 22(3), bbaa186 (2020). https://doi.org/10.1093/bib/bbaa186
18. Qu, J., et al.: Inferring potential small molecule-miRNA association based on triple layer heterogeneous network. J. Cheminform. 10(1), 30:1–30:14 (2018)
19. Zhao, Y., et al.: Snmfsmma: using symmetric nonnegative matrix factorization and kronecker regularized least squares to predict potential small molecule-microRNA association. RNA Biol. 17(2), 281–291 (2020)
20. Wang, C., Chen, X.: A unified framework for the prediction of small molecule-microRNA association based on cross-layer dependency inference on multilayered networks. J. Chem. Inf. Model. 59(12), 5281–5293 (2019)
21. Zhang, W., et al.: A fast linear neighborhood similarity-based network link inference method to predict microRNA-disease associations. IEEE ACM Trans. Comput. Biol. Bioinform. 18(2), 405–415 (2021)
22. Luo, Y., et al.: A network integration approach for drug–target interaction prediction and computational drug repositioning from heterogeneous information. Lect. Notes Comput. Sci. 8(1), 383–384 (2017)
23. Wang, W., et al.: DLS: a link prediction method based on network local structure for predicting drug–protein interactions. Front. Bioeng. Biotechnol. 17(2), 281–291 (2020)
24. Wen, Z., et al.: Predicting drug–disease associations and their therapeutic function based on the drug–disease association bipartite network. Methods 145, 51–59 (2018)
25. Liu, D., et al.: Identifying influential spreaders in large-scale networks based on evidence theory. Neurocomputing 359(24), 466–475 (2019)
26. Zhou, T., Lü, L., Zhang, Y.C.: Predicting missing links via local information. Eur. Phys. J. B 71(4), 623–630 (2009)
27. Wishart, D.S., et al.: Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res. 46(D1), D1074–D1082 (2018)
28. Stefan, P.R.: Supertarget and matador: resources for exploring drug–target relationships. Nucleic Acids Res. 36(1), 919–922 (2008)
29. Wen, M., et al.: Deep-learning-based drug–target interaction prediction. J. Proteome Res. 16(4), 1401–1409 (2017)
30. Tao, Z., et al.: Bipartite network projection and personal recommendation. Phys. Rev. E 76(4), 46115 (2007)
31. Gupta, A.K., Sardana, N.: Prediction of missing links in social networks: feature integration with node neighbour. Int. J. Web Based Commun. 14(1), 38–53 (2018)
32. Ma, C., Cheng, M., Wang, T.: Topic detection based on graph analytical method and cosine similarity. Comput. Eng. ENCE. 41(4), 708–712 (2019)
33. Yan, Z., et al.: Locally private Jaccard similarity estimation. Concurrency Comput. Pract. Ex. 31(24), e4889.1–e4889.15 (2019)
34. Sarwar, B., et al.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web, pp. 285–295. WWW'01.Association for Computing Machinery, New York (2001). https://doi.org/10.1145/371920.372071
35. Herlocker, J.L., et al.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. 22(1), 5–53 (2004)
36. Saadati, M., Shihab, S., Rahman, M.S.: Movie recommender systems: implementation and performance evaluation. CoRR (2019). https://arxiv.org/abs/1909.12749
37. Ge, M., Li, A., Wang, M.: A bipartite network-based method for prediction of long non-coding RNA–protein interactions. Dev Reprod Biol. 14(1), 62–71 (2016). https://europepmc.org/articles/PMC4792848
38. Tong, H., Faloutsos, C., Pan, J.Y.: Fast random walk with restart and its applications. In: Proceedings of the Sixth International Conference on

Data Mining. ICDM'06, pp. 613–622. IEEE Computer Society, USA (2006). https://doi.org/10.1109/ICDM.2006.70
39. Liu, D., et al.: The network representation learning algorithm based on semi-supervised random walk. IEEE Access 8, 222956–222965 (2020)
40. Mingxin, G., Hussein, S.: Walking on a user similarity network towards personalized recommendations. PLoS ONE. 9(12), e114662 (2014)
41. Li, S., et al.: Multi-view graph learning with adaptive label propagation. In: IEEE International Conference on Big Data, pp. 110–115. IEEE Computer Society (2017). https://doi.org/10.1109/BigData.2017.8257918
42. Li, G., et al.: Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity. J. Biomed. Inform. 82(10), 169–177 (2018)
43. Zhang, W., et al.: The linear neighborhood propagation method for predicting long non-coding RNA–protein interactions. Neurocomputing 273(17), 526–534 (2017)
44. Liu, X., et al.: Cause analysis and prevention of organs injury in laparoscopic radical surgery for cervical cancer. Med. J. Wuhan Univ. 38(6), 991–994 (2017)
45. Heijman, J., Heusch, G., Dobrev, D.: Pleiotropic effects of antiarrhythmic agents: dronedarone in the treatment of atrial fibrillation. Clin. Med. Insights Cardiol. 7(7), 127–140 (2013)
46. Dehkordi, N.G., Mirzaei, S.A., Elahian, F.: Pharmacodynamic mechanisms of anti-inflammatory drugs on the chemosensitization of multidrug-resistant cancers and the pharmacogenetics effectiveness. Inflammopharmacology 29(1), 49–74 (2021)
47. Chen, J.F., et al.: Adenosine receptors and caffeine in retinopathy of prematurity. Mol. Asp. Med. 10(55), 118–125 (2017)
48. Pierce, S., et al.: Clinical insights for cervical ripening and labor induction using prostaglandins. AJP Rep. 8(4), 307–314 (2018)
49. Avanzo, D., et al.: Direct and specific activation of human inward rectifier k+ channels by membrane phosphatidylinositol 4,5-bisphosphate. J. Biol. Chem. 285(48), 37129–37132 (2010)
50. Li, J., et al.: A machine learning method for drug combination prediction. Front. Genet. 11(10), 1000 (2020)

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wang, W., et al.: PPDTS: predicting potential drug–target interactions based on network similarity. IET Syst. Biol. 16(1), 18–27 (2022). https://doi.org/10.1049/syb2.12037