Applied Intelligence (2021) 51:8316–8333
https://doi.org/10.1007/s10489-021-02296-4

Modeling polypharmacy effects with heterogeneous signed graph convolutional networks

Taoran Liu[1] · Jiancong Cui[1] · Hui Zhuang[1] · Hong Wang[1,2]

Accepted: 1 March 2021 / Published online: 1 April 2021

Abstract

Pharmaceutical drug combinations can effectively treat various medical conditions. However, some combinations can cause serious adverse drug reactions (ADR). Therefore, predicting ADRs is an essential and challenging task. Some existing studies rely on single-modal information, such as drug-drug interaction or drug-drug similarity, to predict ADRs. However, those approaches ignore relationships among multi-source information. Other studies predict ADRs using integrated multi-modal drug information; however, such studies generally describe these relations as heterogeneous unsigned networks rather than signed ones. In fact, multi-modal relations of drugs can be classified as positive or negative. If these two types of relations are depicted simultaneously, semantic correlation of drugs in the real world can be predicted effectively. Therefore, in this study, we propose an innovative heterogeneous signed network model called SC-DDIS, to learn drug representations. SC-DDIS integrates multi-modal features, such as drug-drug interactions, drug-protein interactions, drug-chemical interactions, and other heterogeneous information, into drug embedding. Drug embedding means using feature vectors to express drugs. Then, the SC-DDIS model is also used for ADR prediction tasks. First, we fuse heterogeneous drug relations, positive/negative, to obtain a drug-drug interaction signed network (DDISN). Then, inspired by social network, we extend structural balance theory and apply it to DDISN. Using extended structural balance theory, we constrain sign propagation in DDISN. We learn final embedding of drugs by training a graph spectral convolutional neural network. Finally, we train a decoding matrix to decode the drug embedding to predict ADRs. Experimental results demonstrate effectiveness of the proposed model compared to several conventional multi-relational prediction approaches and the state-of-the-art deep learning-based Decagon model.

Keywords Polypharmacy effects · Adverse drug reaction · DDI prediction · Graph convolutional neural network · Signed network · Structural balance theory

1 Introduction

Pharmaceutical drugs can treat and relieve symptoms and prevent diseases; however, a single drug often shows limited efficacy, poor safety, and developed drug resistance [14]. Therefore, drug combination therapy has become a very effective way to treat diseases [1, 9]. Many people, particularly the elderly, take multiple drugs simultaneously. However, drug combinations may cause serious adverse drug reactions (ADR) due to drug-drug interactions (DDI). For example, in the United States, 100000 people die of ADRs each year [7]. At the same time, it is difficult to identify the ADRs manually [21]. Therefore, predicting previously unknown ADRs using heterogeneous drug information is a significant and challenging problem.

Traditional ADR detection methods rely on a large number of clinical trials. However, clinical trials are expensive and time-consuming, and sometimes are not possible due to the drug combinations' complexity. Currently, machine learning methods can be applied to ADR detection. Specifically, such methods represent drugs with their chemical structure and biological information and then use machine learning approaches to predict ADRs. Although these methods have been relatively successful [10, 20, 21, 24], many unsolved problems remain. First, such methods only focus on a single type of direct relationship between drugs and ignore the implicit, indirect contact

✉ Hong Wang
111052@sdnu.edu.cn

Extended author information available on the last page of the article.

between drugs, such as drug-chemical interactions, drug-protein interactions (DPI), and drug-disease interactions. Therefore, the characterization of the potential real semantic relationship between drugs is biased. Second, although some studies consider the multi-source heterogeneous relations between drugs [10, 21], the integration methods for these multi-source heterogeneous relations do not take into account the potential semantic information. This problem is even more serious when there are both positive and negative relations between drugs. Third, compared to the positive effects, the negative relations, such as ADRs between drugs, are relatively less frequent; in other words, the drug-drug negative interaction network is very sparse, which has a significantly adverse impact on ADR prediction.

To address these issues, inspired by the social network analysis method [11, 19, 27], we define the multi-source interactions between drugs using a heterogeneous signed network that describes the direct relationships and various indirect semantic relationships between drugs. In addition, the heterogeneous signed network depicts the positive and negative effects of drugs, which conforms to the semantic information about real-world effects. A signed network [27], also known as a polar network, refers to a network where edges are weighted positively and negatively according to their represented semantic relations. Signed networks have been widely used in various research areas, such as social, trust, and traffic control networks [18, 26, 27] and such studies have provided some promising results. However, few studies have investigated ADR-oriented heterogeneous signed networks to the best of our knowledge, and even fewer have considered the spectral convolution on heterogeneous signed networks. Therefore, we propose an innovative method of spectral convolution on drug-drug interactions signed network (SC-DDIS). Specifically, the SC-DDIS method defines different types of semantic information about drugs, such as DDIs, DPIs, and drug-chemical interactions, with positive and negative drug interactions and integrates them. The proposed SC-DDIS predicts drug-drug interaction relationships more accurately and also has excellent interpretability. The primary contributions of the proposed method are summarized as follows.

1) The proposed SC-DDIS is the first ADRs-oriented spectral convolution model on drug-drug interaction heterogeneous signed networks. It provides a more comprehensive description of drug relations because it simultaneously defines and depicts different types of semantic information about drugs. The SC-DDIS considers the real-world semantic information, enriches the information about drug characteristics, enhances the interpretability of the model, and alleviates the adverse effect of network sparsity on the model.
2) The proposed SC-DDIS extends the structure balance theory of the signed network and applies it to learn higher-order sign propagation features and further restricts the sign propagation process in the DDISN. These higher-order sign propagation features capture latent semantic relations implied in the SC-DDIS network, which improves the performance of the proposed SC-DDIS model. The accessibility matrix [28] can be formed by sign propagation. To vividly express the matrix after sign propagation, we will use the sign propagation matrix to refer to the accessibility matrix. We will explain in detail what a sign propagation matrix is in section 3.1.4.
3) We use a variety of methods to optimize the SC-DDIS model. First, we use the proportion of positive and negative samples to allocate different weights to samples to handle the problem of data imbalance in the SC-DDIS and further reduce the loss value. Second, we train a decoding matrix in the SC-DDIS model that can filter out noise, enlarge the critical features, improve the prediction rate of drug side effects, and further improve the model's prediction performance and robustness.
4) Extensive experiments were conducted to verify the effectiveness of the proposed SC-DDIS model. The experimental results demonstrated that the SC-DDIS outperformed a state-of-the-art deep learning method [21] and several other well-known multi-relational link prediction approaches [22, 23, 25, 31].

2 Related work

Previous ADR prediction studies can be classified as traditional and deep learning-based methods. Conventionally, pharmacological, topological, or semantic similarity based on statistical learning is calculated to predict ADRs [4, 15]. Aurel proposed predictive pharmacy interaction networks to predict unknown ADRs using the network structure formed by known DDIs, and various drug classification characteristics [2]. Huang found that integrating the protein-protein interaction (PPI) networks and the drug structures can improve ADR prediction performance [12]. Zhang proposed a matrix perturbation method based on DDI networks combined with similar drug characteristics to predict ADRs [29]. Park proposed the random walk with restart algorithm to simulate signal propagation on protein networks to predict ADRs [24]. Zheng measured and established a framework for drug similarity integration from various per-

spectives and proposed a method to select highly reliable negative samples to predict ADRs [30]. These methods do not rely on deep learning methods with strong learning representation ability, which we call traditional methods for predicting ADR.

With the development of deep learning, Kipf investigated the convolution operation on graphs according to the convolutional neural networks on images. He proposed a graph convolutional neural network (GCN) algorithm based on spectral convolution [17]. The algorithm uses the Laplace feature spectrum to maintain the network structure and the relationship between the Laplace matrix and the Fourier transform, and performs a convolution operation based on the Laplace feature spectrum. The GCN algorithm has achieved excellent results in graph link prediction task studies; therefore, many studies that employ a deep learning method to predict the side effects of drugs primarily use the graph convolution technique. Jure proposed a method to predict ADR by using GCN for DDI, PPI and DPI end-to-end learning [21]. In addition, many algorithms are based on the GCN algorithm to learn the embedded representation of the drug network, and finally predict ADRs [5, 29]. Liu proposed a structural network embedding method based on multi-mode deep automatic coding to predict ADRs [20]. The previously described studies demonstrate that the GCN has achieved excellent results in the embedded representation of drugs. Therefore, this paper also uses the GCN algorithm based on spectral convolution to obtain the embedded expression of drugs. Differing from the described related studies, we use a new method to integrate the drug heterogeneous network to form the DDISN. Then, a higher-order sign propagation network is obtained by constraining the sign propagation of DDISN with the extended structure balance theory, which can increase the amount of available information and compensate for the impact of some data imbalances. On this basis, we use the GCN algorithm to extract the information again to obtain the final drug feature expression.

Although the methods mentioned above have achieved some success, they ignore the rich implicit semantic information between drugs. Therefore, the learned drug node representation is inaccurate and difficult to interpret. Thus, in this paper, a heterogeneous signed network analysis method base on spectral convolution is proposed to predict ADRs in consideration of the multi-source semantic information between drugs, so as to improve interpretability of the model and enhance DDI prediction accuracy. In addition, we employ a variety of techniques to improve the robustness and effectiveness of the model. Specifically, we weight the loss value to solve the problem of imbalance between positive and negative examples in ADRs prediction. We also filter noise data using a training decoding matrix.

3 Methodology

In this section, we introduce the model architecture and related definitions, and use simple examples to explain the definitions.

3.1 SC-DDIS model

3.1.1 Background

In this section, we will briefly introduce some concepts in section 3.1.2 to understand the whole model's architecture. For more specific and detailed definitions of terms used in this paper, we put them in section 3.1.4 for a detailed explanation. DDI networks describe the side effects of drugs and drugs taken together. Biological drug profiles refer to drug subchemical, drug-protein interactions and other information. M-order sign propagation matrix refers to the matrix after the drug sign network has propagated m times, expressed by the formula as $Matrix_s^m$. $Matrix_s$ is a matrix representation of sign networks. GCN is a deep learning method applied to graph, which can effectively learn the representation of nodes in graph.

3.1.2 Model architecture

As shown in Fig. 1, the proposed model is divided into five sub-modules. For convenience, we assume that there are N drugs, and M side effects. In Part A, different side effects are extracted to construct the DDI networks. That is to say, the number of DDI networks is M. In Part B, we fuse DDI networks and biological drug profiles to obtain the M signed networks according to the different semantic relations. In Part C, we constrain the sign propagation according to the enhanced structure balance theory and obtain an m-order sign propagation matrix as the drug feature representation. In Part D, the initial DDI networks and m-order sign propagation matrix are used as input of our SC-DDIS model. Note that the input dimensions of a particular side effect prediction task are $N \times N$ and $N \times N$ is related to a and b in Fig. 1, respectively. There are M side effects; thus, M times inputs are required to calculate each side effect prediction task's score. Then, the embedding representation of drugs is obtained and fed into the GCN spectral convolution component. In Part E, the embedded description of drugs is input into a fully-connected neural network to obtain the reconstructed DDI networks of different side effects to predict ADRs.

3.1.3 Graph convolution network (GCN)

Here, we briefly introduce GCN. Thomas Kpif proposed GCN in the paper Semi-supervised classification with

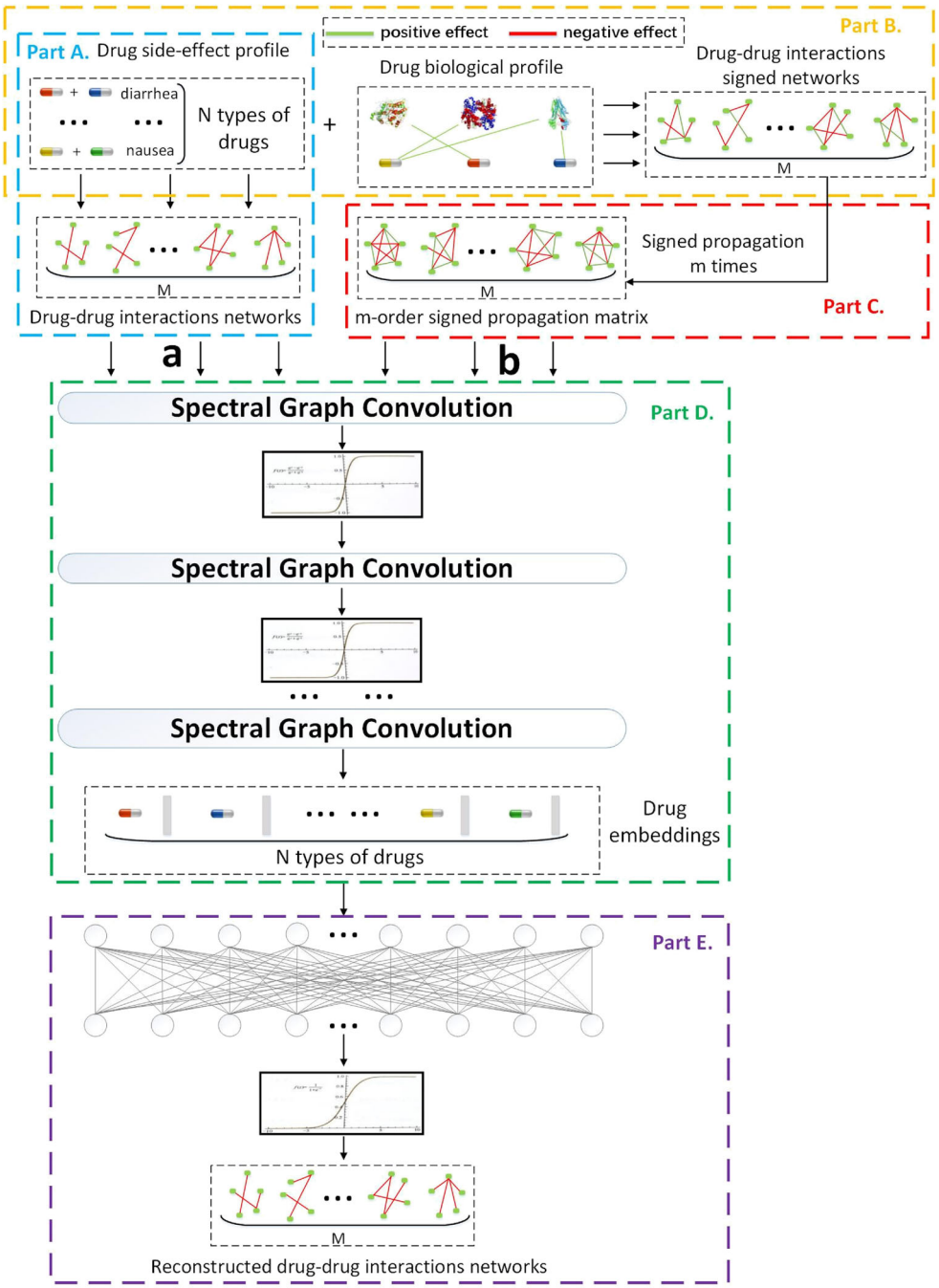

Fig. 1 Proposed side effect prediction model. The SC-DDIS model proposed in this paper consists of five parts: A-E

graph convolutional networks in 2017 [17]. It provides a new idea for graph structure data processing and applies the convolutional neural network commonly used for images in deep learning to graph data. The core of the GCN is based on the characteristic decomposition of the Laplace matrix, which is a semi-positive definite symmetric matrix with many good properties, e.g., a matrix comprising eigenvectors is orthogonal and the eigenvalue is nonnegative. The standard Laplace matrix [6] of the network is given in (1).

$$L = D - A \tag{1}$$

Here, D is the degree matrix (diagonal matrix) of the vertex, the elements on the diagonal are the degrees of each vertex in turn, and A is the adjacency matrix of the graph. The Laplacian matrix is a positive semi-definite symmetric matrix, which can be spectrally decomposed, and has a

particular form after decomposition as follows.

$$L = U \Lambda U^{-1} \tag{2}$$

Here, $U = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n})$ is the eigenvector matrix, and $\mathbf{u_i}$ is the column vector. $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix of n eigenvalues. According to the properties of the Laplace matrix, we know that U is an orthogonal matrix; thus, $UU^T = IUU^T = E$, i.e., $U^{-1} = U^T$. Taking U as the basis of Fourier transform, the rules of Fourier forward transform and inverse transform are defined as follows.

$$\hat{F} = U^T F \tag{3}$$

$$F = U\hat{F} \tag{4}$$

Here, F_i is the vector representation of Drug_i nodes in the DDIN. According to the positive and inverse Fourier transform, we change the DDIN from the spatial domain to the spectral domain via positive Fourier transform $(U^T F)$. Similarly, we change the convolution kernel to the spectral domain using positive Fourier transform $(U^T X)$. The inverse transformation of the product of the two Fourier transforms is obtained by multiplying the product of the two Fourier transforms by U. The final convolution is expressed as follows.

$$(F * X)_G = U((U^T X) \odot (U^T F)) \tag{5}$$

Here, X is the convolution kernel matrix. Equation (5) is a general expression of graph convolution. This paper will use a classical graph convolution neural network structure proposed by kipf [17], such as (6).

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{6}$$

Here, $\tilde{A} = A + I_N$ is the adjacency matrix of an undirected graph with self-connections. I_N is the identity matrix. $\tilde{D_{ii}} = \sum_j \tilde{A_{ij}}$ and $W^{(l)}$ is the trainable weight matrix of the l^{th} layer. $\sigma(\cdot)$ denotes an activation funtion, such as the ReLU(·) = max(0, ·). $H^l \in \mathbb{R}^{N \times D}$ is the expression matrix activated at the l^{th} layer. In a graph, N can represent the number of nodes in the graph, and D means to use d-dimensional data to represent a node in the graph.

3.1.4 Definitions

Here, we define the drug-drug negative effect, the drug-drug positive effect, and the drug-drug interaction network. We then describethe DDI signed network according to semantic information about real-world effects. Then, we explain the extended structure balance theory and sign propagation matrix, which can be applied to DDISN. Finally, the convolution operation on the drug-drug interaction network of side effects is defined.

1) Drug-drug negative effect (DDNE)

DDNE, or ADR, is some undesirable effect caused by the use of a drug. Note that most are natural pharmacological actions. DDNE relation between drugs can be described by a matrix denoted A_{NE}, where $A_{NE}(i, j)$, i.e., an element of A_{NE}, is 0 when the side effects between Drug_i and Drug_j are unknown. If side effects occur between Drug_i and Drug_j, then $A_{NE}(i, j) = 1$. At the same time, because DDNE is mutual, $A_{NE}(j, i) = 1$. In other words, A_{NE} should be a symmetric matrix.

2) Drug-drug positive effect (DDPE)

Let S_{ij} denote a normalized similarity between Drug_i and Drug_j, $\mu \in [0, 1)$ is the threshold value. If $S_{ij} > \mu$, there is a positive effect between Drug_i and Drug_j. If $S_{ij} \leq \mu$, the positive effect between Drug_i and Drug_j is unknown. That is to say, DDPE between two drugs depends on their similarity on biological drug profile. The DDPE is that it is not easy to cause adverse reactions between drugs.

The similarity may be different according to their different relations. At the same time, Jaccard [13] similarity is used to evaluate the similarity in two drugs in a specific relationship. For example, describing the DPIs between Drug_i and Drug_j is based on the Jaccard similarity of a target protein of a drug. In addition, describing the drug-chemical structure interactions between Drug_i and Drug_j is based on the Jaccard similarity of the chemical structure of the drug. Finally, the similarity of the two drugs is obtained by weighting, according to (7).

$$S(i, j) = \sum\nolimits_{rel} \alpha_{rel} \frac{|\Gamma_{rel}(i) \cap \Gamma_{rel}(j)|}{|\Gamma_{rel}(i) \cup \Gamma_{rel}(j)|} \tag{7}$$

Here, $rel \in \{\text{protein}, \text{chemical}, \ldots\}$ is the set of drug relationships, where $\sum_{rel} \alpha_{rel} = 1$. $\Gamma_{rel}(i)$ is the set of rel contacts for Drug_i, $\Gamma_{rel}(j)$ is the set of rel contacts for Drug_j, and $S(i, j)$ is the total similarity between Drug_i and Drug_j.

We take the DPIs as an example to illustrate the positive effects of drugs. Here, we set parameters $rel \in \{\text{protein}\}$ and $\alpha_{protein} = 1$, as shown in (8).

$$S(i, j) = \frac{|\Gamma_{protein}(i) \cap \Gamma_{protein}(j)|}{|\Gamma_{protein}(i) \cup \Gamma_{protein}(j)|} \tag{8}$$

Here, $\Gamma_{protein}(i)$ represents the target protein collection of Drug_i,and $\Gamma_{protein}(j)$ represents the target protein collection of Drug_j.

As shown in Fig. 2, if the target protein similarity of a drug pair is very high, the probability of side effects between that pair is very low (and vice versa). Therefore, when the similarity of the two drugs' target proteins is very high, the probability of side effects is negligible. Here, if the similarity of the two drug's target proteins exceeds the value of μ, there is a positive effect between the drugs, i.e., not easy to have side effects. In the experimental section of this

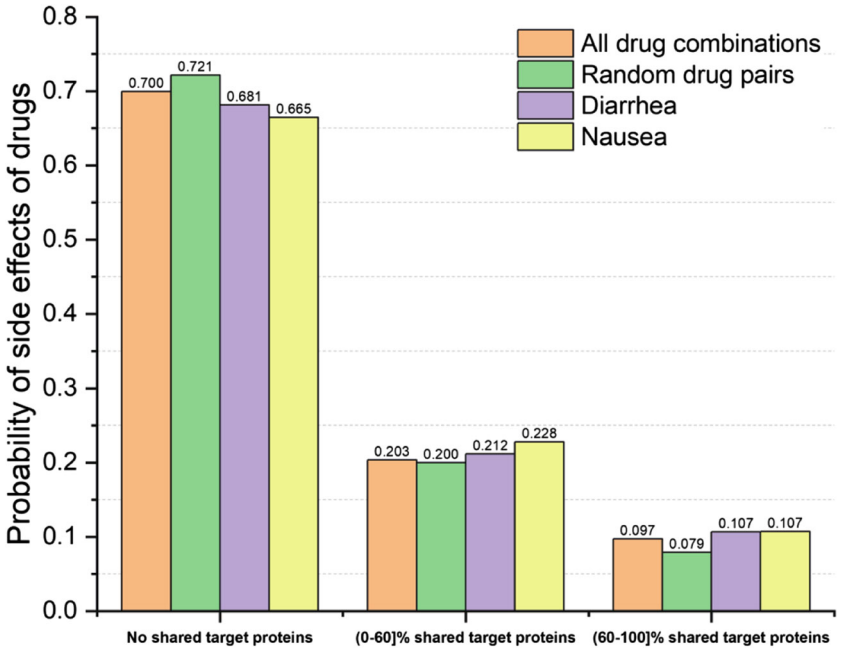

Fig. 2 Relationship between target protein-based DDPE and DDNE. It shows the relation between the target protein similarity of any drug pair and its ADR probability. When the target protein similarity is 0, the probability of side effect of any drug pair is between 66-72%. When the target protein similarity increases to 60%, the probability of side effect decreases to 20-23%. As the target protein similarity continues to rise to 100%, the probability of side effect of any drug pair is lower than 11%. To prove the validity of this relationship, we conduct four group of experiments, all drug combinations, randomly selected drug combinations, drug combinations that cause diarrhea, and drug combinations that cause nausea

article, we validate the effectiveness of defining the positive effects of drugs by the similarity of target proteins. Here, μ is the hyperparameter of the proposed SC-DDIS model, which is obtained experimentally.

3) Drug-drug interactions network (DDIN)

The DDIN is formalized as $G_{DDIN} = (V, E, A_{DDIN})$. Here, V is the set of nodes (nodes represent drugs), E is the set of relationships between nodes, and A_{DDIN} is the adjacency matrix of network G_{DDIN}. In addition, $e(i, j) \in \{0, 1\}$ represents the interaction between nodes $i \in V$ and $j \in V$. If there is a DDNE between Drug_i and Drug_j, then $e(i, j) = 1$. If the DDNE between Drug_i and Drug_j is unknown, then $e(i, j) = 0$. Without loss of generality, we assume $e(i, j) = e(j, i)$; therefore, element $A_{DDIN}(i, j)$ of the adjacency matrix A_{DDIN} of network G_{DDIN} can be expressed as follows.

$$A_{DDIN}(i, j) = \begin{cases} 1, & \text{there is DDNE between } i \text{ and } j \\ 0, & \text{DDNE between } i \text{ and } j \text{ is unknown} \end{cases} \tag{9}$$

4) DDI signed network(DDISN)

The DDISN is formalized as $G_{DDISN} = (V, E, A_{DDISN}, A_{DDIN}, S, \mu)$, where V is a collection of nodes (nodes represent drugs), E is the set of relationships between nodes, and A_{DDISN} is the adjacency matrix of network G_{DDISN}. In addition, A_{DDIN} is the adjacency matrix of the DDIN, S is the drug similarity matrix, $\mu \in [0, 1)$ is the drug similarity threshold, and $e(i, j) \in E$ represents the interactions between $\text{Drug}_i \in V$ and $\text{Drug}_j \in V$. Here, if $S_{ij} > \mu$ and $A_{DDIN}(i, j) = 0$, there is a positive effect between Drug_i and Drug_j, and $S_{ij} \leq \mu$ and $A_{DDIN}(i, j) = 0$ suggests that the effect between Drug_i and Drug_j is unknown. If $A_{DDIN}(i, j) = 1$, there are side effects between Drug_i and Drug_j. Without loss of generality, we assume $e(i, j) = e(j, i)$ such that element $A_{DDISN(i,j)}$ of adjacency matrix A_{DDISN} of network G_{DDISN} can be expressed as follows.

$$A_{DDISN}(i, j) = \begin{cases} 1, & \text{there is DDPE between } i \text{ and } j \\ 0, & \text{DDPE between } i \text{ and } j \text{ is unknown} \\ -1, & \text{there is DDNE between } i \text{ and } j \end{cases} \tag{10}$$

5) Extended structural balance theory

The structural balance theory was put forward by Heider in the 1940s [3]. Some basic rules defined in this theory have been widely used in link signed prediction tasks in signed networks, which has become the basic theory of signed networks. The theory points out that "friends of friends are more likely to be my friends" and "enemies of friends are more likely to be my enemies" are consistent with social psychology. Based on Heider's structural balance theory, we define the extended structural balance theory as follows. If Drug_A and Drug_B have a positive effect and Drug_B and Drug_C have a positive effect, then Drug_A and Drug_C have a positive effect. In addition, if Drug_A and Drug_B have side effects and Drug_B and Drug_C have positive effects, then Drug_A and Drug_C have side effects. In other words, if Drug_A and Drug_B are very similar, and Drug_B and Drug_C are also very similar, then Drug_A and Drug_C are likely very similar, i.e., there is likely no side effect between Drug_A and Drug_C. In addition, if Drug_A and Drug_B have a side effect, and Drug_B and Drug_C are very similar, then Drug_A

and Drug_C are likely to have the same side effect. As shown in the drug balance triangle in Fig. 3, a third virtual edge can be predicted given two real edges.

6) Accessibility Matrix

Accessibility matrix refers to the degree that can be achieved after a certain length of the path between a directed graph's nodes. In the scenario described in this article, let's give a new name to the Accessibility Matrix, called Sign Propagation Matrix (SPM). The SPM is a matrix obtained from A_{DDISN} using a propagation operation according to the extended structural balance theory. The SPM is initialized first.

$$SPM_0 = I \tag{11a}$$

$$SPM_1 = A_{DDISN} \tag{11b}$$

Here, $I \in \mathbb{R}^{N\times N}$ is an identity matrix. The propagation mode of the SPM is defined as follows.

$$\begin{aligned} SPM_m &= \text{Sign}(SPM_{m-1} \times SPM_1) \\ &= \text{Sign}\left(\sum_{k=0}^{n-1}[SPM_{m-1}(i,k) \times SPM_1(k,j)]\right) \quad with\ \ m \geq 2 \end{aligned} \tag{12}$$

Sign(x) in (13) is defined as a sign function. Based on the extended structure balance theory, the propagation of the sign in the DDISN is constrained. The sign propagation method is defined by formal language to obtain m-order sign propagation matrix (SPM_m) in the DDISN.

$$\text{Sign}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases} \tag{13}$$

The element of SPM_m is defined as follows.

$$SPM_m(i,j) = \begin{cases} 1, & \text{positive effect between } i \text{ and } j \\ 0, & \text{unknown effect between } i \text{ and } j \\ -1, & \text{side effect between } i \text{ and } j \end{cases} \tag{14}$$

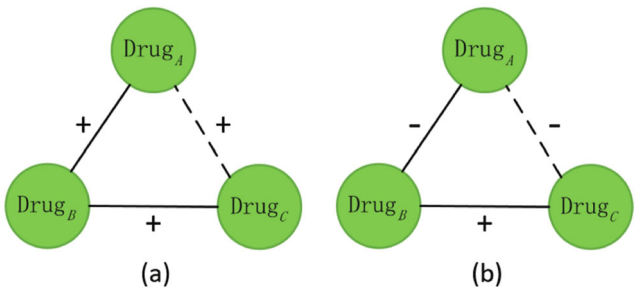

Fig. 3 Drug balance triangle based on extended structure balance theory. Drug_A, Drug_B, and Drug_C represent drugs. + indicates a positive effect between drugs, and − indicates a side effect between drugs. The solid line indicates the known relationship, and the dotted line indicates the relationship predicted according to the solid line

According to the extended structure balance theory, we can perform signed propagation in the DDISN and obtain high-order drug features. For example, the drug balance triangle in Fig. 3 can predict the third virtual edge when two real sides are known. With the above theory, the signs in the DDISN are propagated to obtain SPM_m. Through SPM_m, we can mine potential DDIs to obtain the higher order of DDI information, which plays an essential role in the convolution neural network.

7) Convolution on the DDIN of side effect r

We will apply the classic graph convolutional neural network structure to the prediction of drug side effects. Then the formula can be expressed by some concepts defined in this article.

$$H^{(l)} \leftarrow \sigma\left(\tilde{D}_r^{-\frac{1}{2}}\tilde{A}_r\tilde{D}_r^{-\frac{1}{2}}H^{(l-1)}W_r^{(l-1)}\right) \tag{15}$$

Here, $\tilde{A}_r = A_{rDDIN} + I_r$ is the adjacency matrix of the DDIN, $G_{rDDIN} = (V, E, A_{rDDIN})$ with added self-connections, and I_r is the N-order identity matrix. Here, $\tilde{D}_r(i,i) = \sum_j \tilde{A}_r(i,j)$. $W_r^{(l)}$ is the trainable weight matrix of the l^{th} layer. $\sigma(\cdot)$ denotes an activation funtion. $H^l \in \mathbb{R}^{N\times D}$ is the drug expression matrix activated at the l^{th} layer. In a DDISN, N can represent the number of drug nodes in the graph, and D means to use d-dimensional data to represent a drug node in the graph.

3.1.5 Simple application

Here, we use an actual drug heterogeneity network to apply the above theory to facilitate a comprehensive understanding of the theory. The drug heterogeneity network is shown in Fig. 4.

Figure 4 helps us understand the above theory. First, we only focus on the solid red lines, from which we can extract the known DDIN, which can be represented by the following adjacency matrix.

$$A_{DDIN} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{16}$$

Here, $\mu = 0.5, rel = \{\text{protein}\}$, and $\alpha_{protein} = 1$. It is easy to calculate $S_{AB} = \frac{2}{3} > \mu$ and $S_{CD} = S_{AA} = S_{BB} = S_{CC} = S_{DD} = 1 > \mu$ by observing the drug-protein network in Fig. 4. In other words, there is a positive effect between Drug_A and Drug_B, and Drug_C and Drug_D. Similarly, the green dotted lines in Fig. 5 indicate a positive effect. In addition, the similarity between each drug and itself is 1, which has a positive effect. Note that this is an obvious conclusion; thus, it is not identified in Fig. 5 According to the known DDIN, the adjacency matrix is used

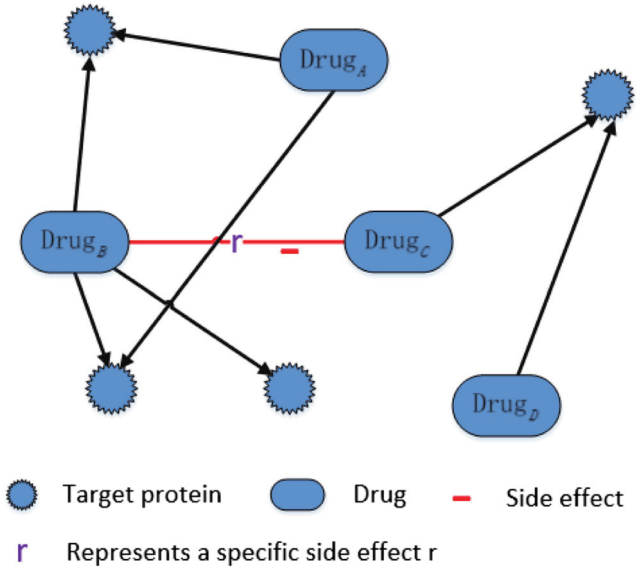

Fig. 4 Drug initial heterogeneity network. The figure is a heterogeneous DDI network with one kind of side effect

to represent the DDISN as follows.

$$A_{DDISN} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \tag{17}$$

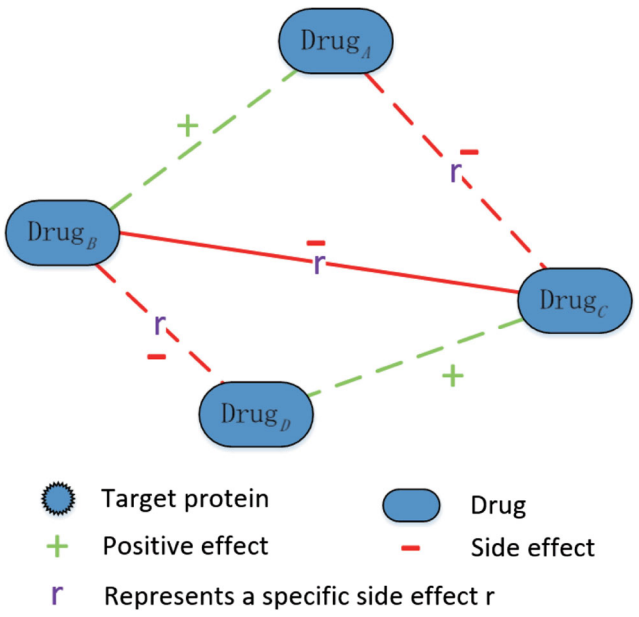

Fig. 5 Drug final DDI network. The solid line represents the known DDIN, and the dotted line represents the predicted DDIN based on the above definition

The SPM is initialized according to the above definition, as shown in (18).

$$SPM_1 = A_{DDISN} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \tag{18}$$

Using the signed propagation, the second-order SPM is obtained as follows.

$$SPM_2 = \text{Sign}(SPM_1 \times SPM_1) = \begin{pmatrix} 1 & 1 & \mathbf{-1} & 0 \\ 1 & 1 & -1 & \mathbf{-1} \\ \mathbf{-1} & -1 & 1 & 1 \\ 0 & \mathbf{-1} & 1 & 1 \end{pmatrix} \tag{19}$$

Through the second-order SPM, we can easily find that two more pairs of side effects than the initial DDISN, i.e., Drug$_A$ and Drug$_C$, Drug$_B$ and Drug$_D$ in Fig. 5 (red dotted lines). We find many potential reasonable relationships in SPM$_2$. Then, SPM$_2$ can be used as the initial expression of drug characteristics, which achieved good results in our experiment.

3.2 Process and algorithm

Here, we will elaborate on the process and algorithm of the proposed SC-DDIS.

3.2.1 Process

The process is described as follows.

Step 1. A DDIN with specific side effect r is extracted from the heterogeneous drug network, and adjacency matrix $A_{rDDIN} \in \mathbb{R}^{N\times N}$ is used to represent the DDIN, $G_{rDDIN} = (V, E, A_{rDDIN})$.

Step 2. Drug similarity threshold hyperparameter μ is set, and the DDISN is constructed based on semantic information. Here, adjacency matrix $A_{rDDISN} \in \mathbb{R}^{N\times N}$ is used to represent the DDISN $G_{rDDISN} = (V, E, A_{rDDISN}, A_{rDDIN}, S, \mu)$.

Step 3. The extended structural balance theory is used as a constraint condition for sign propagation, and the sign in the DDISN is propagated to obtain SPM_m as a higher-order expression of the drug feature.

Step 4. A_{rDDIN} is used as the initial adjacency matrix, and SPM_m is taken as the initial vector representation of the drug. Here, we use Xavier [8] to initialize the convolution kernel. Z_r of the drug feature is extracted using a graph convolution neural network.

Step 5. A decoding matrix of the drug side effect r is trained, and the decoding matrix $X_r \in \mathbb{R}^{N\times N}$ is

used to predict the side effect r between drugs. The decoding process is expressed as follows.

$$\hat{A}_r = \sigma(Z_r X_r Z_r^T) \text{ with } Z_r = \text{SC-DDIS}(SPM_m, A_{rDDIN}) \tag{20}$$

Here, $\hat{A}_r$ is the reconstructed adjacency matrix. $\hat{A}_r^{ij}$ is the probability of side effects between Drug_i and Drug_j, and σ is the sigmoid activation function, i.e., $\sigma(x) = \frac{1}{1+e^{-x}}$. By training the decoding matrix of side effect r, the noise data can be filtered out, and the vital drug feature information can be retained, thereby enhancing the effectiveness and robustness of the model.

Step 6. The loss function value is calculated. Note that the number of side effects due to DDIs is less than the total number of drug combinations. In other words, the dataset of drug combinations that produce side effects is imbalanced, i.e., there are fewer positive samples. In other words, there are fewer edges with ADR. To solve this problem, we use a weighted cross-entropy loss function to calculate the loss value. The loss of positive samples is increased according to the proportion of positive and negative examples. As a result, the overall cost is adjusted to minimize the impact of the dataset's imbalance on the proposed SC-DDIS.

$$Loss_r = \sum -\beta A_{rDDIN}^{ij} log\left(\hat{A}_r^{ij}\right) - \left(1 - A_{rDDIN}^{ij}\right) log\left(1 - \hat{A}_r^{ij}\right) \tag{21}$$

Here, $\beta = |Sa_-| / |Sa_+|$ is the ratio of negative samples to positive samples. $|Sa_-|$ is the number of negative samples, i.e., the number of edges with unknown effects in the DDIN. $|Sa_+|$ is the number of positive samples, i.e., the number of edges known to generate DDNE in the DDIN. Finally, the Adam algorithm [16] is used to optimize and complete the model's training task. The pseudocode for the proposed SC-DDIS's algorithm is presented in Algorithm 1.

Algorithm 1 Spectral convolution on drug-drug interactions signed network.

Input: Similarity threshold hyperparameters μ; Drug-drug interactions network adjacency matrix A_{rDDIN}; The number of iterations Ep of training model; The layer number L of graph convolution neural network; Heterogeneous information of drugs.

Output: Reconstituted adjacency matrix $\hat{A}_r$ of drug-drug interaction network

1: **for** $i = 0 \rightarrow n-1$ **do**
2: **for** $j = 0 \rightarrow n-1$ **do**
3: $\mathbf{S_{ij}} \leftarrow 0$
4: **for** rel in {protein, chemistry, ... } **do**
5: $\mathbf{S_{ij}} \leftarrow \alpha_{rel} |\Gamma_{rel}(i) \cap \Gamma_{rel}(j)| / |\Gamma_{rel}(i) \cup \Gamma_{rel}(j)|$
6: **end for**
7: **if** $S_{ij} > \mu$ and $A_{rDDIN}[i][j] = 0$ **then** $\mathbf{A_{rDDISN}}[i][j] \leftarrow 1$
8: **else**
9: **if** $S_{ij} \leq \mu$ and $A_{rDDIN}[i][j] = 0$ **then** $\mathbf{A_{rDDISN}}[i][j] \leftarrow 0$
10: **else** $\mathbf{A_{rDDISN}}[i][j] \leftarrow -1$
11: **end if**
12: **end if**
13: **end for**
14: **end for**
15: $\mathbf{SPM_0} \leftarrow I$
16: $\mathbf{SPM_1} \leftarrow A_{rDDISN}$
17: **for** $i = 2 \rightarrow m$ **do**
18: $\mathbf{SPM_i} \leftarrow SPM_{i-1} * SPM_1$
19: **end for**
20: $\mathbf{H^0} \leftarrow SPM_m$
21: $\beta \leftarrow |Sa_-| / |Sa_+|$
22: $\tilde{\mathbf{A}}_r \leftarrow A_{rDDIN} + I_r$
23: $\tilde{\mathbf{D}}_r(i,i) \leftarrow \sum_j \tilde{A}_r(i,j)$
24: **for** $epoch = 1 \rightarrow Ep$ **do**
25: **for** $l = 1 \rightarrow L$ **do**
26: $\mathbf{H^{(l)}} \leftarrow \sigma\left(\tilde{D}_r^{-\frac{1}{2}} \tilde{A}_r \tilde{D}_r^{-\frac{1}{2}} H^{(l-1)} W_r^{(l-1)}\right)$
27: **end for**
28: $\mathbf{Z_r} \leftarrow H^{(L)}$
29: $\hat{\mathbf{A}}_r \leftarrow \sigma\left(Z_r X_r Z_r^T\right)$
30: $\mathbf{Loss_r} \leftarrow \sum -\beta A_{rDDIN}^{ij} log\left(\hat{A}_r^{ij}\right) - \left(1 - A_{rDDIN}^{ij}\right) log\left(1 - \hat{A}_r^{ij}\right)$
31: Updating parameters(W_r and X_r) matrix by Adam.
32: **end for**

Here, lines 1 to 17 construct the DDISN based on semantic information by fusing a multi-source heterogeneous network of drugs and setting drug similarity threshold μ. Lines 18 to 22 perform signed propagation on the DDISN

based on the extended structural balance theory to obtain the SPM_m. Lines 23 to 26 initialize the SC-DDIS model parameters, and lines 27 to 35 perform Ep rounds of iterative training on the DDIN using the proposed SC-DDIS model. Lines 28 to 31 obtain the low-dimensional vector representation of the drug, and line 32 passes the decoding matrix to decode the drug code to obtain a reconstructed DDIN. Finally, lines 33 to 34 calculate the loss value of the proposed SC-DDIS model and update the parameter matrix.

4 Experiment

4.1 Datasets

In this study, we used the dataset in the Decagon model (http://snap.stanford.edu/decagon). This benchmark dataset includes 645 drug nodes, 7,795 protein nodes, 4,576,785 DDIs, and 18,690 DPIs. The dimension of DDIs is $645 \times 645 \times 964$ (964 ADR events), and the dimension of DPIs is $645 \times 7,795$ (18,690 interactions). If ADR occurs, the corresponding element in the DDIs dataset is marked as 1. Here, we focused on 964 common side effects, and each side effect occurs in at least 500 drug combinations; among these side effects, such as anaemia, diversity breathing, nausea and so on. Drug data statistics are shown in Table 1.

To test the generalization of the model, we introduce a new data set. The dataset can be obtained from the https://github.com/ltrbless/Dataset/tree/master website. In this dataset, 548 drug nodes and 97168 DDIs were obtained from TWOSIDES, and 881 chemical substructure types were obtained from PubChem. The type of side effect is not specified in this data set. If there are side effects between two drugs, the corresponding element in the DDIs dataset is marked as 1.

4.2 Experimental settings

We used TensorFlow 1.13.1 to implement the proposed SC-DDIS, and we used the Adam adaptive learning rate optimizer to train the model. Note that effective parameter values were determined experimentally. We then set the drug similarity threshold $\mu = 0.8$, parameter $rel \in \{\text{protein}\}$, and the proportion of target protein similarity weight $\alpha_{protein} = 1$. In addition, the learning rate was set to 0.01, we used a second-order SPM, and iteratively trained 200 times for each side effect. Here, we used a two-layer deep neural network with 32 and 16 dimensions, and the tanh function was used as the activation function between the first hidden layer and second hidden layer to extract the features of the SPM.

For the second new dataset, we will define the positive effect in terms of the similarity of chemical substructures. We set the drug similarity threshold $\mu = 0.8$, parameter $rel \in \{\text{chemical}\}$, and the proportion of target protein similarity weight $\alpha_{chemical} = 1$. Other settings are the same as above.

4.3 Baseline

We also used the latest Decagon model based on deep learning and other classic multi-relational link prediction approaches to evaluate and compare the proposed SC-DDIS.

1) Decagon model [21]: This graph convolution neural network is developed to predict multi-relation links in heterogeneous networks. In this model, end-to-end learning of DDIs, DPI, and protein-protein interactions are performed via graph convolution to obtain a drug feature expression. Then, ADRs are predicted.
2) RESCAL [22]: This is a relational learning method based on the factorization of a three-way tensor, which obtains predictions by decomposing the matrix of drug side effect r.
3) DEDICOM [23]: This technique uses tensor decomposition to provide useful potential information from DDIs for the prediction of ADRs. It is decomposed into $X_r = AU_rTU_rA^T$ by the DDIN X of known drug side effects r. The probability of side effects from Drug_i and Drug_j is calculated as $p_{ij} = a_iU_rTU_ra_j$.
4) DeepWalk [25, 31]: This technique is based on a biased random walk, and a low-dimensional feature representation of drug nodes is obtained by learning the neighborhood nodes of the drug-drug interaction networks. Then, the representations of each drug pair are spliced to represent the characteristics of each pair

Table 1 Drug data statistics

Data type	Data	Dimension
Node	Drug node	645
Node	Protein node	7795
Interaction	DDIs	645 × 645 (964 ADR events and 4,576,785 interactions)
Interaction	Target protein	645 × 7795 (18,690 interactions)

of drugs. Finally, an independent logistic regression classifier is trained for each side effect to predict the ADRs.

5) Concatenated drug features [21]: In this method, principal component analysis is used to reduce the dimension of DPI network, which is used as the expression of the initial drug characteristics. Then, the features of each pair of drugs are spliced as the characteristics of each pair of drugs. Finally, a gradient boosting tree classifier is employed to predict the ADRs of each drug pair.

4.4 Evaluation

We used the 10-fold cross-validation method to compare these algorithms. For a particular side effect, we randomly took 10% of the drug pairs with and without side effects as the test set and removed the known drug interaction relationship from the dataset. The other 90% of drug pairs was used as a training set to train the model. The final evaluation score for a specific side effect was taken as the average of all results repeated 10 times under different random partitions of the dataset. Finally, the average rating of 964 side effects was used as the final evaluation score.

4.5 Metrics

We use three commonly used metrics to evaluate the performance of the model: AUROC (area under the receiver operating characteristic curve), AUPRC (area under the precision-recall curve), and AP@K (average precision at K). The three evaluation criteria and related concepts are defined as follows.

Firstly, for a binary classification problem, the result of classification is positive (P) or negative (N). There are four scenarios in the forecast.

1) True Positive (TP): the prediction value is P, and the actual value is also P.
2) False Positive (FP): the prediction value is P, but the actual value is N.
3) True Negative (TN): the prediction value is N, and the actual value is also N.
4) False Negative (FN): the prediction value is N, but the actual value is P.

True positive rate (TPR) is the probability of positive samples among all positive samples, that is, model's sensitivity to positive samples. False positive rate (FPR) is the probability of positive samples in all negative samples, that is, model's sensitivity to negative samples. TPR and FPR are defined as follows.

$$TPR = \frac{TP}{TP + FN} \tag{22}$$

$$FPR = \frac{FP}{FP + TN} \tag{23}$$

Taking FPR as the x-axis and TPR as the y-axis, the Receiver Operating Characteristic Curve (ROC) can be obtained by setting different classification thresholds. AUROC is the area under the ROC curve. The larger the value of AUROC, the better the discrimination performance of the model. When the distribution of positive and negative samples in the test set changes, the ROC curve can keep stable, making the value of AUROC stable, so the AUROC index is robust.

Secondly, The concept of Precision is the proportion of correctly classified samples to the total samples. The definition of Recall is the same as that of TPR as the probability of positive samples. The definition of Precision and Recall is shown as follows.

$$Precision = \frac{TP}{TP + FP} \tag{24}$$

$$Recall = \frac{TP}{TP + FN} \tag{25}$$

With Recall as the x-axis and precision as the y-axis, the precision-recall curve (PRC) can be obtained by setting different classification thresholds. AUPRC is the area under PRC. It is noted that the AUPRC value can reflect the quality of the classifier more effectively than that of the AUROC, as the AUPRC can reflect the actual performance of the classification when the proportion of positive and negative samples is quite different.

Thirdly, the definition of average precision at K (AP @ K) is shown as follows.

$$AP@K = \frac{\sum_{i=1}^{K} Precision(i)}{min(L, K)} \tag{26}$$

$Precision(i)$ is the precision before position i in the test set's sorted prediction result. L is the total number of drug combinations with ADRs in the test set.

5 Results

5.1 Prediction results

As discussed previously, proposed SC-DDIS model is compared to the latest Decagon model and other classic multi-relational link prediction methods. The results are given in Table 2. Experimental results show that SC-DDIS performs better than other prediction methods under

Table 2 Average AUROC, AUPRC, and AP@50 in the prediction of 964 side effects

Approach	AUROC	AUPRC	AP@50
Concatenated drug features	0.793	0.764	0.712
DeepWalk neural embeddings	0.761	0.737	0.658
DEDICOM tensor factorization	0.705	0.637	0.567
RESCAL tensor factorization	0.639	0.613	0.476
Decagon	0.872	0.832	0.803
SC-DDIS	0.947	0.930	0.895

the three scoring methods and confirm the SC-DDIS correctness and effectiveness.

We further compare the performance of our SC-DDIS model with that of the Decagon model in detail, as the Decagon model performs the best in our baseline models. Concretely, we compare their best and worst performance in predicting 10 sorts of representative ADRs, respectively. As there is a massive proportion gap between the positive and negative samples in the dataset, the AUPRC value is used to evaluate the prediction results, as shown in Table 3.

Tables 2 and 3 show that the proposed SC-DDIS outperformed the compared models in all aspects, which verifies its effectiveness. For the second new dataset, the results are given in Table 4. Experimental results show that SC-DDIS performs better than other prediction methods and confirms the SC-DDIS correctness and effectiveness. To better reflect the difference in AP score, we set K to 2000. Due to the lack of protein-protein interactions in the new data set, the decagon method's results cannot be calculated. However, this does not affect the effectiveness of the proposed method. This is because we have proved that the SC-DDIS method outperforms the Decagon method in data sets hundreds of times richer than the new data set. The main purpose of introducing a second new dataset is to verify the effectiveness of using other drugs' biological profiles to calculate the drug-drug positive effect and verify our sign propagation's effectiveness and test the SC-DDIS model's generalization.

Table 4 shows that the SC-DDIS model proposed in this paper is superior to the comparison model in all aspects, which verifies its generalization, the effectiveness of using other drugs' biological profiles to calculate the drug-drug positive effect, and the effectiveness of our sign propagation.

5.2 Parameters analysis

In this section, we mainly investigate two parameters, the order of sign propagation m and the threshold μ. We select three side effects with fewer edges among the 964 types of side effects. These three side effects have relatively few edges; thus, they contain less information. As a result, they better reflect the performance of the proposed SC-DDIS model. Data for the three side effects are shown in Table 5.

We first test the effect of m to our model. Here, the experimental parameter settings were the same as above except for m. When $m = (0, 1, \ldots, 9)$, we calculate the change in AUROC and AUPRC scores for these three side effects, as shown in Fig. 6 and 7.

When $m = 0$, the SPM is the identity matrix, and, when $m = 1$, the SPM is a DDISN. It can be seen from the experimental results that when $m = 1$, that is, after

Table 3 Best/Worst AUPRC scores for Decagon model and SC-DDIS model

Best AUPRC	Decagon	SC-DDIS	Worst AUPRC	Decagon	SC-DDIS
Mumps	0.964	0.996	Bleeding	0.679	0.827
Carbuncle	0.949	0.994	Increased body temperature	0.680	0.839
Coccydynia	0.943	0.996	Emesis	0.693	0.856
Tympanic membrane perfor	0.941	0.988	Renal disorder	0.694	0.825
Dyshidrosis	0.938	0.997	Leucopenia	0.695	0.829
Spondylosis	0.929	0.980	Diarrhea	0.705	0.860
Schizoaffective disorder	0.919	0.993	Icterus	0.707	0.847
Breast dysplasia	0.918	0.978	Nausea	0.711	0.860
Ganglion	0.909	0.997	Itch	0.712	0.842
Uterine polyp	0.908	0.990	Anaemia	0.712	0.926

Table 4 AUROC and AUPRC, and AP@2000 scores

Scenario	AUROC	AUPRC	AP@2000
Concatenated drug features	0.845	0.851	0.829
DeepWalk neural embeddings	0.913	0.918	0.984
DEDICOM tensor factorization	0.775	0.799	0.824
RESCAL tensor factorization	0.776	0.800	0.827
SC-DDIS$_{(m=0)}$	0.863	0.896	0.981
SC-DDIS$_{(m=1)}$	0.905	0.923	0.990
SC-DDIS$_{(m=2)}$	0.928	0.935	0.990

Table 5 Number of edges in three sorts of ADRs data

Side effect name	The number of edges	The number of edge when $e(i, j) = e(j, i)$
Avascular necrosis	498	964
Carcinoma of the cervix	656	900
Phlebitis superficial	583	929

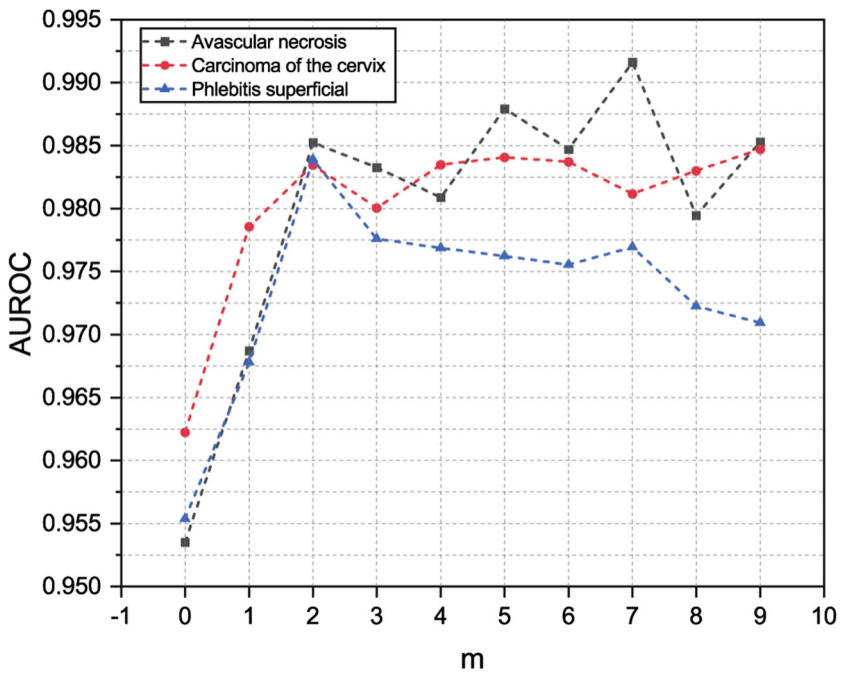

Fig. 6 AUROC scores of three side effects

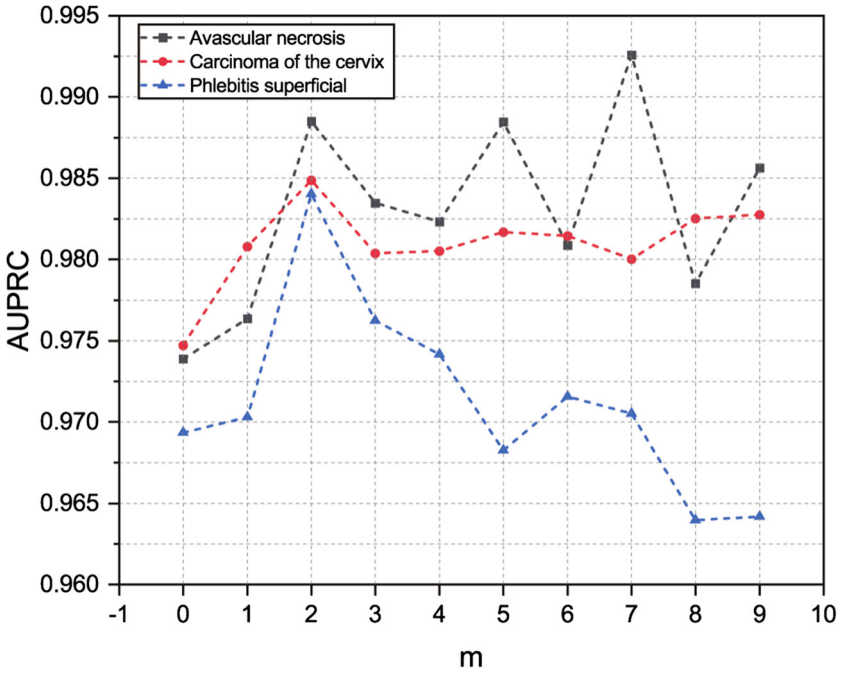

Fig. 7 AUPRC scores of three side effects

adding a positive effect between drugs to form a DDISN, the product is significantly improved. The empirical results show that considering the semantic information in the real world, medicines' characteristic information can be enriched, the side effect relationship network's sparseness can be reduced, and the prediction performance can be improved (Table 6).

When $m = 2$, the SPM is the first-order sign propagation based on the structural balance theory in the DDISN. As shown in Figs. 6 and 7, improvement to the score is obvious and stable for $m = 2$. According to the experimental results, we can find that after the first-order sign propagation, the effect has been significantly improved. The experience result also proves once again that the proposed SC-DDIS extends the structure balance theory of the signed network and applies it to learn higher-order sign propagation features is correct.

However, when the order is greater than 2, non-critical information may be introduced, thus making the the model's performance in-stable. Therefore, m was generally set to 2 to enhance the effectiveness and robustness of the model. Note that parameter m can be controlled flexibly for different side effects to achieve the best model effect. For example, the optimal prediction result for the avascular necrosis side effect is obtained when $m = 7$.

Then, we test the impact of μ on the model's performance. We used the above three side effects to test the impact varying μ change on the model's performance. Here, the parameter settings were the same as that of the original trial except for μ. When $\mu = (0.05, 0.1, \ldots, 0.99)$, we obtained changes to the AUROC and AUPRC scores for these three side effects.

From Figs. 8 and 9, we see that both scores are satisfactory enough when the value of $\mu > 0.8$. It is practically significant, as the larger the μ value,the higher the similarity of two drugs. Here, we consider that the two drugs have a positive effect when $S_{ij} > 0.8$ Therefore, a DDISN with larger μ is frequently more reliable. However, if the μ value is close to 1, less information about the positive effects between drugs is obtained thus making model performance decrease. This is why we set $\mu = 0.8$ in our experiments. For the second new dataset, we calculate

Table 6 AUROC and AUPRC scores of three side effects

Scenario	Avascular necrosis		Carcinoma of the cervix		Phlebitis superficial	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
m = 0	0.956	0.974	0.962	0.975	0.955	0.969
m = 1	0.969	0.976	0.979	0.981	0.968	0.970
m = 2	0.985	0.989	0.983	0.985	0.984	0.984

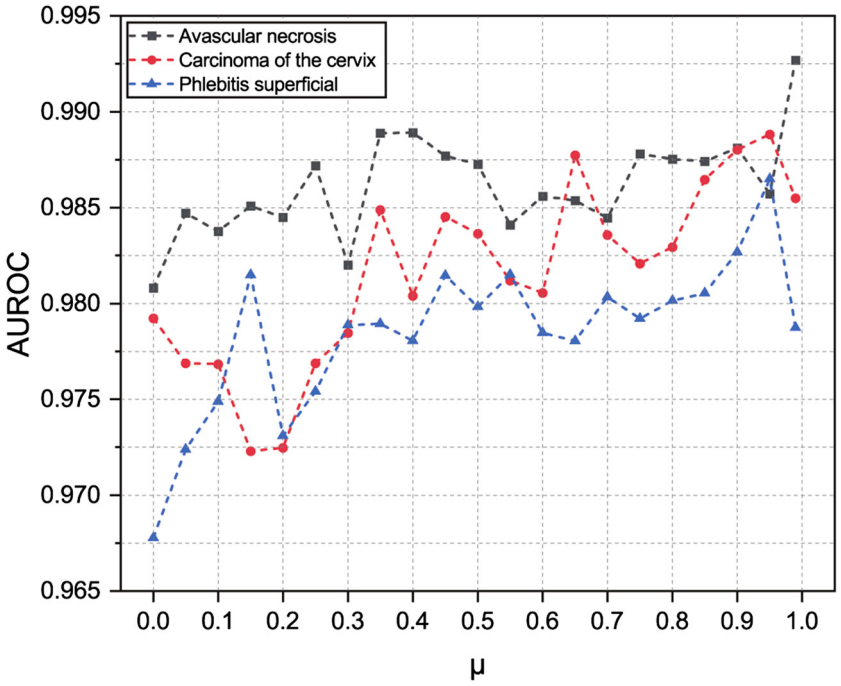

Fig. 8 AUROC scores of three side effects

the AUROC and AUPRC scores by changing the value of *m* to verify the effectiveness of the sign propagation matrix, as shown below (Figs. 10 and 11).

When m = 1, that is, the positive interaction between drugs is considered. After adding the positive action between drugs, the drug sign network will be formed. There were 17,146 pairs of positive reactions between drugs and 97,168 pairs of side effects in the drug sign network. The first-order drug sign network is used as the initial feature expression of drugs. According to the experimental results, it can be found that the performance of the model has been dramatically improved. Once again proved the generalization and effectiveness of our method. The reason for the performance improvement is the introduction of semantic information between drugs in the real world. That is, there are positive effects and side effects of drugs. When

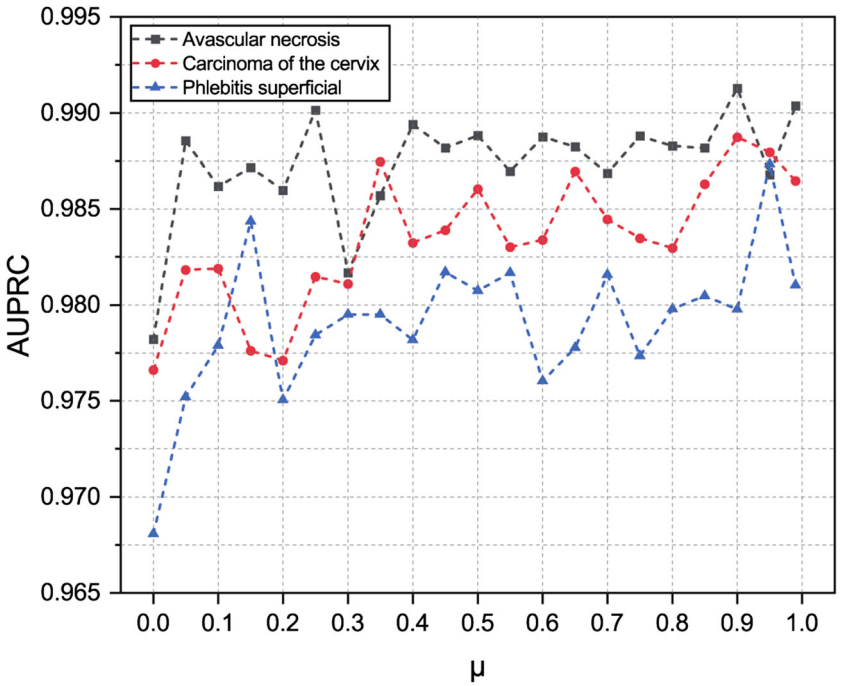

Fig. 9 AUPRC scores of three side effects

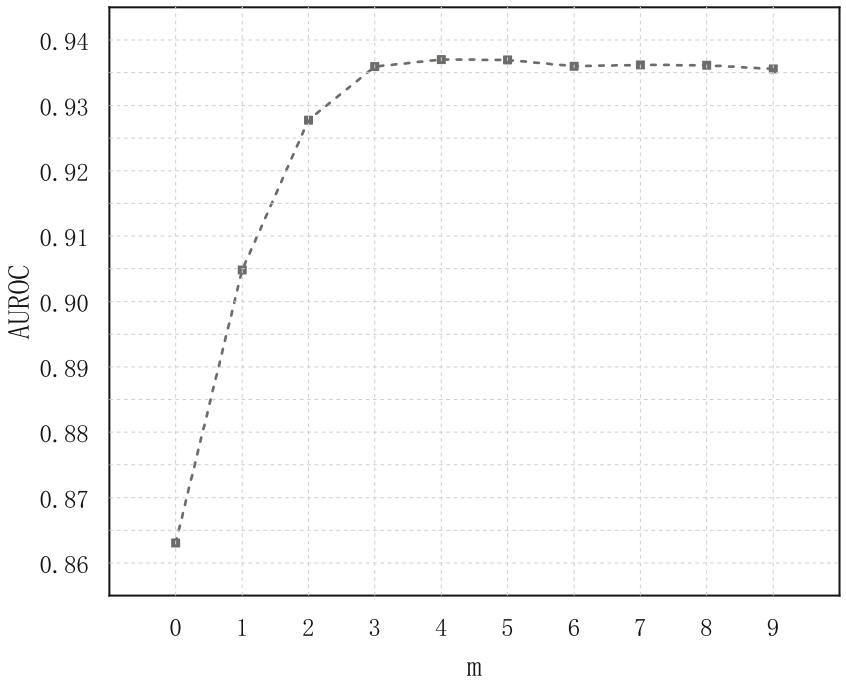

Fig. 10 AUROC scores

m = 2, based on extends the structure balance theory of the signed network, it is applied to learn the high-order symbol propagation characteristics. According to the experimental results, it is proved again that we can obtain high-order semantic information according to the extent of the signed network's structural balance theory, which can significantly improve the effect.

6 Conclusions

In this paper, we proposed the use of semantic information to integrate drug heterogeneous networks. Then, we applied an extended structure balance theory and sign propagation technology to improve the performance of predicting adverse drug reactions. A heterogeneous information

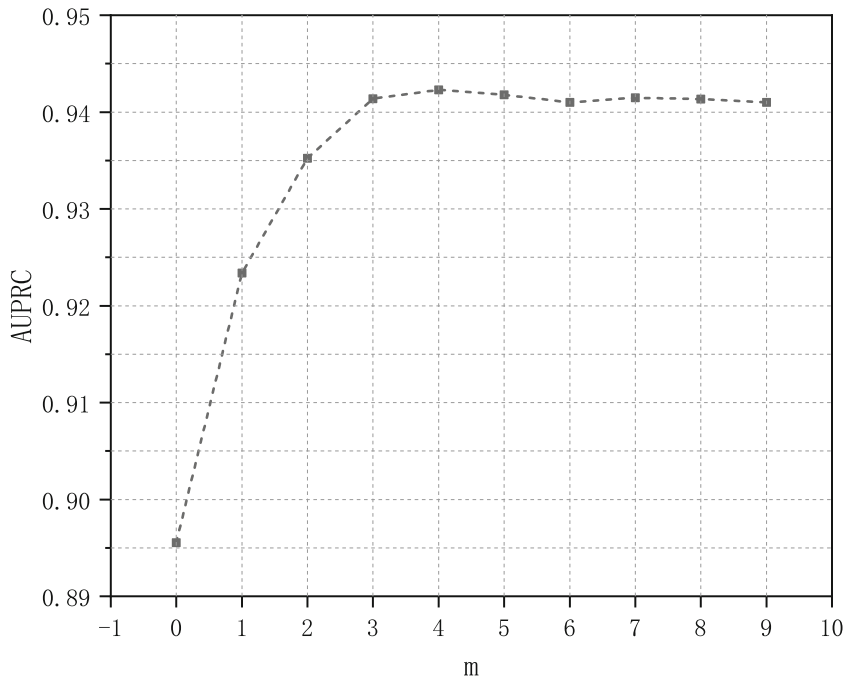

Fig. 11 AUPRC scores

network embedding method (i.e., the proposed SC-DDIS) based on spectral convolution was designed to fuse multi-source drug information. We developed a general fusion method based on Jaccard similarity to obtain sign network integrated information for different heterogeneous information. Then, the sign network's learning is performed via the extended structure balance theory and sign propagation to obtain more abundant information in drug networks, which significantly improves the model's prediction performance. A large number of experiments demonstrate that model parameters $\mu = 0.8$, and $m = 2$ yield excellent prediction performance in most cases. We further improve model robustness by training a decoding matrix to predict ADR, and a weighted cross-entropy loss function is used to calculate the loss value to reduce the negative impact of sample imbalance on the proposed SC-DDIS. We performed many ADR prediction task experiments and compared the results to those obtained by Decagon model and other classic multi-relational link prediction methods, and the results proved the effectiveness of SC-DDIS. We also analyzed the essential parameters of the proposed SC-DDIS model and factors that affect parameter selection.

In this study, we only considered the relationship between drugs and other information. Thus, in future, we plan to investigate how to add protein-protein interactions while ensuring rationality. To further solve the critical problems in ADR prediction, we also plan to consider how to make full use of semantic information to improve ADR prediction results' interpretability.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant 61672329 and 81273704, in part by the Project of the Shandong Provincial Project of Education Scientific Plan (No.SDYY18058).

References

1. A community computational challenge to predict the activity of pairs of compounds. Nature Biotechnology 32(12), 1213–1222
2. Cami A, Manzi S, Arnold A, Reis BY (2013) Pharmacointeraction network models predict unknown drug-drug interactions. PLOS ONE 8(4):1–9. https://doi.org/10.1371/journal.pone.0061468
3. Cartwright D, Harary F (1977) Structural balance: A generalization of heider's theory 1. Soc Netw 63(5):9–25
4. Çelebi R, Mostafapour V, Yasar E, Gümüs O, Dikenelli O (2015) Prediction of drug-drug interactions using pharmacological similarities of drugs. In: 2015 26th International workshop on database and expert systems applications (DEXA), pp 14–17
5. Chen X, Liu X, Wu J (2019) Drug-drug interaction prediction with graph representation learning. In: 2019 IEEE International conference on bioinformatics and biomedicine (BIBM), pp 354–361
6. D'Informatique D, Ese N, Esent P, Au E, Gers F, Hersch P, Esident P, Frasconi P (2001) Long short-term memory in recurrent neural networks. Epfl 9(8):1735–1780
7. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y (2007) When good drugs go bad. Nature 446(7139):975–977
8. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp 249–256
9. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC (2017) Synergistic drug combinations for cancer identified in a crispr screen for pairwise genetic interactions. Nature Biotechnology
10. Hu B, Wang H, Wang L, Yuan W (2018) Adverse Drug Reaction Predictions Using Stacking Deep Heterogeneous Information Network Embedding Approach. Molecules. 23(12):3193. https://doi.org/10.3390/molecules23123193. https://www.mdpi.com/1420-3049/23/12/3193
11. Hu B, Wang H, Yu X, Yuan W, He T (2017) Sparse network embedding for community detection and sign prediction in signed social networks. Journal of Ambient Intelligence & Humanized Computing 10 (1)1–12. https://doi.org/10.1007/s12652-017-0630-1
12. Huang LC, Wu X, Chen JY (2013) Predicting adverse drug reaction profiles by integrating protein interaction networks with drug structures. Proteomics 13(2):313–324
13. Jaccard P (1912) The distribution of flora in the alpine zone. N Phytol 11(2):37–50
14. Jia J, Zhu F, Ma X, Cao ZW, Li YX, Chen YZ (2009) Mechanisms of drug combinations: interaction and network perspectives. Nat Rev Drug Discov 8(6):516–516
15. Kastrin A, Ferk P, Leskošek B (2018) Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. PLOS ONE 13(5):1–23. https://doi.org/10.1371/journal.pone.0196865
16. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. Computer Science
17. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks
18. Liu H, Liu B, Zhang H, Li L, Qin X, Zhang G (2018) Crowd evacuation simulation approach based on navigation knowledge and two-layer control mechanism. Inform Sci 436–437:247–267
19. Liu R, Wang H, Yu X (2018) Shared-nearest-neighbor-based clustering by fast search and find of density peaks. Inform Sci 450:200–226
20. Liu S, Huang Z, Qiu Y, Chen YP, Zhang W (2019) Structural network embedding using multi-modal deep auto-encoders for predicting drug-drug interactions. In: 2019 IEEE International conference on bioinformatics and biomedicine (BIBM), pp 445–450
21. Marinka Z, Monica A, Jure L (2018) Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics 34(13):i457–i466
22. Nickel M, Tresp V, Kriegel HP (2011) A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th international conference on machine learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011
23. Papalexakis EE, Faloutsos C, Sidiropoulos ND (2016) Tensors for Data Mining and Data Fusion: Models, Applications, and Scalable Algorithms. Association for Computing Machinery, New York, NY, USA 8(2) 44. 2157–6904. https://doi.org/10.1145/2915921

24. Park K, Kim D, Ha S, Lee D (2015) Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks. Plos One 10(10):e0140816
25. Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '14, 701–710, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2623330.2623732
26. Qin X, Liu H, Zhang H, Liu B (2018) A collective motion model based on two-layer relationship mechanism for bi-direction pedestrian flow simulation. Simul Modell PractTheory 84:268–285
27. Tang J, Chang Y, Aggarwal C, Liu H (2015) A survey of signed network mining in social media. ACM Computing Surveys
28. Wang XX, Li JB (2005) Method of computing accessibility matrix from adjacency matrix. Journal of Jilin Institute of Chemical Technology
29. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X (2017) Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. Bmc Bioinformatics 18(1):18
30. Zheng Y, Peng H, Ghosh S, Lan C, Li J Inverse similarity and reliable negative samples for drug side-effect prediction. BMC Bioinformatics 19(13) 1471–2105. https://doi.org/10.1186/s12859-018-2563-x
31. Zong N, Hyeoneui K, Victoria N, Olivier H (2017) Deep mining heterogeneous networks of biomedical linked data to predict novel drugtarget associations (15)15. Bioinformatics 33(15):1367–4803. https://doi.org/10.1093/bioinformatics/btx160

Affiliations

Taoran Liu[1] · Jiancong Cui[1] · Hui Zhuang[1] · Hong Wang[1,2]

Taoran Liu
ltrbless@163.com

Jiancong Cui
201711010104@sdnu.edu.cn

Hui Zhuang
sdnu_zh@163.com

[1] School of Information Science and Engineering, Shandong Normal University, Jinan, 250358, China

[2] Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, Shandong Normal University, Jinan, 250358, China