A Network Embedding Based Approach to Drug-Target Interaction Prediction Using Additional Implicit Networks

Han Zhang[1], Chengbin Hou[1,2], David McDonald[1], and Shan He[1]

[1] School of Computer Science, University of Birmingham, Birmingham, UK.
[2] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China.
{hxz325,dxm237,s.he}@cs.bham.ac.uk, chengbin.hou10@foxmail.com

Abstract. Identifying novel drug-target interactions (DTIs) is a crucial step in drug discovery. Since experimentally determining DTIs is expensive and time-consuming, it becomes popular to employ computational methods for providing promising candidate DTIs. However, in the existing computational methods, the drug implicit network and target implicit network constructed from a DTI network (a bipartite network) have been ignored in the DTI prediction problem, while such implicit networks constructed from a bipartite network have been proven useful in other problems, e.g., the link prediction task in a bipartite network. Motivated by that, we propose a novel DTI prediction method which considers the implicit networks in addition to drug structure similarity network and target sequence similarity network. The experiments over five real-world DTI datasets demonstrate the competitive performance of the proposed method compared to the state-of-the-art methods. The code is available at https://github.com/BrisksHan/NE-DTIP.

Keywords: Drug-Target Interaction Prediction, Network Embedding, Implicit Networks, Network Topology

1 Introduction

Identifying novel drug-target interactions (DTIs) is a crucial step in drug discovery. Since experimentally determining DTIs is expensive and time-consuming [10], it is desirable to develop computational methods to identify promising candidate DTIs to accelerate the speed of drug discovery.

Over the years, many computational methods have been proposed to identify novel DTIs. The existing methods could be categorised into three groups. The first group is the target structure based methods [15]. These methods simulate the docking process of drugs. However, the 3D structures of the proteins are required as the input, yet, the 3D structures of many proteins are unavailable. The second group is the ligand similarity based methods [14]. These methods use the structural similarity between ligands to predict interactions, however, the sequence similarities between targets are ignored in predicting DTIs. The

third group is the machine learning based methods [2, 16, 17, 19, 26, 28, 33, 34, 36]. They often take the DTI network, drug structural similarity network (DSSN), and target sequence similarity network (TSSN) as the inputs to train a machine learning model for predicting DTIs. Note that, the machine learning based methods have attracted a lot of research interests in recent years. Most of them focus on identifying novel DTIs from known DTIs, as a drug might bind to more than one target [33], which could lead to successful drug repositioning.

Although many computational methods have been proposed to tackle the DTI prediction problem, the implicit networks constructed from the known DTI network are ignored in existing works. However, it has been shown the implicit networks constructed from a bipartite network can improve the performance of the link prediction task in the bipartite network [29]. Besides, the implicit relations can also improve the performance of recommender systems [35].

Motivated by the success of using implicit networks or relations in other problems, we suggest to also consider the *implicit networks* extracted from the DTI network (a bipartite network) in the DTI prediction problem. The implicit networks, i.e., drug implicit network (DIN) and target implicit network (TIN), are constructed using the second-order proximity in the DTI network. It is worth noticing that, an edge in DIN indicates that two drugs would bind to at least one common target and an edge in TIN indicates that two targets would bind to at least one common drug, while DSSN (or TSSN) represents the structural similarity directly calculated between two drugs (or two targets). The topological structures of DIN and TIN would be different from the topological structures of TSSN and DSSN, as they are computed from different perspectives.

To incorporate the implicit networks, we propose a method termed as Network Embedding based Drug-Target Interaction Prediction (NE-DTIP). Specifically, NE-DTIP is a machine learning based method and it includes two stages: the feature vector construction stage and the DTI classification stage. During feature vector construction stage, DIN and TIN are constructed from the DTI network. After that, drug embeddings and target embeddings are learned from DIN, TIN, DSSN, and TSSN using a network embedding method. *Unlike previous methods, the proposed method additionally considers two homogeneous networks, i.e., TIN and DIN, both of which are generated based on the implicit relations of a given DTI network.* During DTI classification stage, the drug-target pairs in DTI network are regarded as positive samples, while randomly sampled unknown drug-target pairs in the DTI network are regarded as negative samples. The feature vector of each training sample (for a drug-target pair) is constructed via concatenating the corresponding drug embeddings and target embeddings. Finally, all samples are used to train a classifier for DTI predictions.

We employ five real-world DTI datasets to evaluate the performance of NE-DTIP. Comparing against four state-of-the-art methods, NE-DTIP outperforms the existing methods on three out of five datasets. We also conduct a case study to verify the top-20 DTI predictions by NE-DTIP on the latest dataset, and it is interesting to find that six out of twenty novel DTIs (i.e., new DTIs not recorded in the dataset) are supported by recent studies.

2 Related Work

The machine learning based DTI prediction methods can be categorised into three groups. The first group is the distance based methods [16, 34, 36]. Those methods embed the drugs and targets into a unified space based on the known DTIs and the similarities between drugs and targets. Then, the novel DTI predictions are made based on the distance between the drugs and targets in the unified space. The second group is the bipartite local prediction based methods [2, 19, 32]. Those methods learn two classifiers. The first classifier predicts targets for a given drug and the second classifier predicts drugs for a given target. Given a drug-target pair, the two classifiers are jointed together to make a prediction. The third group is the feature vector based methods [17, 26, 28]. Those methods construct feature vectors for drugs and targets. Then, the known DTIs are treated as positive samples and unknown DTIs are treated as negative samples. The feature vectors of drug-target pairs are constructed via concatenating the corresponding drugs and targets. Then, a classifier is employed to predict DTIs.

3 Notation and Problem Definition

Homogeneous Network: All of DIN, DSSN, TIN, and TSSN are homogeneous networks, i.e., all nodes of a network are the same type. Let $G = (V, E)$ be a homogeneous network, where V denotes a set of nodes and $E \subseteq V \times V$ denotes edges. The number of nodes in V is denoted with $|V|$. For any node pair (v_i, v_j), where $i \in [1, ..., |V|]$ and $j \in [1, ..., |V|]$, there is a non-negative edge weight w_{ij} which describes the strength of connection between the two nodes. The weight is 0 if two nodes are not connected in E. All the edge weights are represented in a $|V| \times |V|$ matrix $W = [w_{ij}]$. Note that, there is no self-loop in this work, i.e., $w_{ii} = 0 \ \forall i \in [1, ..., |V|]$ in all homogeneous networks.
DTI Network: DTI network is a bipartite network, i.e., nodes are in two sets and there is no edge between nodes in the same set. Let $B = (V^D, V^T, E^{DTI})$ be a DTI network, where V^D denotes a set of drugs, V^T denotes a set of targets and $E^{DTI} \subseteq V^D \times V^T$ denotes edges, as there is no drug-drug edge or target-target edge in the DTI network. For any node pair (v_i^D, v_j^T), where $i \in [1, ..., |V^D|]$ and $j \in [1, ..., |V^T|]$, w_{ij}^{DTI} is used to describes the strength of interaction, and the weight is 0 if there is no interaction between a drug-target pair. All the edge weights are represented in a $|V^D| \times |V^T|$ matrix $W^{DTI} = [w_{ij}^{DTI}]$.
DTI Prediction Problem: The aim of DTI prediction is to infer a DTI prediction matrix $M \subseteq V^D \times V^T$ for all drug-target pairs, given the inputs, i.e., DTI network, DSSN, and TSSN. Each entry in the outputs M should reflect the possibility of the existence of the interaction between a drug-target pair.

4 Method

The proposed method as shown in Fig. 1 includes two stages. The feature vector construction stage is discussed in Section 4.1-4.3. The DTI classification stage is discussed in Section 4.4. And the pseudocode is summarised in Algorithm 1.

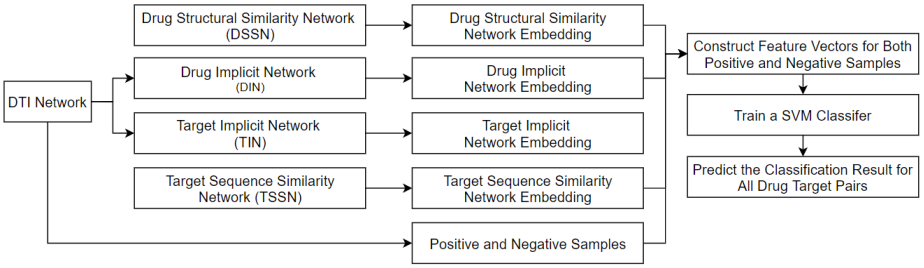

Fig. 1. The framework of the proposed method.

4.1 Network Construction

Construction of Implicit Networks for DIN and TIN. The implicit networks $G^{DIN} = (V^D, E^{DIN})$ and $G^{TIN} = (V^T, E^{TIN})$ are constructed from the DTI network $B = (V^D, V^T, E^{DTI})$. Both of the implicit networks are homogeneous networks, each node represents a drug in G^{DIN} and a target in G^{TIN}, and an edge between two nodes indicates the corresponding nodes share at least one common neighbour in the DTI network. Following the definition in [8], for any node pair in DIN and TIN, the edge weight is defined as:

$$w_{ij}^{DIN} = \sum_{k \in [1,...,|V^T|]} w_{ik}^{DTI} w_{jk}^{DTI} \ \forall i, j \in [1, ..., |V^D|] \text{ and } i \neq j \tag{1}$$

$$w_{ij}^{TIN} = \sum_{k \in [1,...,|V^D|]} w_{ki}^{DTI} w_{kj}^{DTI} \ \forall i, j \in [1, ..., |V^T|] \text{ and } i \neq j \tag{2}$$

where, w_{ij}^{DIN} denotes the edge weight for node pair (v_i^D, v_j^D) in DIN and w_{ij}^{TIN} denotes the edge weight for node pair (v_i^T, v_j^T) in TIN.

Construction for DSSN and TSSN. The edge weight in DSSN is calculated via applying either Tanimoto coefficient [30] or [12] to pairs of drug molecular structures, while the edge weight in TSSN is calculated via applying Smith-Waterman algorithm to pairs of protein sequences [24]. The DSSN and TSSN are fully connected as each node pair would have a positive similarity score as weight. However, many edges are not informative as the corresponding edge weights are very small. By removing edges with small weights, the speed of network embedding process can be significantly improved. Since it is difficult to set cut-off scores for both drug structural similarity score and target sequence similarity score, we introduce an edge density parameter α to remove the non-informative edges, as it has been shown the DSSN shows strong community structures via only keeping the edges with high similarity scores [27]. In both DSSN and TSSN, we rank edges by weights and use the edge density parameter α to denote the proportion of edges to keep, i.e., $\alpha = 0.1$ means only the top 10 percent of edges are kept.

4.2 Network Embedding

A network embedding method is employed to learn low dimensional node embeddings for constructing feature vectors. The reasons are that the rows and columns of the weight matrix of a network only contain the first-order proximity information, and they are sparse and high dimensional. Those factors made the weight matrix sub-optimal for the downstream machine learning task.

The modified DeepWalk [22, 13] is employed to learn the node embeddings, as it has been widely recognised as a simple, efficient, and effective algorithm in the network embedding field [9]. The modified DeepWalk includes three parts: 1) Random walks that capture the topological structure of a network; 2) A sliding window that encodes the node similarity from node sequence into the node pairs D; 3) Skip-gram negative sampling model (SGNS) [20] that learns the node embeddings based on the frequency of node pairs in D and SGNS is employed instead of the original method in DeepWalk to reduce the computational cost.

The modified DeepWalk is applied to DSSN, DIN, TSSN, and TIN to learn drug and target embeddings. The embeddings from the four networks can be denoted as $Z^{DSSN} \in \mathbb{R}^{|V^D| \times d}$, $Z^{DIN} \in \mathbb{R}^{|V^D| \times d}$, $Z^{TSSN} \in \mathbb{R}^{|V^T| \times d}$, and $Z^{TIN} \in \mathbb{R}^{|V^T| \times d}$ respectively, where d is the embedding dimension.

4.3 Feature Vector Construction for Training Samples

The feature vectors of training samples are constructed based on the learned embeddings: Z^{DSSN}, Z^{DIN}, Z^{TSSN}, and Z^{TIN}. The positive samples are the edges from the DTI network. For the $m-th$ drug-target pair $(v_i^D, v_j^T) \in E^{DTI}$, the feature vector of positive samples $Z_m^{positive}$ is constructed by:

$$Z_m^{positive} = Z_i^{DSSN} \oplus Z_i^{DIN} \oplus Z_j^{TSSN} \oplus Z_j^{TIN} \tag{3}$$

For $m-th$ positive sample, there is a corresponding label $y_m^{positive} = 1$. The negative samples are constructed from the unknown pairs in the DTI network as those pairs have not be experimentally validated, hence, they are unlikely to have interactions. For the $h-th$ randomly sampled drug-target pair $(v_i^D, v_j^T) \notin E^{DTI}$, the feature vector $Z_h^{negative}$ can be constructed by:

$$Z_h^{negative} = Z_i^{DSSN} \oplus Z_i^{DIN} \oplus Z_j^{TSSN} \oplus Z_j^{TIN} \tag{4}$$

the corresponding label of the negative sample is $y_h^{negative} = -1$. In the training samples, the ratio of positive samples to negative samples is 1:10.

4.4 DTI Prediction

A support vector machine (SVM) with *Platt scaling* [23] classifier is trained to separate positive and negative samples as SVM has been shown to be accurate

Algorithm 1 NE-DTIP: Network Embedding based DTI Prediction

Input: DTI network; DSSN; TSSN; edge density parameter α; number of walks per node r; walk length l; window size w; number of negative samples q; embedding dimension d; kernel parameter γ for the SVM classifier; tolerance parameter c for training the SVM classifier;

Output: DTI prediction matrix $M \in \mathbb{R}^{|V^D| \times |V^T|}$.

1: Construct DIN and TIN (two implicit networks) from DTI network. The edge weights in DIN and TIN are constructed by Eq (1) and Eq (2) respectively.
2: Keep the top α proportion of edges in DSSN and TSSN.
3: **for** DIN, TIN, DSSN, and TSSN **do**
4: Apply the modified DeepWalk to learn the node embeddings.
5: **End for**
6: Construct feature vectors for positive samples (from known DTIs) and randomly sampled negative samples (from unknown DTIs) as training samples.
7: Train a SVM classifier using training samples, which yields a trained SVM classifier.
8: Construct feature vectors for all drug-target pairs.
9: Feed all feature vectors (from the last step) into the trained SVM classifier to learn the DTI prediction matrix M.
10: **return** DTI prediction matrix M.

and robust [1]. For a drug-target pair, the probability of classifying a feature vector as a DTI edge can be written as:

$$p(y = 1|x) = \frac{1}{1 + \exp(-(\sum_{k=1}^{p} y_k a_k K(x, x_k) + b))} \tag{5}$$

where p is the number of training samples, K indicates the kernel function, $a = (a_1, ..., a_p)$ and b are learnable parameter. The kernel function provides the non-linear ability to classify samples without having to project the features of samples into higher dimensional space. In this work, we use Radial basis function kernel [3] to calculate the distance between samples. The parameters in Eq (5) are *learned* using matrix decomposition via transforming it to a quadratic programming problem [6], in which a tolerance parameter c is introduced.

After all the parameters being learned, we start to *infer* the DTI prediction matrix M. For all drug-target pairs, the feature vectors are constructed as described before. After that, all feature vectors for all drug-target pairs are fed into the trained decision function to infer the DTI prediction matrix M.

5 Experiments

5.1 Datasets

The experiments are conducted on five benchmark datasets. Four of those datasets are published in [34] and they are nuclear receptor (NR), G-protein-coupled receptors (GPCR), ion channel (IC), and enzyme (E). All of those four datasets

Table 1. The statistics of five DTI datasets.

	NR	GPCR	IC	E	DT-IN
# of DTI edges (in a DTI network)	90	635	1476	2926	4978
# of drug nodes (in both DTI and DIN networks)	54	223	210	445	732
# of targets nodes (in both DTI and TIN networks)	26	95	204	664	1915
# of implicit drug-drug edges (in a DIN network)	218	2748	2546	5137	25628
# of implicit target-target edges (in a TIN network)	54	668	8843	15497	46843

are obtained from DrugBank [31]. The fifth dataset drug-target inhibition (DT-IN) is published in [17], in which, the DTIs are obtained from multiple sources, and only the DTIs with the binding threshold below 10μM are kept.

For all the benchmark datasets, DSSNs, and TSSNs are pre-computed. There is a difference in constructing the DSSNs between those datasets. For NR, GPCR, IC, and E, the drug structural similarity scores are constructed using [12], while for DT-IN, the drug structural similarity score is constructed using Tanimoto coefficient. The target sequence similarity scores in all of the five datasets are constructed using Smith-Waterman algorithm [24].

5.2 Baselines

The DTI prediction task can be viewed as classifying an imbalanced datasets. The datasets are imbalanced as there is far more unknown edges than the known edges in the DTI network. Following previous works [17, 28], the metric of evaluating the performance of predicting DTI is the area under precision and recall curve (AUPR), as it is more suitable for imbalanced datasets. Other popular metrics such as the area under receiver operating characteristic curve (AUC) would give an optimistic evaluation of the prediction [7].

Five independent ten-fold cross validations are conducted to evaluate the performance of the DTI prediction methods on each dataset. To calculate the AUPR of the testing edges, the randomly sampled unknown edges are treated as negative samples. Following the experimental setting in NeoDTI [28], the ratio of the testing edges to the randomly sampled unknown edges is 1:10.

5.3 Compared to Other Methods

The compared methods are Netlaprls [32], BLM-NII [19], NRLMF [16], and NeoDTI [28], which have been introduced in related works. Although there are newer methods, those methods take additional networks such as drug-drug interaction and target-target interaction as inputs, yet those additional networks may not available in some datasets. For fairness, the inputs are DTI network, DSSN, and TSSN for all methods. The hyper-parameters of the compared methods are obtained via conducting a grid search on DT-IN using a ten-fold cross validation. The set of hyper-parameters with the highest performance in AUPR is selected to conduct all experiment on all five datasets.

Table 2. The AUPR scores on five benchmark datasets.

	Netlaprls		BLM-NII		NRLMF		NeoDTI		NE-DTIP	
	AUPR	st.d.	AUPR	st.d.	AUPR	st.d.	AUPR	st.d.	AUPR	st.d.
NR	0.2288	0.0486	**0.3617**	0.0984	0.3336	0.0629	0.2308	0.0835	0.2906	0.0992
GPCR	0.4149	0.0358	0.4578	0.0434	**0.4979**	0.0392	0.4966	0.0624	0.4398	0.0628
IC	0.4704	0.0291	0.4763	0.0276	0.5201	0.0278	0.5841	0.0356	**0.6077**	0.0397
E	0.6930	0.0265	0.7299	0.0285	0.7352	0.0294	0.7844	0.0267	**0.7963**	0.0248
DT-IN	0.7816	0.0230	0.8024	0.0231	0.8484	0.0186	0.8560	0.0161	**0.8610**	0.0145

The best results are in bold. The st.d. is the abbreviation for standard deviation.

The experiments of our method are conducted using the following hyper-parameters unless otherwise specified. The edge density hyper-parameter α for DSSN and TSSN is set as 0.1. For the hyper-parameters in the modified DeepWalk, the hyper-parameters for embedding dimension, number of walks, walk length l, window size w, and negative sample number q are set to 128, 10, 80, 10, and 5 respectively, as these parameters have been shown the good performance in most cases according to [13, 22]. For the hyper-parameters in SVM, γ and the tolerance hyper-parameter c are set to $1/(11|E^{DTI}|)$ and 1 respectively, both of which are according to [6].

The result is shown in Table 2, from which, we have two observations. First, our method shows very competitive performance comparing to other methods, as our method consistently outperforms other methods over the three largest datasets (see Table 1 for the size). Second, the performance of our method is affected by *implicit relations* between drugs and targets. It can be seen from Table 1, there are far more edges in implicit networks constructed from IC, E, DT-IN than that of the implicit networks constructed from NR and GPCR. Due to the sparseness of implicit networks in NR and GPCR, many nodes are isolated in the corresponding implicit networks. As a result, the embeddings learned from the DIN and TIN with less implicit relations (in NR and GPCR datasets) are less informative, which degrades the performance of our method.

5.4 The Effect of Implicit Networks

To evaluate the performance gain by adding the implicit networks in predicting DTIs, we investigate the DTI prediction performance by using node embeddings from DSSN and TSSN to construct to feature vectors and this method is named as DSSN+TSSN. The experiments are conducted on DT-IN with five independent ten-fold validations. The embedding dimension d of the DSSN+TSSN is set as 128 and 256, while the embedding dimension d of NE-DTIP is set as 128 as previous. The reason for the additional dimension setting $d = 256$ for DSSN+TSSN is to make the dimensions of feature vectors of DSSN+TSSN and NE-DTIP to be the same. The experimental result is draw in Table 3. From it, we can conclude there is a performance gain by adding implicit networks, as NE-DTIP significantly outperforms DSSN+TSSN in both experimental settings.

Table 3. The effect of implicit networks.

DSSN+TSSN d=128		DSSN+TSSN d=256		NE-DTIP	
AUPR	st.d.	AUPR	st.d.	AUPR	st.d.
0.8075	0.0171	0.8063	0.0172	0.8610	0.0145

It demonstrates the effectiveness of the proposed method by additionally incorporating the implicit networks. The reason is that the topological structures of DIN and TIN are different from that of the DSSN and TSSN. As a result, the node embeddings from DIN and TIN could provide additional useful features for the prediction model to improve its performance.

5.5 Parameter Sensitivity

For the proposed method, an important hyper-parameter α is investigated. For the parameter sensitivity analysis, α is set to 0.01, 0.04, 0.1, 0.4, and 1.0 with other parameters fixed. The effect to the DTI prediction is shown in Table 4. It can be seen that the proposed method performs the best with $\alpha = 0.04$ and $\alpha = 0.1$, while there is a performance loss if α is set to a too low or too high value. The reason for the performance loss of setting $\alpha = 0.01$ is that many informative relations are removed. However, the reason for the performance loss of setting $\alpha = 0.4$ and 1.0 is that most uninformative relations are kept. By keeping those uninformative edges, node pairs with very small similarity scores could co-exist in the training node pair set D. As a result, the distance between the node embeddings do not reflect the structure/sequence similarity between the nodes. This is further verified by an additional performance loss in setting $\alpha = 1$ compared to setting $\alpha = 0.4$.

5.6 A Case Study of Novel DTI Prediction

We conduct a case study on DT-IN dataset to find whether the novel DTIs (excluding the known DTIs recorded in the dataset) predicted by our method can be supported by recent studies. In this experiment, all the known DTIs are first used to train our model. The trained model is then used to predict novel DTIs. The top-100 novel DTIs predicted by our method are shown in Fig. 2.

We search for the supporting studies over the top-20 predicted DTIs. It is interesting to find that six of them are supported by recent studies. Sunitinib inhibits EPHB2 [18], GSK3B [5], and SYK [21]. The treatment of using Sunitinib substantially increases ErbB3 [11]. Bosutinib is an inhibitors to KDR [4]. Haloperidol would down regulate CHRM2 [25].

Table 4. The effect of edge density parameter α.

α=0.01		α=0.04		α=0.1		α=0.4		α=1	
AUPR	st.d.	AUPR	st.d.	AUPR	st.d.	AUPR	st.d.	AUPR	st.d.
0.8547	0.0140	0.8608	0.0153	0.8610	0.0145	0.8536	0.0151	0.8491	0.0159

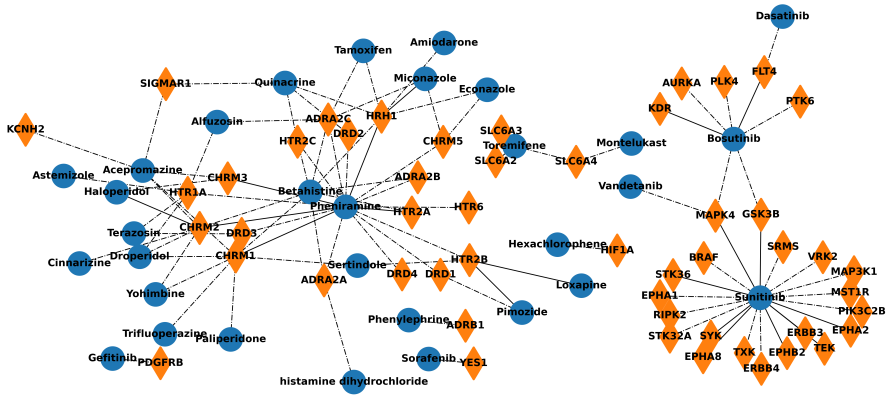

Fig. 2. The top-100 DTIs predicted by the proposed method on DT-IN dataset. Each circle represents a drug. Each diamond represents a target. The edges between nodes indicate novel DTI predictions (i.e., new DTIs not recorded in the dataset). The top-20 DTI predictions are in solid lines, while the remaining ones are in dashed lines.

6 Conclusion

In this work, we propose a novel DTI prediction method, namely NE-DTIP, by additionally considering the implicit networks. Experiments over five benchmark datasets demonstrate the competitive performance of the proposed method compared to other state-of-the-art methods, especially on the datasets with densely connected implicit networks. Further experiments suggest that there is a significant performance gain in incorporating the implicit networks, which however, are ignored in the previous related works. Finally, a case study indicates that the proposed method is capable of predicting novel DTIs.

The future work can be conducted from the following directions. The first direction is to increase the robustness of utilising the implicit relations. If the implicit network is too sparse, the features generated by network embedding module of NE-DTIP would not be informative. As a result, the performance of DTI prediction would be degraded. The second direction is to further incorporate other networks such as drug-drug interaction network, target-target interaction network and drug-side-effect network, since these networks with the more information may increase the performance of DTI prediction.

References

1. Archana, S., Elangovan, K.: Survey of classification techniques in data mining. International Journal of Computer Science and Mobile Applications **2**(2), 65–71 (2014)
2. Bleakley, K., Yamanishi, Y.: Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics **25**(18), 2397–2403 (2009)
3. Broomhead, D.S., Lowe, D.: Radial basis functions, multi-variable functional interpolation and adaptive networks. Tech. rep., Royal Signals and Radar Establishment Malvern (United Kingdom) (1988)

4. Brown, S.A., Nhola, L., Herrmann, J.: Cardiovascular toxicities of small molecule tyrosine kinase inhibitors: an opportunity for systems-based approaches. Clinical Pharmacology & Therapeutics **101**(1), 65–80 (2017)
5. Calero, R., Morchon, E., Johnsen, J.I., Serrano, R.: Sunitinib suppress neuroblastoma growth through degradation of mycn and inhibition of angiogenesis. PloS one **9**(4), e95628 (2014)
6. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST) **2**(3), 1–27 (2011)
7. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proceedings of the 23rd ICML. pp. 233–240 (2006)
8. Deng, H., Lyu, M.R., King, I.: A generalized co-hits algorithm and its application to bipartite graphs. In: Proceedings of the 15th ACM SIGKDD. pp. 239–248 (2009)
9. Goyal, P., Ferrara, E.: Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems **151**, 78–94 (2018)
10. Haggarty, S.J., Koeller, K.M., Wong, J.C., Butcher, R.A., Schreiber, S.L.: Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. Chemistry & biology **10**(5), 383–396 (2003)
11. Harvey, P.A., Leinwand, L.A.: Oestrogen enhances cardiotoxicity induced by sunitinib by regulation of drug transport and metabolism. Cardiovascular research **107**(1), 66–77 (2015)
12. Hattori, M., Okuno, Y., Goto, S., Kanehisa, M.: Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. Journal of the American Chemical Society **125**(39), 11853–11865 (2003)
13. Hou, C., Zhang, H., He, S., Tang, K.: Glodyne: Global topology preserving dynamic network embedding. IEEE Transactions on Knowledge and Data Engineering (2020)
14. Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J., Shoichet, B.K.: Relating protein pharmacology by ligand chemistry. Nature biotechnology **25**(2), 197–206 (2007)
15. Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B., et al.: Predicting new molecular targets for known drugs. Nature **462**(7270), 175–181 (2009)
16. Liu, Y., Wu, M., Miao, C., Zhao, P., Li, X.L.: Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLoS computational biology **12**(2), e1004760 (2016)
17. Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, J., Chen, L., Zeng, J.: A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature communications **8**(1), 1–13 (2017)
18. Martinho, O., Silva-Oliveira, R., Miranda-Goncalves, V., Clara, C., Almeida, J.R., Carvalho, A.L., Barata, J.T., Reis, R.M.: In vitro and in vivo analysis of rtk inhibitor efficacy and identification of its novel targets in glioblastomas. Translational oncology **6**(2), 187–IN20 (2013)
19. Mei, J.P., Kwoh, C.K., Yang, P., Li, X.L., Zheng, J.: Drug–target interaction prediction by learning from local information and neighbors. Bioinformatics **29**(2), 238–245 (2013)
20. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. NeurIPS **26**, 3111–3119 (2013)

21. Noé, G., Bellesoeur, A., Thomas-Schoemann, A., Rangarajan, S., Naji, F., Puszkiel, A., Huillard, O., Saidu, N., Golmard, L., Alexandre, J., et al.: Clinical and kinomic analysis identifies peripheral blood mononuclear cells as a potential pharmacodynamic biomarker in metastatic renal cell carcinoma patients treated with sunitinib. Oncotarget **7**(41), 67507 (2016)
22. Perozzi, B., Al-Rfou, R., Skiena, S.: DeepWalk: online learning of social representations. In: ACM SIGKDD. pp. 701–710 (2014)
23. Platt, J., et al.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers **10**(3), 61–74 (1999)
24. Smith, T.F., Waterman, M.S., et al.: Identification of common molecular subsequences. Journal of molecular biology **147**(1), 195–197 (1981)
25. Swathy, B., Banerjee, M.: Haloperidol induces pharmacoepigenetic response by modulating mirna expression, global dna methylation and expression profiles of methylation maintenance genes and genes involved in neurotransmission in neuronal cells. PloS one **12**(9), e0184209 (2017)
26. Thafar, M.A., Olayan, R.S., Ashoor, H., Albaradei, S., Bajic, V.B., Gao, X., Gojobori, T., Essack, M.: Dtigems+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques. Journal of Cheminformatics **12**(1), 1–17 (2020)
27. Vogt, M., Stumpfe, D., Maggiora, G.M., Bajorath, J.: Lessons learned from the design of chemical space networks and opportunities for new applications. Journal of computer-aided molecular design **30**(3), 191–208 (2016)
28. Wan, F., Hong, L., Xiao, A., Jiang, T., Zeng, J.: Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. Bioinformatics **35**(1), 104–111 (2019)
29. Wang, Y., Jiao, P., Wang, W., Lu, C., Liu, H., Wang, B.: Bipartite network embedding via effective integration of explicit and implicit relations. In: International Conference on Database Systems for Advanced Applications. pp. 435–451 (2019)
30. Willett, P., Barnard, J.M., Downs, G.M.: Chemical similarity searching. Journal of chemical information and computer sciences **38**(6), 983–996 (1998)
31. Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., Woolsey, J.: Drugbank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research **34**(suppl_1), D668–D672 (2006)
32. Xia, Z., Wu, L.Y., Zhou, X., Wong, S.T.: Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. In: BMC systems biology. vol. 4, p. S6. Springer (2010)
33. Xie, L., Xie, L., Kinnings, S.L., Bourne, P.E.: Novel computational approaches to polypharmacology as a means to define responses to individual drugs. Annual review of pharmacology and toxicology **52**, 361–379 (2012)
34. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M.: Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. Bioinformatics **24**(13), i232–i240 (2008)
35. Yu, L., Zhang, C., Pei, S., Sun, G., Zhang, X.: Walkranker: A unified pairwise ranking model with multiple relations for item recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
36. Zong, N., Kim, H., Ngo, V., Harismendy, O.: Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations. Bioinformatics **33**(15), 2337–2344 (2017)