

Article

Adverse Drug Reaction Predictions Using Stacking Deep Heterogeneous Information Network Embedding Approach

Baofang Hu ^{1,2,3} , Hong Wang ^{1,3,*} , Lutong Wang ^{1,3} and Weihua Yuan ^{1,3}

¹ School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China; hubaofang@sdwu.edu.cn (B.H.); wanglutong1002@163.com (L.W.); weihuayuan_qingdao@126.com (W.Y.)

² School of Data and Computer Science, Shandong Women's University, Jinan 250014, China

³ Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, Shandong Normal University, Jinan 250014, China

* Correspondence: wanghong106@163.com; Tel.: +86-531-8961-0769

Academic Editor: Xiangxiang Zeng

Received: 5 November 2018; Accepted: 30 November 2018; Published: 4 December 2018



Abstract: Inferring potential adverse drug reactions is an important and challenging task for the drug discovery and healthcare industry. Many previous studies in computational pharmacology have proposed utilizing multi-source drug information to predict drug side effects and have achieved initial success. However, most of the prediction methods mainly rely on direct similarities inferred from drug information and cannot fully utilize the drug information about the impact of protein–protein interactions (PPI) on potential drug targets. Moreover, most of the methods are designed for specific tasks. In this work, we propose a novel heterogeneous network embedding approach for learning drug representations called SDHINE, which integrates PPI information into drug embeddings and is generic for different adverse drug reaction (ADR) prediction tasks. To integrate heterogeneous drug information and learn drug representations, we first design different meta-path-based proximities to calculate drug similarities, especially target propagation meta-path-based proximity based on PPI network, and then construct a semi-supervised stacking deep neural network model that is jointly optimized by the defined meta-path proximities. Extensive experiments with three state-of-the-art network embedding methods on three ADR prediction tasks demonstrate the effectiveness of the SDHINE model. Furthermore, we compare the drug representations in terms of drug differentiation by mapping the representations into 2D space; the results show that the performance of our approach is superior to that of the comparison methods.

Keywords: adverse drug reaction prediction; heterogeneous information network embedding; stacking denoising auto-encoder; meta-path-based proximity

1. Introduction

Adverse drug reactions (ADRs) are side effects caused by the use of one or several drugs. Some ADRs may be part of the natural pharmacological action of a drug that cannot be avoided, but more often, they may be unpredictable at the development stage. ADRs have caused a global and substantial burden that accounts for considerable mortality and morbidity [1]. Before clinical application of a drug, it should go through two ADR detection stages, including preclinical in vitro safety profiling and clinical drug safety trials. However, since so many side effect types and drug combinations exist, many potential side effects cannot be detected during the early drug development stage [2].

Recently, with the development of data mining and computational prediction methods, researchers have collected extensive drug data from the literature, and reports and have utilized

these data to predict unknown ADRs [3–6]. ADR predictions based on computational methods can point drug safety tests in the right direction and consequently shorten the time requirement and save financial costs during drug development. A large number of machine learning methods have been proposed to predict potential ADRs [7–9]. Vilar et al. [10] utilized known side effect information of drugs to construct an associated matrix of drugs and adverse effects and adopted the matrix completion method to predict unknown side effects. LaBrute et al. [11] processed multi-source drug target information to find association relationships between ADRs and drug targets. These prediction methods are based on single drug information, and these mined drug datasets usually contain much noise. For example, the SIDER dataset [5], which was extracted from the public ADR reports, may contain some fake or unconfirmed noise data. Researchers have established different drug databases that describe drug features from different aspects, including chemical, biological, phenotypic, and interaction relationships [4–6,12,13]. It is more logical to combine different drug information to reduce the prediction error. Integrating this useful complex drug information to obtain more accurate ADR predictions is more effective. Yamanishi et al. [7] used multi-source drug data from the SIDER, PubChem, DrugBank, and Matador databases to predict side effects. The prediction method they adopted was based on multiple kernel regression and canonical correlation analysis. Zhang et al. [14] integrated different drug information to calculate drug similarities and utilized the linear neighborhoods method to transform the similarities into the side effect space and predict side effects. These prediction methods are mainly for the side effects caused by a single drug. However, in real life, many patients, especially the elderly, are on multiple prescriptions to treat different diseases. Drug–drug interactions (DDIs) may change the effects of drugs and cause some potential ADRs. Therefore, predicting the potential side effects induced by DDIs is imperative. Segura-Bedmar et al. [15] utilized a text mining method to predict the occurrence of DDIs based on a shallow language learning model. Jin et al. [16] formulated the DDI type prediction problem as a multi-task dyadic regression problem and utilized the model to predict the side effect types induced by DDIs. Zhang et al. [17] collected a variety of drug data that might influence DDIs and adopted an ensemble learning method to predict the occurrence of DDIs. Motivated by the success of deep learning in many areas, Zitnik et al. [18] developed a new graph convolutional neural network for multi-relational link prediction in multimodal networks to predict the DDI types.

Although the above methods have achieved great success, the methods are mostly designed for specific ADR prediction tasks and lack generic abilities. With the development of the network embedding, learning combined characteristic embeddings of drugs has attracted great attention from researchers [19–21]. Every drug can be embedded into a low-dimensional feature vector, which integrates different drug information, including chemical, biological, phenotypic, and interaction relationships. The drug representations are more general and can be used for different ADR prediction tasks. Li et al. [22] proposed a matrix completion method to integrate multiple sources of drug data and predicted ADRs. Ma et al. [23] proposed a drug embedding method based on multi-view deep auto-encoders to predict ADRs. However, their works only considered immediately relevant information of side effects and neglected potential indirect information. There are some potential association relationships between different biological data. For example, drug targets propagate to another protein through the protein–protein interactions (PPI) network, because the biological function signal cascade propagates through different proteins via PPI [24,25]. When one drug acts on a known target protein, it may change another potential target protein through protein–protein interaction effects and consequently cause potential ADRs.

In this work, we propose a general drug embedding method to learn the representations of drugs and predict different types of ADRs. The flowchart for ADR prediction is shown in Figure 1. We firstly modeled different drug information in a drug heterogeneous information network (drug HIN) framework and then proposed a stacking deep heterogeneous information network embedding approach based on semantic meta-paths. The generated drug embedding integrates multi-source drug information and multi-relationship side effect information to improve the ADR prediction accuracy.

Especially, we utilized the target propagation strategy to recognize the potential drug targets and improve the prediction accuracy. At the target propagation stage, we need to search for the proteins that are more obviously affected by the known targets of the drug. Finding the nearest node of one node is challenging because tens of thousands of nodes exist in the PPI network. We propose using transition probability based on the random walk procedure [26,27], which is common in recommendation systems, to calculate the target propagation proximity and reconstruct the drug-target network based on the target propagation meta-path.

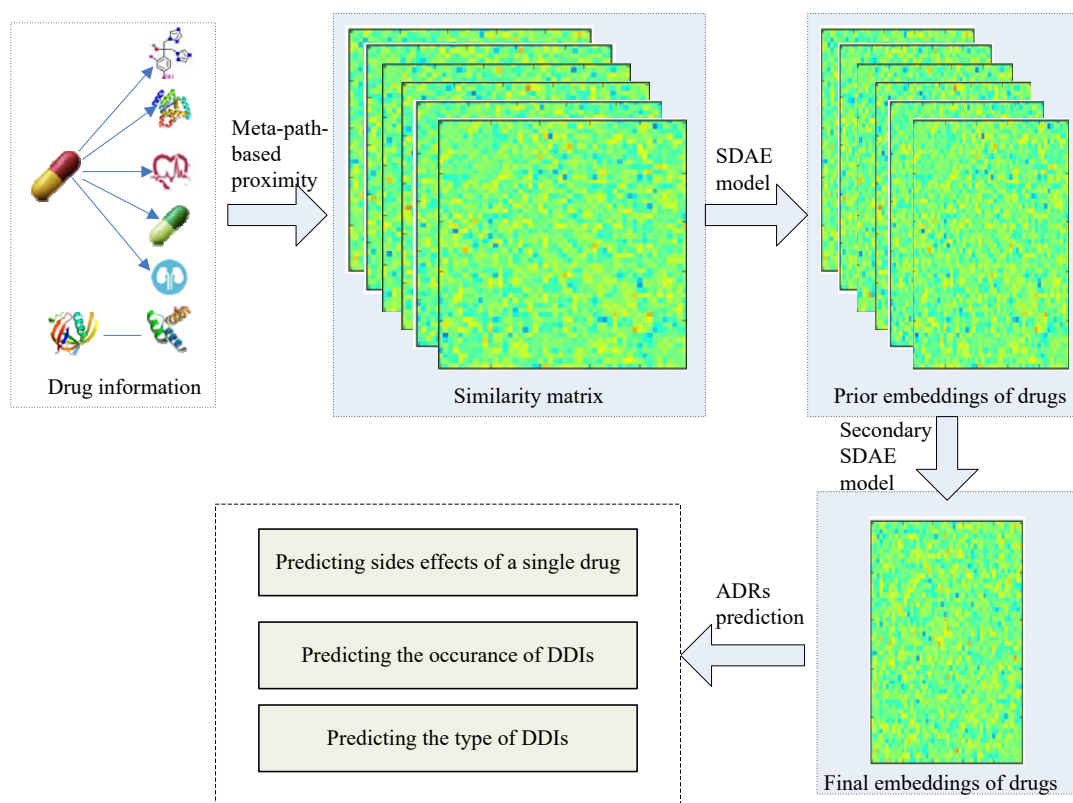


Figure 1. The flowchart for adverse drug reactions' (ADRs) prediction. SDAE, stacked denoising auto-encoder; DDIs, Drug-Drug interaction.

2. Datasets and Method

2.1. Datasets

We collected six types of drug data from seven public databases.

- Drug-Drug interaction information (DDI): Tatonetti [12] mined side effects induced by DDIs from the FDA Adverse Event Reporting System (FAERS, <http://www.fda.gov/cder/aers/default.htm>) and developed a database called "TWO SIDES". The database contains 645 drugs and ADRs caused by 63,473 combinations of different drugs.
- Protein-Protein interaction data (PPI): We downloaded the PPI network data from the Human Protein Reference Database (HPRD, <http://www.hprd.org>). The dataset contains 9519 proteins and 37,062 protein-protein interactions.
- Other drug information: We also obtained other drug information from four online drug information databases (DrugBank [4], the PubChem Compound database [6], the SIDER database [5], and the OFFSIDES database [12]). DrugBank is a widely-used public drug information database. From the DrugBank database, we collected drug target protein and disease treatment information. The PubChem system generates a binary substructure fingerprint

for chemical structures. From the PubChem database, we searched every drug's chemical substructure. We also extracted drug side effect information from the SIDER and OFFSIDES databases. These two databases include most associations between drugs and side effects, and we integrated the drug-side effect data obtained from the two databases.

We mapped drug ids in the TWOSIDES dataset to the other aforementioned datasets and finally constructed an integrated dataset that contained multi-source drug information, including DDIs, drug chemical substructures, drug targets, drug side effects, drug treatment, and PPI. The dataset used in this work is shown in Table 1. In this article, we did not consider the probabilities of ADR events. If a type of ADR occurred, the corresponding element in the DDI dataset or side effect dataset was labeled 1.

Table 1. Description of the drug data. PPI, protein–protein interaction.

Data Type	Data	Data Source	Dimension
Chemical	Substructures	PubChem	548 × 881
Biological	Target protein	DrugBank	548 × 695
Phenotypic	Treatment disease	DrugBank	548 × 718
Phenotypic	Side effect	SIDER, OFFSIDES	548 × 1318 (1318 ADR events)
Interaction	DDIs	TWOSIDES	548 × 548 × 1318 (1318 ADR events)
Interaction	PPI	HPRD	9519 × 9519 (37,062 interactions)

2.2. Drug HIN

Multi-source drug information describes different aspects of drugs and forms a typical heterogeneous information network (HIN). An HIN is a network that contains multiple types of objects or multiple types of relationships [28]. The drug HIN consists of five types of objects: drug (D), chemical substructure (C), protein (P), side effect (S_E), and disease (D_I). The five types of objects are connected through six types of links (as shown in Figure 2). A drug–drug link indicates a type of drug–drug interaction, whereas the link between a drug and its chemical substructure indicates that the drug consists of some type of chemical substructure. In Table 2, we present the semantics of the different link types in the drug HIN.

Table 2. Semantics of link types in the drug heterogeneous information network (HIN).

Link Types	Abbreviated Form	Semantics of Link Types
Drug–Drug	$D-D$	Drug–drug interactions
Drug–Chemical	$D-C$	The chemical substructure of a drug
Drug–Protein	$D-P$	The target protein of a drug
Protein–Protein	$P-P$	Protein–protein interactions
Drug–Disease	$D-D_I$	The therapeutic effect between a drug and a disease
Drug–Side Effect	$D-S_E$	The side effect between a drug and a disease

In HINs, two objects connect via different link types, which are called semantics meta-paths [29,30]. Given an HIN, a meta-path is a sequence of objects connected by different link types. Different types of meta-paths in the drug HIN are shown in Figure 3. Because our final goal is to learn drug representations, we only consider the meta-paths in which the starting objects are all drugs. The detailed meta-paths used in this study are summarized in Table 3.

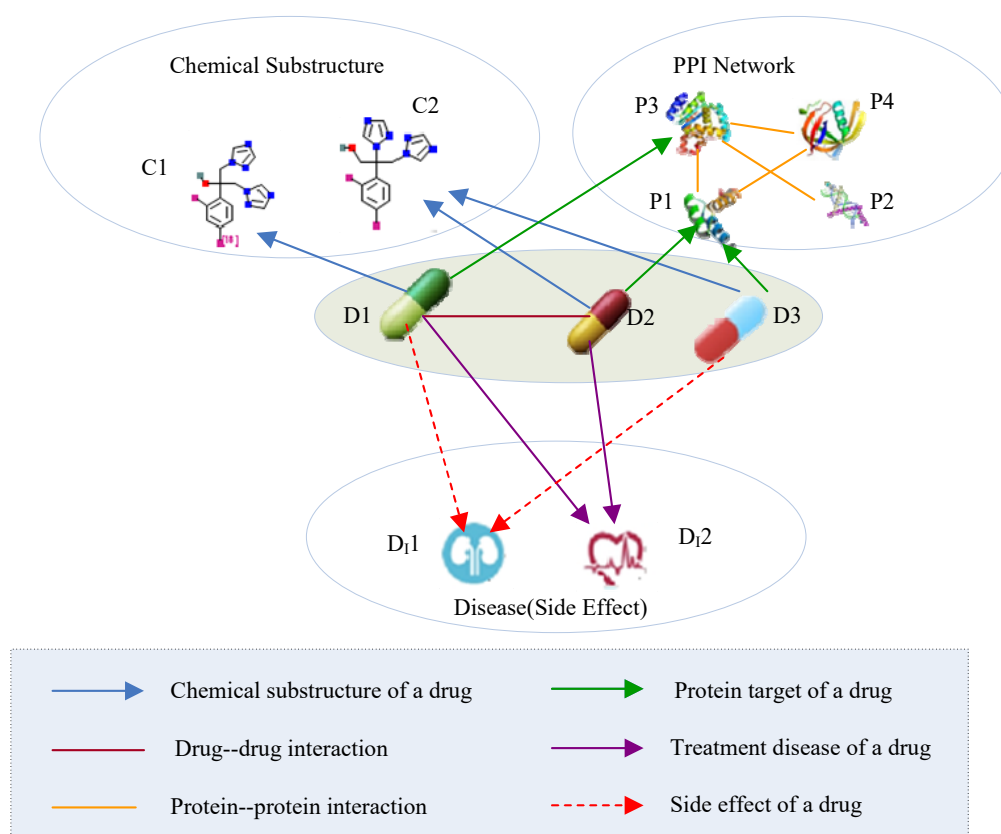


Figure 2. Heterogeneous drug information. PPI, protein-protein interaction.

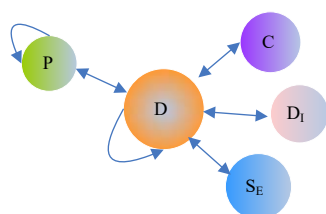


Figure 3. Meta-paths in drug HIN.

Table 3. Meta-paths in drug HIN.

Meta-Paths	Abbreviated Form	Semantics of Meta-Paths
Drug-Drug	DD	Drug-Drug interactions (at the drug embedding stage, interaction types are not considered).
Drug-Chemical-Drug	DCD	Two drugs have a similar chemical substructure.
Drug-Protein-Drug	DPD	Two drugs have the same target protein.
Drug-Protein-...-Protein-Drug	$DP^{(n)}D$ ($n \geq 2$)	There are protein-protein interactions between the targets of two drugs. For example, the path $D_1P_1P_2D_2$ in Figure 4 indicates that the targets of D_1 and D_2 are P_1 and P_2 , respectively. Meanwhile, there is an interaction between P_1 and P_2 (in meta-path $DP^{(n)}D$, there are $n - 1$ protein-protein interactions).
Drug-Disease-Drug	DD_1D	Two drugs have the same therapeutic effect.
Drug-Side Effect-Drug	$DS_E D$	Two drugs have the same side effect.

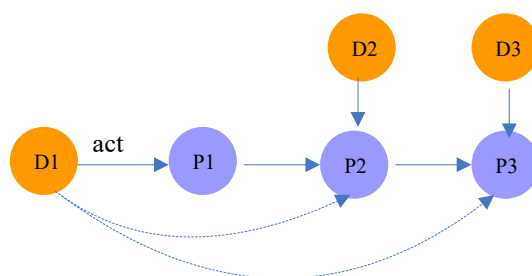


Figure 4. Illustration of meta-path $DP^{(n)}D$.

2.3. Stacking Deep HIN Embedding

Our goal is to learn the low-dimensional vector representations of drugs that highly summarize the drug information and preserve the original proximity of drugs in different drug relationships, and then to predict the different types of ADRs. In this work, we proposed a semi-supervised deep model SDHINE to perform HIN embedding; the framework of the model is shown in Figure 5. In detail, first, we defined the meta-path-based proximities and constructed several homogeneous sub-networks based on the defined proximities. Then, we adopted semi-supervised stacked denoising auto-encoders (SDAE) to encode each sub-network. The supervision information is the meta-path-based proximity in every sub-network. Next, we concatenated the drug embeddings together and further learned the final drug embeddings through the secondary encoding process.

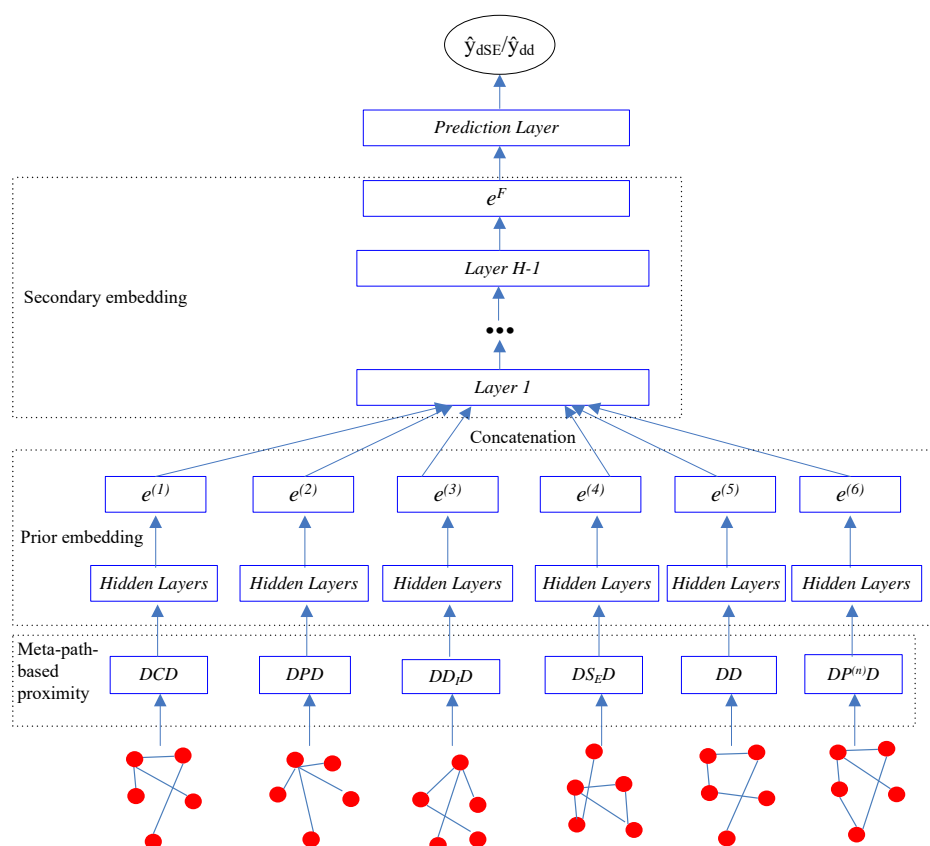


Figure 5. The framework of our proposed semi-supervised deep model SDHINE.

2.3.1. Meta-Path-Based Proximity

We defined three types of meta-path-based proximities and constructed corresponding sub-networks based on the defined proximities.

- Constructing drug-drug interaction sub-network.

For the Drug-Drug (*DD*) meta-path shown in Table 3, in which the nodes are all drug nodes, we utilized the Jaccard similarity as the edge weights to construct the Drug-Drug interaction sub-network. Notably, we do not consider the interaction types at the drug embedding stage. The proximity between drug *i* and drug *j* based on meta-path *DD* can be calculated as shown in Equation (1).

$$s(i, j) = \frac{|D_i \cap D_j|}{|D_i \cup D_j|} \quad (1)$$

where D_i is a column vector with 0 and 1 elements that represent the Drug-Drug interactions between the *i*th drug and other drugs.

- Constructing sub-networks using PathSim proximity.

The meta-paths Drug-Chemical-Drug (*DCD*), Drug-Protein-Drug (*DPD*), DD_1D , and DS_ED in drug HIN contain different types of nodes and path semantic information. For example, meta-path $D_1C_1D_2$ indicates that drug D_1 and drug D_2 have the same chemical substructure C_1 and there is a path from D_1 to D_2 via C_1 . Therefore, we adopted PathSim [29] as the proximity measure in these meta-paths. The PathSim proximity $S(i, j)$ is defined in Equation (2).

$$s(i, j) = \frac{2 \times |\{p_{i \rightarrow j} : p_{i \rightarrow j} \in P\}|}{|\{p_{i \rightarrow i} : p_{i \rightarrow i} \in P\}| + |\{p_{j \rightarrow j} : p_{j \rightarrow j} \in P\}|} \quad (2)$$

where $p_{i \rightarrow j}$ is a path instance between *i* and *j*, $p_{i \rightarrow i}$ is a path instance between *i* and *i*, and $p_{j \rightarrow j}$ is a path instance between *j* and *j*.

The proximities of drugs under meta-paths *DCD*, *DPD*, DD_1D , and DS_ED (as shown in Table 3) are directly calculated using PathSim. Then, we constructed corresponding sub-networks using the proximities as the edge weights to form corresponding sub-networks.

- Reconstructing the drug-target sub-network using the target propagation method.

One innovation of our proposed approach is calculating target protein transition probabilities based on the PPI network and reconstructing the drug-target sub-network. As previously mentioned, potential association relationships exist between different biological data, especially for drug target information. When one target protein is activated by a drug, another potential protein may be activated by protein-protein interactions and consequently cause an unreported ADR. Therefore, we should reconstruct the drug-target sub-network using the target propagation strategy according to the meta-path $DP^{(n)}D$ ($n \geq 2$). The target propagation in the meta-path $DP^{(k)}D$ can be seen as a random walk procedure. A walker walks on the PPI network and achieves the destination protein via $k - 1$ steps. Suppose node transition probabilities in the PPI network converge after *n* steps. The global proximity based on meta-paths $DP^{(n)}D$ ($n \geq 2$) is:

$$s(i, j) = \sum_{k=2}^n S^{(k)}(i, j) = \sum_{k=2}^n (DP^{(k)}D)_{ij} \quad (3)$$

In random walk theory, the *k*-step random walk transition probability is the *k*th power of the transition probability matrix *P*. For example, as shown in Figure 4, the probability that drug D_1 acts on protein P_3 is equal to the product of the transition probability from P_1 to P_2 and the probability from P_2 to P_3 .

The proximity between drug i and drug j based on meta-path $DP^{(k)}D$ can be unfolded as follows:

$$s^{(k)}(i, j) = (DP^{(k)}D)_{ij} = (DP^kD)_{ij} = D_i P P^{k-2} (D_j P)^T = D_i P \times \underbrace{P \times P \times \dots \times P}_{k-2} \times (D_j P)^T \quad (4)$$

where P is the transition probability matrix in the PPI network, D is the original drug-target matrix, and $D_i P$ represents the target row vector of the i th drug.

$$s(i, j) = \sum_{k=2}^n S^{(k)}(i, j) = \sum_{k=2}^n (DP^kD)_{ij} = D_i P \left(\sum_{k=2}^n P^{k-2} \right) (D_j P)^T \quad (5)$$

Given tens of thousands of nodes in the PPI network, calculating P^k is very difficult. However, P^k is very common in the random walk theory. During the random walk procedure, a walker starts from an initial node and moves to neighbors with probability μ and back to the initial node with probability $1 - \mu$. Based on the Katz model [31], which is a method of computing similarities between nodes in a graph, taking into account not only the direct edges but also the indirect edges, Equation (5) can be rewritten as follows:

$$s(i, j) \approx D_i P \left[\sum_{k=2}^n (\mu P)^{k-2} \right] (D_j P)^T = D_i P [(I - \mu P)^{-1} - I] (D_j P)^T \quad (6)$$

where I is the identity matrix and the damping factor μ usually is 0.98. The inverse of the matrix in Equation (6) can be calculated using the SVD-based matrix factorization method.

2.3.2. Prior Drug Embedding

A stacking auto-encoder is a multi-layer deep neural network based on layer-wise training in which different multi-granularity data features are learned layer by layer and higher complex features are learned in higher layers. To enhance the robustness of sub-network embedding, we adopted stacked denoising auto-encoders (SDAE) in which the input neurons in every layer were randomly discarded by assigning some of the input neurons to 0 with a certain probability.

Traditional SDAE is an unsupervised model, which is composed of the encoder stage and decoder stage. At the encoder stage, the input data x_i are mapped into representation vector space, whereas at the decoder stage, the output data \hat{x}_i are the reconstructed data from x_i . The optimizer objective function of the SDAE is to minimize the reconstruction error of the output and input. The loss function is shown as follows:

$$L_1 = \sum_{i=1}^n \|\hat{x}_i - x_i\|_2^2 \quad (7)$$

Here, to protect the meta-path-based proximity of every sub-network, we adopted a semi-supervised SDAE framework [32]. The different meta-path-based proximities are the supervision information that preserves the proximity of the representation of two nodes. The optimizer objective function for this goal is defined as follows:

$$L_2 = \sum_{i,j=1}^n S^{(p)}_{ij} \left\| e^{(p)}_i - e^{(p)}_j \right\|_2^2 \quad (8)$$

where $S^{(p)}_{ij}$ is the proximity of drug i and drug j based on meta-path p and $e^{(p)}_i$ is the embedding of x_i based on the corresponding meta-path.

The objective function of the semi-supervised SDAE model, which combines Equations (7) and (8), is as follows:

$$L = L_1 + \alpha L_2 + \beta L_{reg} \quad (9)$$

where L_{reg} (as shown in Equation (12)) is an L_2 -norm regularizer term to prevent overfitting and α and β are hyperparameters. $W^{(k)}$ and $\hat{W}^{(k)}$ are the k th layer weight matrices at the encoder and decoder stages, respectively.

$$L_{reg} = \sum_{k=1}^K (\|W^{(k)}\|_F^2 + \|\hat{W}^{(k)}\|_F^2) \quad (10)$$

2.3.3. Secondary Drug Embedding

After obtaining the sub-network embeddings, we concatenated the embeddings together and used the secondary semi-supervised stacking denoising auto-encoder to obtain the final drug embedding. Given the embedding node $e^{(p)}_i$ of drug i in a different meta-path p , we concatenated them to obtain a new representation vector e_i . Then, we utilized auto-encoder layers to learn the final embedding e^F_i of drug i (as shown in Equation (11)).

$$\begin{aligned} e^{F(1)}_i &= \sigma(W^{F(0)}e_i + b^{F(0)}) \\ e^{F(2)}_i &= \sigma(W^{F(1)}e^{F(1)}_i + b^{F(1)}) \\ &\dots \\ e^F_i &= e^{F(h)}_i = \sigma(W^{F(h-1)}e^{F(h-1)}_i + b^{F(h-1)}) \end{aligned} \quad (11)$$

Here, we continued to adopt a semi-supervised SDAE framework to protect the original proximities of drugs in every sub-network. The supervision information in the optimizer objective function is defined as follows:

$$L_2 = \sum_{i,j=1}^n \frac{\sum_{p=1}^K \alpha^{(p)} S^{(p)}_{ij}}{\sum_{p=1}^K \alpha^{(p)}} \|e^F_i - e^F_j\|_2^2 \quad (12)$$

where $S^{(p)}_{ij}$ is the proximity of drug i and drug j based on meta-path p and $\alpha^{(p)}$ is a hyperparameter, which is the weight coefficient of the meta-path p . At the experimental stage, the best hyperparameters $\alpha^{(p)}$ are learned using 10-fold cross-validation on 10% labeled data with a grid search over $\alpha^{(p)} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

The objective function of this part is shown in Equation (13).

$$L_G = \sum_{i=1}^n \|\hat{e}_i - e_i\|_2^2 + \alpha L_2 + \beta L_{reg} \quad (13)$$

2.4. Prediction Formulation

For prediction tasks, learning a classifier that can be generalized to unknown ADRs is desirable. We predict the labels on training data using a fully-connected layer $y = h(e^F_i) = \sigma(W^P e^F_i + b)$. The prediction loss is formulated by Equation (14).

$$L = \sum_{y \in Y_{train}} [-y \ln y' - (1 - y) \ln(1 - y')] + \lambda \|W^P\|_F^2 \quad (14)$$

3. Experiment

3.1. Implementation and Evaluation Strategy

Proposed model: We implemented the proposed model with TensorFlow 1.2 and trained the model using the adaptive learning rate optimizer Adam [33]. All neurons were activated by the

sigmoid function. We optimized the hyperparameters in the model using validation data and then fixed them for all denoising auto-encoder layers.

Baseline: In addition, we implemented the following three network embedding baselines for comparison:

- Concatenate drug features: This method is a simple original HIN embedding method [28]. The approach constructs a feature vector for each drug by concatenating the PCA representation of each correlation matrix, which represents one aspect of the drug character.
- GraphCNN [34]: GraphCNN is a recently-proposed network embedding method based on spectral convolutional operation and achieves state-of-the-art performance on important prediction problems in recommender systems. Here, first, we linearly integrated similarity matrices based on all meta-paths except the target propagation meta-path $DP^{(n)}D$ and then learned the drug embeddings using the same GraphCNN structure described in [35].
- metapath2vec++ [36]: metapath2vec++ is a heterogeneous information network embedding method based on a meta-path-guided random walk strategy.

For further validation of the impact of target propagation on improving the quality of ADR predictions, we designed a network embedding algorithm without regard for the impact of protein-protein interactions and discarded the PPI dataset; this algorithm was named SDHINE-no-target propagation.

The subsequent ADR prediction methods after the network embedding stage were all based on the same loss function in every prediction task.

Evaluation: We evaluated and compared these algorithms using a 10-fold cross-validation methodology. We randomly selected a fixed percentage (10%) of drugs as the test set and moved all ADRs associated with these drugs from the dataset. The side effects and DDIs of these drugs were all set to 0. The other 90% of the drugs were further divided into the training set and validation set. The training set was formed with 95% of the remaining drugs and was used to train the model. The validation set was formed with the other 5% of the drugs and was used to test the model performance. The independent validation experiments were repeated 30 times with different random divisions of the data for the three sets.

The metrics used to evaluate the model performance were two common ranking metrics: mean average precision at K (MAP@K) and area under the receiver operating characteristic curve (ROC-AUC).

Average precision at K (AP@K) reflects the accuracy of the top-ranked ADRs by a model and can be computed as the mean of Precision@k for each drug or drug pair in the test set. The formula for computing AP@K is given as follows:

$$AP@K = \sum_{k=1}^K \text{Precision}(k) / \min(L, K) \quad (15)$$

where Precision(k) is the precision at cut-off k in the return list. L is the total number of true ADRs for the test drug or drug pairs.

3.2. Experimental Results

3.2.1. Visualization Results

First, we compared the performances of all network embedding approaches for the visualization task, which aimed to layout the drug HIN in a 2-dimensional vector space. We mapped the representation vectors of drugs obtained from all comparison approaches to a 2D vector space using the t-SNE [37]. Once a drug is successfully developed, the chemical substructure is fixed. The targets and side effects of a drug are all affected by the chemical substructures of the drug. Therefore, to compare

the dimensional reduction performance of different network embedding approaches, the drugs are firstly clustered into different clusters based on their chemical substructures.

The results are shown in Figure 6, in which drugs belonging to the same cluster are represented by the same color. The concatenated drug features method and metapath2vec++ could not separate drugs from different groups. GraphCNN and SDHINE-no-target propagation basically separated drugs from different groups, but some dark green points were mixed with the other groups. The results obtained with SDHINE were the best among these methods, because it separated most of the drugs from the different groups. This result was consistent with the fact that deep integration of different characteristics can effectively eliminate noise from data and recover the original signal.

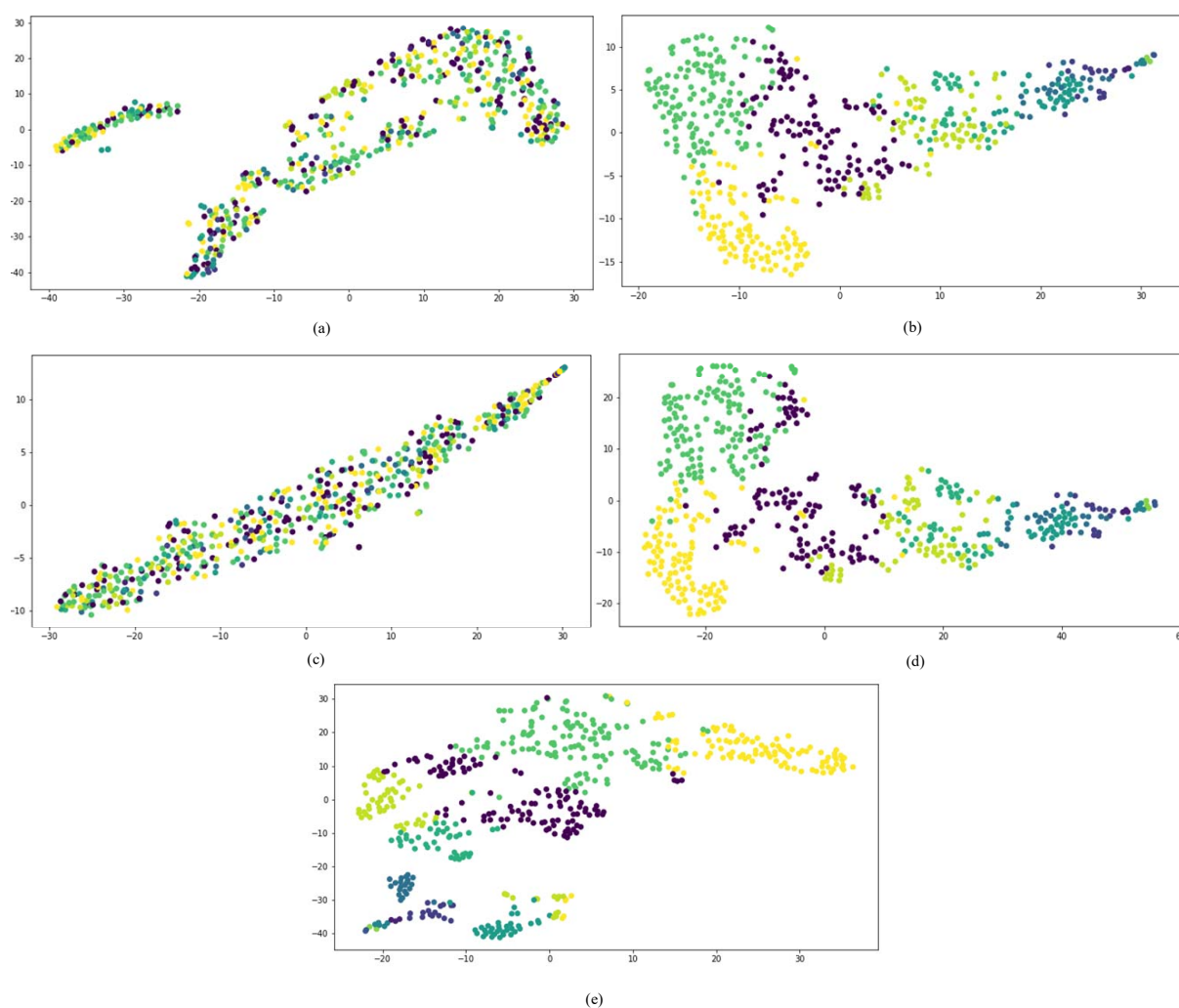


Figure 6. Visualization of the different representations: (a) concatenate drug features; (b) GraphCNN; (c) metapath2vec++ ; (d) SDHINE-no-target propagation; (e) SDHINE.

3.2.2. Prediction Results

Our experiments further evaluated the drug embeddings obtained through different network embedding methods on different tasks, including side effect predictions for a single drug, binary predictions of the occurrence of DDIs, and multi-label predictions of specific DDI types.

- Task 1: Predicting side effects of a single drug.

To demonstrate the side effect prediction performance based on our network embedding approach, we performed comparison experiments with the aforementioned three baselines and our two proposed

models. Predicting the types of side effects caused by one single drug can be formulated into a multi-label classification problem. The output value y of the prediction formulation in Section 2.4 is a column vector with 1318 dimensions, and W^P is a weighted matrix. Each type of side effect was trained one by one. The negative sampling method was adopted to settle the sample unbalanced problem.

Detailed comparisons of the experimental results are shown in Table 4. Our model based on target propagation clearly performed best compared with other models without a target propagation process in terms of MAP@20 and MAP@100. It was also very close to the best result in terms of MAP@50. Analogously, our approach improved ROC-AUC by 5.87% (84.07% vs. 78.20%) compared to the worst result. From the perspective of the approaches based on deep architecture, GraphCNN and the two SDHINE models performed better than the other models. Meanwhile, the model based on target propagation clearly improved the performance by 3.86% (84.07% vs. 80.21%) compared with the similar model without a target propagation process in terms of ROC-AUC.

Table 4. Side effect identification performance comparison.

Models	MAP@20	MAP@50	MAP@100	ROC-AUC
Concatenate drug features	0.5590	0.5475	0.5310	0.7820
GraphCNN	0.6510	0.6493	0.6321	0.8190
metapath2vec++	0.5835	0.5760	0.5628	0.7845
SDHINE-no-target propagation	0.6508	0.6416	0.6356	0.8021
SDHINE	0.6653	0.6479	0.6361	0.8407

- Task 2: Binary prediction of the occurrence of DDIs.

When one drug is administered with another drug, the effect of the drug may be changed, and an unknown side effect may be caused by the DDI. Detecting the occurrence of DDIs is preparation for further research on the ADRs induced by DDIs. When predicting the occurrence of a DDI without regard for the type of DDI, the prediction task can be modeled as a binary classification problem. In this situation, a probability value can be the output layer of the prediction formulation in Section 2.4. W^P can be written as a weighted vector. The input layer is formed by the embedding vectors of the two drugs. Table 5 shows a detailed comparison of the experimental results obtained from the binary prediction task of the occurrence of DDIs. The model based on target propagation performed better compared with the similar model without a target propagation process in terms of the mean average precision at k and ROC-AUC. The target propagation strategy and deep architecture were still useful for improving the prediction of DDI occurrence.

Table 5. DDI occurrence identification performance comparison.

Models	MAP@20	MAP@50	MAP@100	ROC-AUC
Concatenate drug features	0.6122	0.5624	0.5432	0.7409
GraphCNN	0.6874	0.6715	0.6219	0.7918
metapath2vec++	0.6542	0.6326	0.5986	0.7332
SDHINE-no-target propagation	0.6813	0.6718	0.6211	0.7814
SDHINE	0.7015	0.6854	0.6328	0.8124

- Task 3: Multi-label prediction of specific adverse DDI types.

Compared with the prediction of DDI occurrence, most often, we need to address which types of side effects are caused by the DDI. This issue is a multi-label classification problem in which the output layer y in the prediction formulation is a column vector with 1318 DDI events. The input layer is concatenated by two drug representation vectors, and W^P is a weighted matrix.

Detailed comparison experimental results for specific adverse DDI type identification tasks are shown in Table 6. The model based on target propagation and deep architecture was superior to

the models without a target propagation process or deep architecture in terms of not only the mean average precision at k , but also the ROC-AUC value.

Table 6. DDI type identification performance comparison.

Models	MAP@20	MAP@50	MAP@100	ROC-AUC
Concatenate drug features	0.6596	0.6144	0.5045	0.74322
GraphCNN	0.6823	0.6681	0.6137	0.7851
metapath2vec++	0.6766	0.6567	0.5118	0.7543
SDHINE-no-target propagation	0.6804	0.6622	0.6119	0.7996
SDHINE	0.6881	0.6745	0.6126	0.8175

Based on the results of the three prediction tasks, the network embedding approach with target propagation performance was superior to the approaches without target propagation processing. Moreover, approaches based on deep architecture performed better than the other linear network embedding methods and the combination methods. This result indicates the feasibility of predicting ADRs based on target propagation and proves that the deep learning process is effective at heterogeneous information network embedding.

3.3. Performance Comparison of Different Embedding Dimensions

To examine the impact of embedding size on prediction performance, we compared SDHINE and SDHINE-no-propagation with different dimensions of drug embeddings for three prediction tasks in terms of ROC-AUC. The results are shown in Figure 7. The prediction performances gradually increased with the increase of embedding dimension and reached the top when embedding dimensions were 64. The prediction performances at 256 dimensions were worse than that at 64 dimensions. This is because the higher dimensional embedding reduced the drug's differentiability. From the results, we also can find that SDHINE performed better than the same model without the target propagation process at the same embedding dimension on all three prediction tasks. It further verified our assumption that target propagation based on PPI can improve ADRs' prediction performance.

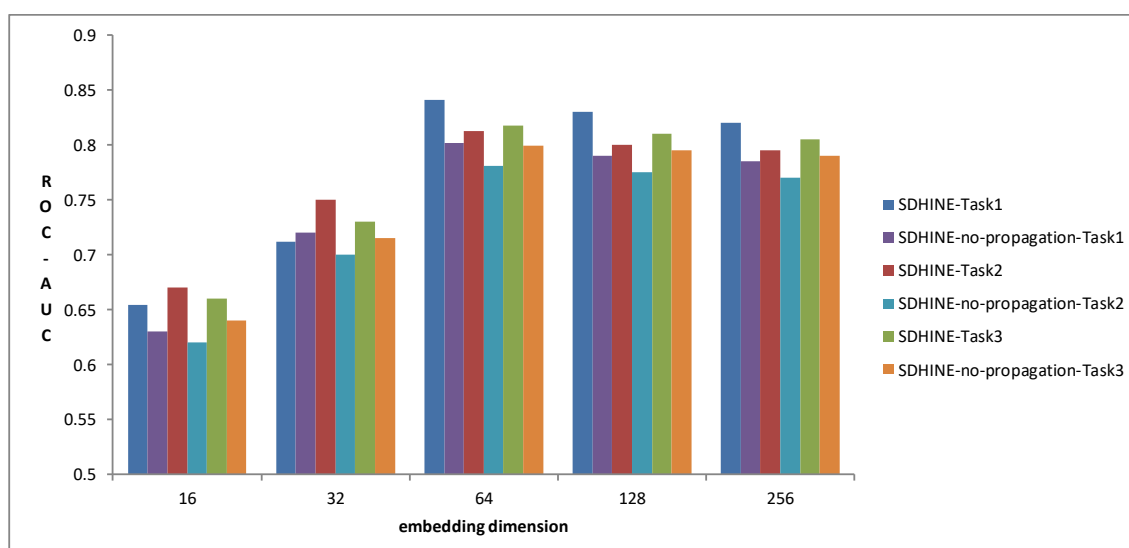


Figure 7. Performance comparison of different embedding dimensions.

3.4. Case Studies

We examined how the proposed network embedding method predicted potential unknown side effects based on learned drug embeddings. In this article, we only can query the drugs that are in

the selected datasets by inputting their id. We took triamcinolone, which is an intermediate-acting synthetic glucocorticoid given orally, by muscular or intra-articular injection, or as a topical ointment or cream and is used to treat various medical conditions (e.g., eczema and ulcerative colitis), as an example. It has been confirmed that it may cause many kinds of side effects, such as cough, headache, influenza, and so on. In the SIDER database, triamcinolone may cause 147 kinds of side effects, and 37 of them have been confirmed in preclinical in vitro safety profiling and clinical drug safety trials.

Table 7 shows the top 10 side effects of triamcinolone based on our model. We found from the results that most of the returned side effects were confirmed and that only two side effects were not confirmed. We analyzed the reasons for this result by taking eye redness as an example. First, we analyzed the target protein of triamcinolone from the DrugBank database and found that triamcinolone activated the target protein NR3C1. Then, we calculated the similarity based on the target propagation meta-path and found that NR3C1 might walk to protein EGFR with a high transition probability in the PPI network. EGFR is the only target protein of the drug gefitinib. We searched the side effects of gefitinib in the SIDER and OFFSIDES databases and found that gefitinib had an eye redness side effect. Thus, we found that eye redness is a potential side effect of triamcinolone based on the above logical inference. The works in [38–40] reported that peribulbar injections of triamcinolone may cause intraocular pressure (IOP) elevation, keratitis, and cataract. It is reasonable to believe that eye redness could be one of the ADRs of triamcinolone.

Table 7. Prediction of the top 10 side effects for triamcinolone based on SDHINE.

Top K	Side Effect	Confirmation
K = 1	headache	yes
K = 2	cough	yes
K = 3	fever	yes
K = 4	eye redness	no
K = 5	sneezing	yes
K = 6	nausea	yes
K = 7	rash	yes
K = 8	fatigue	yes
K = 9	dry skin	no
K = 10	conjunctivitis	yes

4. Conclusions

In this work, we proposed to utilize the impact of protein–protein interactions on drug targets to improve the prediction performance of adverse drug reactions. We designed a meta-path-based heterogeneous information network embedding approach (SDHINE) to integrate multi-source drug information, especially the PPI network. Different meta-path-based proximity calculation methods are designed for different semantic meta-paths. We adopted a semi-supervised stacked denoising auto-encoder to learn drug embeddings in each type of meta-path and integrated them into a second auto-encoder to learn the final drug embeddings. Extensive experiments were performed to compare our algorithm with several state-of-the-art network embedding methods for three ADR prediction tasks, which demonstrated the effectiveness of SDHINE. We also verified the ability of SDHINE to distinguish side effect types and performed a case study by examining the impact of protein-protein interactions on side effects.

In this work, we only considered the meta-paths in which the start and end nodes are all drugs. In future work, we will investigate how to use meta-paths starting from other objects (e.g., side effect nodes) under the guarantee of rationality and interpretability. As a major issue in ADR prediction, we will also consider how to further enhance the interpretability of prediction methods and results based on the semantics of meta-paths.

Author Contributions: H.W. guided the research; B.H. designed the experiments and wrote the manuscript; W.Y. analyzed the data; L.W. performed the experiments.

Funding: The work is partially supported by the National Natural Science Foundation of China (Nos. 61672329, 61373149, 61472233, 61572300, 81273704), the Shandong Province Science and Technology Plan Supported Project (No. 2014GGX101026), the Project of Shandong Province Higher Educational Science and Technology Program (No. J18KA370), and the Taishan Scholar Fund of Shandong Province (No. TSHW201502038, 20110819). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the TITAN X GPU used for this research and the support of Shandong provincial key laboratory for distributed computer software novel technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giacomini, K.M.; Krauss, R.M.; Dan, M.R.; Eichelbaum, M.; Hayden, M.R.; Nakamura, Y. When good drugs go bad. *Nature* **2007**, *446*, 975–977. [[CrossRef](#)] [[PubMed](#)]
2. Whitebread, S.; Hamon, J.; Bojanic, D.; Urban, L. Keynote review: In vitro safety pharmacology profiling: An essential tool for successful drug development. *Drug Discov. Today* **2005**, *10*, 1421–1433. [[CrossRef](#)]
3. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38*, 355–360. [[CrossRef](#)] [[PubMed](#)]
4. Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V. DrugBank 3.0: A comprehensive resource for ‘Omics’ research on drugs. *Nucleic Acids Res.* **2011**, *39*, D1035. [[CrossRef](#)] [[PubMed](#)]
5. Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L.J.; Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **2010**, *6*, 343. [[CrossRef](#)] [[PubMed](#)]
6. Li, Q.; Cheng, T.; Wang, Y.; Bryant, S.H. PubChem as a public resource for drug discovery. *Drug Discovery Today* **2010**, *15*, 1052–1057. [[CrossRef](#)] [[PubMed](#)]
7. Yamanishi, Y.; Pauwels, E.; Kotera, M. Drug side effect prediction based on the integration of chemical and biological spaces. *J. Chem. Inf. Model.* **2012**, *52*, 3284–3292. [[CrossRef](#)]
8. Li, J.; Zheng, S.; Chen, B.; Butte, A.J.; Swamidass, S.J.; Lu, Z. A survey of current trends in computational drug repositioning. *Brief. Bioinf.* **2016**, *17*, 2–12. [[CrossRef](#)]
9. Xu, B.; Shi, X.F.; Zhao, Z.H.; Zheng, W. Leveraging Biomedical Resources in Bi-LSTM for Drug Drug Interaction Extraction. *IEEE Access* **2018**, *17*, 33432–33439. [[CrossRef](#)]
10. Vilar, S.; Tatonetti, N.P.; Hripcsak, G. 3D Pharmacophoric Similarity improves Multi Adverse Drug Event Identification in Pharmacovigilance. *Sci. Rep.* **2015**, *5*, 8809. [[CrossRef](#)]
11. Labute, M.X.; Zhang, X.; Lenderman, J.; Bennion, B.J.; Wong, S.E.; Lightstone, F.C. Adverse drug reaction prediction using scores produced by large-scale drug-protein target docking on high-performance computing machines. *PLoS ONE* **2014**, *9*, e106298. [[CrossRef](#)] [[PubMed](#)]
12. Tatonetti, N.P.; Ye, P.P.; Daneshjou, R.; Altman, R.B. Data-Driven Prediction of Drug Effects and Interactions. *Sci. Transl. Med.* **2012**, *4*, 125–131. [[CrossRef](#)] [[PubMed](#)]
13. Ping, Z.; Fei, W.; Hu, J. Towards Drug Repositioning: A Unified Computational Framework for Integrating Multiple Aspects of Drug Similarity and Disease Similarity. *AMIA Annu. Symp. Proc.* **2014**, *2014*, 1258–1267.
14. Zhang, W.; Chen, Y.; Tu, S.; Liu, F.; Qu, Q. Drug side effect prediction through linear neighborhoods and multiple data source integration. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, Shenzhen, China, 15–18 December 2016; pp. 427–434.
15. Segura-Bedmar, I. Using a shallow linguistic kernel for drug–drug interaction extraction. *J. Biomed. Inf.* **2011**, *44*, 789–804. [[CrossRef](#)] [[PubMed](#)]
16. Jin, B.; Yang, H.; Xiao, C.; Zhang, P.; Wei, X.; Wang, F. Multitask Dyadic Prediction and Its Application in Prediction of Adverse Drug–Drug Interaction. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 1367–1373.
17. Zhang, W.; Chen, Y.; Liu, F.; Luo, F.; Tian, G.; Li, X. Predicting potential drug–drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinf.* **2017**, *18*, 18. [[CrossRef](#)] [[PubMed](#)]
18. Zitnik, M.; Agrawal, M.; Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* **2018**, *34*, i457–i466. [[CrossRef](#)] [[PubMed](#)]

19. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.J.; Yang, Q.; Lin, S. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 40–51. [[CrossRef](#)]
20. Cao, S.; Lu, W.; Xu, Q. Deep neural networks for learning graph representations. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 1145–1152
21. Huang, Z.; Mamoulis, N. Heterogeneous Information Network Embedding for Meta Path based Proximity. Available online: <https://arxiv.org/abs/1701.05291> (accessed on 19 Jan 2017).
22. Li, R.; Dong, Y.; Kuang, Q.; Wu, Y.; Li, Y.; Zhu, M.; Li, M. Inductive matrix completion for predicting adverse drug reactions (ADRs) integrating drug–target interactions. *Chemom. Intell. Lab. Syst.* **2015**, *144*, 71–79. [[CrossRef](#)]
23. Ma, T.; Xiao, C.; Zhou, J.; Wang, F. Drug Similarity Integration Through Attentive Multi-view Graph Auto-Encoders. Available online: <https://arxiv.org/abs/1804.10850> (accessed on 28 Apr 2018).
24. Kelley, B.P.; Sharan, R.; Karp, R.M.; Sittler, T.; Root, D.E.; Stockwell, B.R.; Ideker, T. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Nat. Acad. Sci. USA* **2003**, *100*, 11394–11399. [[CrossRef](#)]
25. Yeh, C.Y.; Yeh, H.Y.; Arias, C.R.; Soo, V.W. Pathway Detection from Protein Interaction Networks and Gene Expression Data Using Color-Coding Methods and A* Search Algorithms. *Sci. World J.* **2012**, *2012*, 315797. [[CrossRef](#)]
26. Codling, E.A.; Plank, M.J.; Benhamou, S. Random walk models in biology. *J. R. Soc. Interface* **2008**, *5*, 813–834. [[CrossRef](#)] [[PubMed](#)]
27. Zou, Q.; Li, J.; Song, L.; Zeng, X.; Wang, G. Similarity computation strategies in the microRNA-disease network: A Survey. *Brief. Funct. Genom.* **2016**, *15*, 55–64. [[CrossRef](#)] [[PubMed](#)]
28. Shi, C.; Li, Y.; Zhang, J.; Sun, Y.; Yu, P.S. A Survey of Heterogeneous Information Network Analysis. *IEEE Trans. Knowl. Data. Eng.* **2016**, *29*, 17–37. [[CrossRef](#)]
29. Shakibian, H.; Charkari, N.M. Mutual information model for link prediction in heterogeneous complex networks. *Sci. Rep.* **2017**, *7*, 44981. [[CrossRef](#)] [[PubMed](#)]
30. Chang, S.; Han, W.; Tang, J.; Qi, G. J.; Aggarwal, C.C.; Huang, T.S. Heterogeneous Network Embedding via Deep Architectures. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, New South Wales, Australia, 10–13 August 2015; pp. 119–128.
31. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [[CrossRef](#)]
32. Wang, D.; Cui, P.; Zhu, W. Structural Deep Network Embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1225–1234.
33. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 22 December 2014).
34. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. Available online: <https://arxiv.org/abs/1609.02907> (accessed on 9 September 2016).
35. Kipf, T.N.; Welling, M. Variational Graph Auto-Encoders. Available online: <https://arxiv.org/abs/1611.07308> (accessed on 21 November 2016).
36. Dong, Y.; Chawla, N.V.; Swami, A. In metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2017; pp. 135–144.
37. Maaten, L.V.D.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
38. Hashizume, K.; Nabeshima, T.; Fujiwara, T.; Machida, S.; Kurosaka, D. A case of herpetic epithelial keratitis after triamcinolone acetonide subtenon injection. *Cornea* **2009**, *28*, 463–464. [[CrossRef](#)] [[PubMed](#)]
39. Suarez-Figueroa, M.; Contreras, I.; Noval, S. Side-effects of triamcinolone in young patients. *Arch. Soc. Esp. Ophthalmol.* **2006**, *81*, 405–407. [[PubMed](#)]
40. Chew, E.Y.; Glassman, A.R.; Beck, R.W. Ocular side effects associated with peribulbar injections of triamcinolone acetonide for diabetic macular edema. *Retina* **2011**, *31*, 284. [[CrossRef](#)]

