Mathematical Biosciences and Engineering

MBE, 19(6): 5754–5771.
DOI: 10.3934/mbe.2022269
Received: 10 January 2022
Revised: 20 February 2022
Accepted: 07 March 2022
Published: 06 April 2022

http://www.aimspress.com/journal/MBE

*Research article*

# Identification of drug side effects with a path-based method

**Meng Jiang[1,†], Bo Zhou[2,†] and Lei Chen[1,*]**

[1] College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China
[2] Shanghai University of Medicine & Health Sciences, Shanghai 201318, China

† These authors contributed equally to this work.

* **Correspondence:** Email: chen_lei1@163.com; Tel: +862138282825; Fax: +862138282800.

**Abstract:** The study of drug side effects is a significant task in drug discovery. Candidate drugs with unaccepted side effects must be eliminated to prevent risks for both patients and pharmaceutical companies. Thus, all side effects for any candidate drug should be determined. However, this task, which is carried out through traditional experiments, is time-consuming and expensive. Building computational methods has been increasingly used for the identification of drug side effects. In the present study, a new path-based method was proposed to determine drug side effects. A heterogeneous network was built to perform such method, which defined drugs and side effects as nodes. For any drug and side effect, the proposed path-based method determined all paths with limited length that connects them and further evaluated the association between them based on these paths. The strong association indicates that the drug has a side effect with a high probability. By using two types of jackknife test, the method yielded good performance and was superior to some other network-based methods. Furthermore, the effects of one parameter in the method and heterogeneous network was analyzed.

**Keywords:** drug; side effect; heterogeneous network; path; chemical-chemical interaction

## 1. Introduction

Adverse drug reactions (ADRs), also called side effects, refer to the pharmacological effects of drugs beyond the purpose of treatment that occur after the standard dose of the drug is taken. The unaccepted side effects of drugs mainly cause the failure of drugs in clinical trials and withdrawal from

the market. In each year, approximately 2 million people are affected by the side effects of drugs [1,2]. According to some public information, a drug is released to the consumer market after 10–15 years from the initial stage. The entire process is highly expensive. Although the drug research and development market has invested and progressed in the past 20 years, the number of new drugs approved each year remains low [3]. To save time, reduce cost and risk, and accelerate the procedures of drug discovery, researchers should study drug side effects. However, the traditional experiment and clinical methods are very time-consuming and laborious. Therefore, fast, efficient, and low-cost methods should be designed urgently.

In recent years, several essential progresses have been made in computer technology. These newly proposed technologies provide strong supports to deal with some realities, including drug side effect prediction. The accumulative knowledge on drugs and side effects, which have been stored in some public databases, provides strong data support to develop efficient computational methods for the identification of drug side effects. Several computational methods have been set up to identify drug side effects. In some studies, a binary classifier for each side effect has been built [4–8]. Considering the increasing number of side effects that have been reported, the efficiency of such type of methods is a problem. Several other studies developed multi-label classifiers to identify drug side effects caused by the multiple side effects of several drugs [9–14]. Some studies paired drugs and side effects as samples and built a uniform binary classifier involving all side effects [15–17]. The prediction of drug side effects was also modelled as recommended systems in some studies [2,18,19]. Most previous studies seldom considered the associations of drugs and side effects. Networks have been used to organize drugs or side effects in some methods. However, the networks for both drugs and side effects have not been included, hindering the deep investigation on the associations of drugs or side effects. The heterogeneous network is a good form to include drugs and side effects simultaneously. This study adopted such form and further designed a novel computational method to identify drug side effects.

In this study, a new computational method was built to identify drug side effects. A heterogeneous network that defines drugs and side effects as nodes was constructed. On such network, a path-based method was designed. For each drug-side effect pair, we obtained all paths connecting two nodes, corresponding to the drug and side effect, with limited length. Based on these paths, we computed a measurement to evaluate the strength of the association between such drug and side effect. The strong association suggested that the drug had the side effect with a high probability. Two types of jackknife test were employed to assess the performance of the method, inducing satisfied AUROC and AUPR. The path-based method was also superior to some methods that incorporate random walk with restart (RWR) [20] or Laplacian heat diffusion (LHD) [21] algorithms. Finally, novel side effects of some drugs yielded using the path-based method were analyzed.

## 2. Materials and methods

### *2.1. Materials*

We sourced the drugs and their side effects from Side Effect Resource (SIDER) database (http://sideeffects.embl.de/, version 4.1) [22], a public database containing information about marketed medicines and their recorded ADRs. Based on the original downloaded file "meddra_all_se.tsv.gz", 1,556 drugs represented by STITCH compound IDs and 6,123 different side effects were obtained. This investigation was conducted with a heterogeneous network, drugs and side effects were defined

as nodes, and drugs and side effects that were not included in such network were excluded. After such operation, 1,489 drugs and 6,028 side effects remained. From these drugs and side effects, 164,047 validated drug-side effect pairs were obtained and called as positive samples. The remaining drug-side effect pairs were not been labelled, and whether the drug had a side effect was not determined. These pairs were termed as negative samples when evaluating the performance of methods.

### *2.2. Heterogeneous network construction*

A network is a well-defined scheme used to organize several objects at the system level. Under this form, the relationships between objects are clearly displayed to mine hidden information of different objects. Here, we adopted such form to investigate drug side effects.

The constructed network defined drugs and side effects as nodes. Such type of network is called a heterogeneous network in computer science. Two types of nodes are included in this network, and edges in such network can be classified into the three following types: (1) edges connecting two drugs; (2) edges connecting two side effects; and (3) edges connecting one drug and one side effect. These three edge types induced three sub-networks, namely, drug network, side effect network, and drug–side effect network. The construction procedures are as follows.

#### 2.2.1. Drug network

This network was constructed by downloading the chemical–chemical interaction (CCI) information collected in STITCH (http://stitch.embl.de/, version 4.0) [23]. Each CCI is composed of two chemicals, represented by STITCH compound IDs, and one confidence score, ranging from 1 to 999. This score was obtained by integrating four scores, which measure associations between chemicals based on their structures, activities, reactions, and co-occurrence in literature. The score derived from chemical structures is defined as the Tanimoto 2D chemical similarity score based on chemicals' Simplified Molecular Input Line Entry System (SMILES) [24] strings, which is widely used in chemical- or drug-related studies. Thus, the integrated score reported in STITCH can extensively measure the associations of chemicals. Based on the downloaded CCI information, we extracted those involving 1,489 drugs mentioned in Section 2.1, and 151,316 CCIs were obtained. Each of these CCIs formed an edge in the drug network, that is, two chemicals were adjacent if and only if they can constitute a CCI. Furthermore, each edge was assigned a weight, defining as the confidence score of the corresponding CCI divided by 1,000.

#### 2.2.2. Side effect network

This network contained 6,028 side effect nodes. Edges were determined in this network by measuring the associations between any two side effects. For side effect $s$, let $D(s)$ be a set containing drugs having such side effect. Accordingly, the associations between two side effects $s_i$ and $s_j$ can be evaluated based on their drug sets, and the expression is as follows:

$$w(s_i, s_j) = \frac{\left|D(s_i) \bigcap D(s_j)\right|}{\left|D(s_i) \bigcup D(s_j)\right|} \tag{1}$$

We set a threshold of 0.1 in such association to define edges in the side effect network, in which two side effects $s_i$ and $s_j$ were connected by an edge if and only if $w(s_i, s_j) > 0.1$. In addition, $w(s_i, s_j)$ was assigned to the edge as its weight. This side effect network contained 596,165 edges.

#### 2.2.3. Drug-side effect network

Finally, the drug–side effect network defined drugs and side effects as nodes, and each edge connected one drug node and one side effect node. We used the positive samples, namely, the validated drug-side effect pairs, mentioned in Section 2.1 to define edges. One drug node and one side effect node were connected if they can constitute a positive sample. The edges in this network indicated the current known relationships between drugs and side effects. Each edge was assigned a weight of 1. This network contained 164,047 edges in total.

The three networks above were combined to comprise the heterogeneous network. A total of 911,528 edges were contained in such network. For convenience, this network was denoted as $N_H$.

### *2.3. Path-based method*

In Section 2.2, a heterogeneous network $N_H$ was constructed, from which the associations between any drug and any side effect can be assessed. Path is an important definition in network theory. Two nodes connected by a path imply the special associations between them. Such association can be measured according to the paths that connect them. Each of these paths provides its contribution, which is influenced by its length. Let $P_1$ be a path with length one connecting two nodes, in which two nodes are directly connected by an edge. This path indicates the direct relationship between these two nodes. Let $P_2$ be another path with length two that connect them, this path indicates indirect relationship between them. Evidently direct relationship is stronger than the indirect relationship. Thus, when measuring the association of these two nodes, path $P_2$ provides less contribution than $P_1$. This finding indicates that the contribution of a path generally follows a decreasing trend with the increasing of its length. Considering that each path connecting two given nodes can provide its contribution, the association of two nodes can be measured by aggregating the contributions of all paths connecting them. In the presence of several paths that connect two nodes, the aggregation of the contributions of these paths is large, indicating high association between them. However, when computing the contribution of each path, the path length should be considered as above arguments. Therefore, we designed the path-based method in the following manner.

For one drug node $d$ and one side effect node $s$, the paths connecting them in $N_H$ were picked up. However, obtaining all paths that connect two nodes is an NP-hard problem. We set a limitation on the path length, which was denoted by $L$, that is, only paths with length of no more than $L$ were considered. Long paths slightly contribute to measure the associations of two nodes. Accordingly, such limitation was set. Let $P_L$ be a set containing all paths connecting nodes $d$ and $s$ and with length of no more than $L$. For each path $P$ in $P_L$, the weight is defined as the product of weights of edges in $P$. Such weight was denoted as $w(P)$. Considering that the edge weights were all between 0 and 1, $w(P)$ was much small if its length is high, thus supporting that short paths result in higher contribution for measuring the associations of two nodes than long paths. Accordingly, the associations of $d$ and $s$ were measured using the following expression:

$$Q(s,d) = \sum_{p \in P_L} (w(p))^{F_{decay}(p)}, \tag{2}$$

where $F_{deca}(p)$ represents the decay function, which can further increase the influence of path length, as defined below.

$$F_{decay}(p) = \theta \cdot l(p), \tag{3}$$

where $\theta$ = 2.26 as suggested in [25,26,27,28], and $l(p)$ represents the length of path $P$. The higher the $Q(s,d)$ value, the stronger the associations between $s$ and $d$.

### *2.4. Jackknife test*

Jackknife test was employed to evaluate the performance of the path-based method. Two cases were used for the prediction of drug side effects in terms of whether the side effects for a given drug are known or not. In the first case, some side effects of a given drug were validated, the method was used to discover novel side effects of this drug. For the second case, the side effects of an input drug were completely unknown, and the method was used to identify its side effects based on the side effects of other drugs. Thus, the performance of path-based method under two cases should be evaluated.

For the first case, we singled out each drug-side effect pair one by one. When one drug-side effect pair was picked up, the corresponding edge present in the heterogeneous network $N_H$ was deleted. Then, the path-based method was performed on such network to evaluate the association of the drug and side effect measured using Eq (2). The test under such case was called local jackknife test. For the second case, the pairs of one drug and all its side effects were all singled out. Under such case, the edges that connect this drug node and all side effect nodes were deleted from $N_H$. Then, the proposed path-based method was performed on this network to assess the associations between this drug and all side effects, which were also measured by Eq (2). The test under the second case was termed as global jackknife test.

### *2.5. Performance measurement*

Through the local or global jackknife test, each drug-side effect was assigned a measurement (Eq (2)). Accordingly, all drug-side effect pairs can be sorted in decreasing order of their measurements. For a given threshold, a prediction can be made for each drug-side effect pair, and the pair with higher measurement than the threshold was predicted to be positive; otherwise, it was predicted to be negative. Accordingly, the true positive rate (same as recall), false positive rate, and precision can be computed. By setting a series of thresholds, a group of the above measurements can be obtained. Receiver operating characteristic (ROC) was plotted by setting the true positive rate into the Y-axis and false positive rate into the X-axis. Furthermore, the precision-recall (PR) curve was drawn with precision in the Y-axis and recall in the X-axis. We calculated the area under these two curves, called AUROC and AUPR, to assess the performance of path-based method. High AUROC or AUPR indicated the good performance of the method. However, these two measurements may lead to different results. Accordingly, we computed the mean of AUROC and AUPR to fully evaluate the path-based method with different parameters and other methods. Subsequently, a uniform result can be obtained.

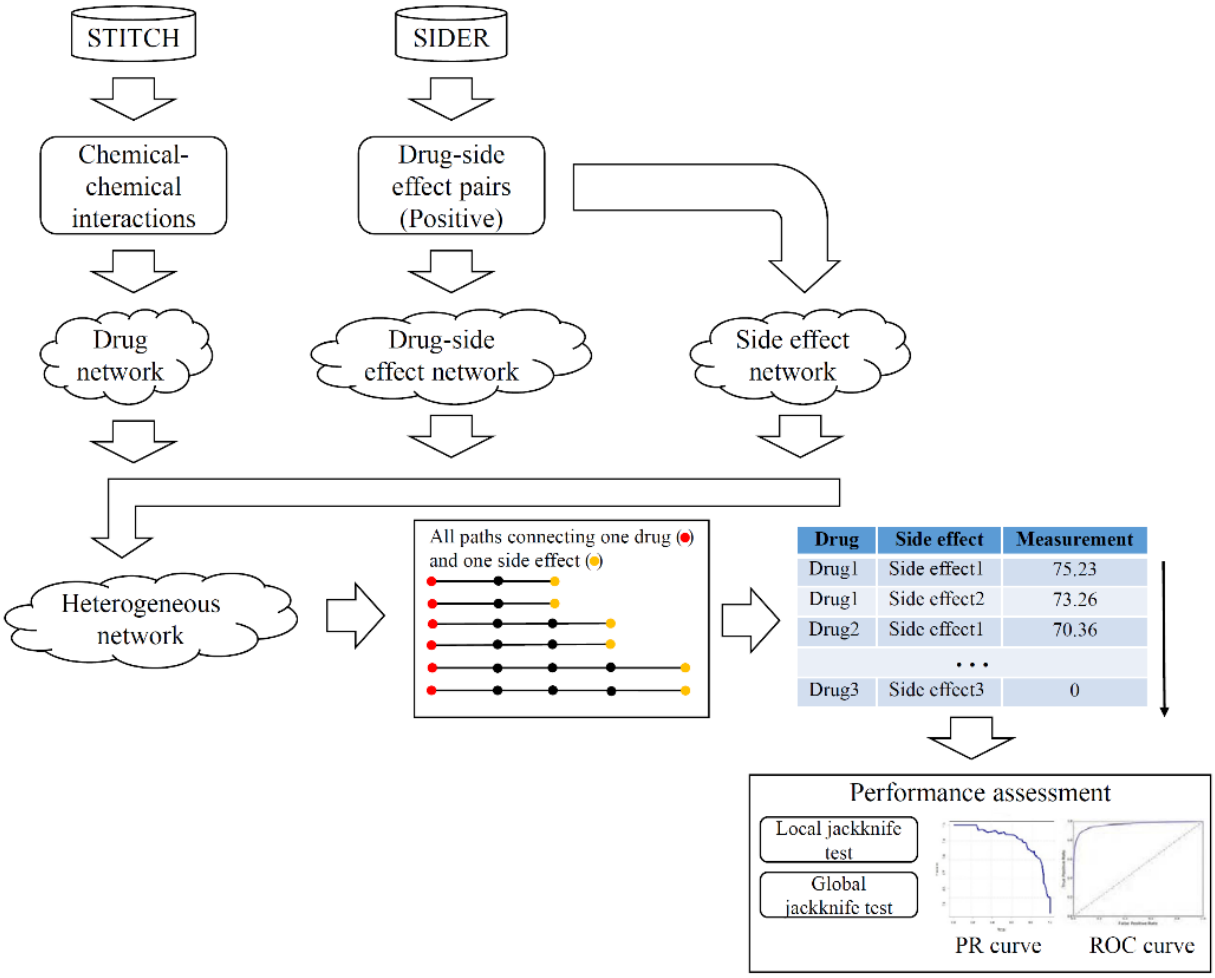


**Figure 1.** Entire procedures of the path-based method for the identification of drug side effects. A heterogeneous network containing three networks, namely, drug network, drug-side effect network, and side effect network, was constructed. For one drug and one side effect, all paths with limited length connecting them are extracted from the network. These paths were used to evaluate the association between the drug and side effect. Two types of jackknife test were adopted to evaluate the performance of path-based method, inducing ROC and PR curves.

## 3. Results

In the present study, a novel path-based method was built to identify drug side effects. The procedures are shown in Figure 1. This section provides the evaluation results and elaborated the superiority of the path-based method.

### *3.1. Performance of the method*

Before executing the path-based method, the parameter $L$ on the limitation of path length should be determined. In the present study, we set $L = 2$, that is, we only considered the paths that connect any drug and side effect with length of no more than 2. Two types of jackknife test were employed to

evaluate the performance of such method. The results induced one ROC and one PR curve, as illustrated in Figures 2 and 3. Furthermore, we calculated the area under these two curves, as shown in Figures 2 and 3 and Table 1. Based on the local jackknife test, the path-based method yielded AUROC of 0.9116 and AUPR of 0.4533. Based on the global jackknife test, the AUROC was 0.8662 and AUPR was 0.3353. Therefore, the performance under local jackknife test was better than that under global jackknife test. This finding is reasonable, because the side effect information of the drug in the testing sample was included in the training dataset under local jackknife test, whereas such information was excluded under the global jackknife test.

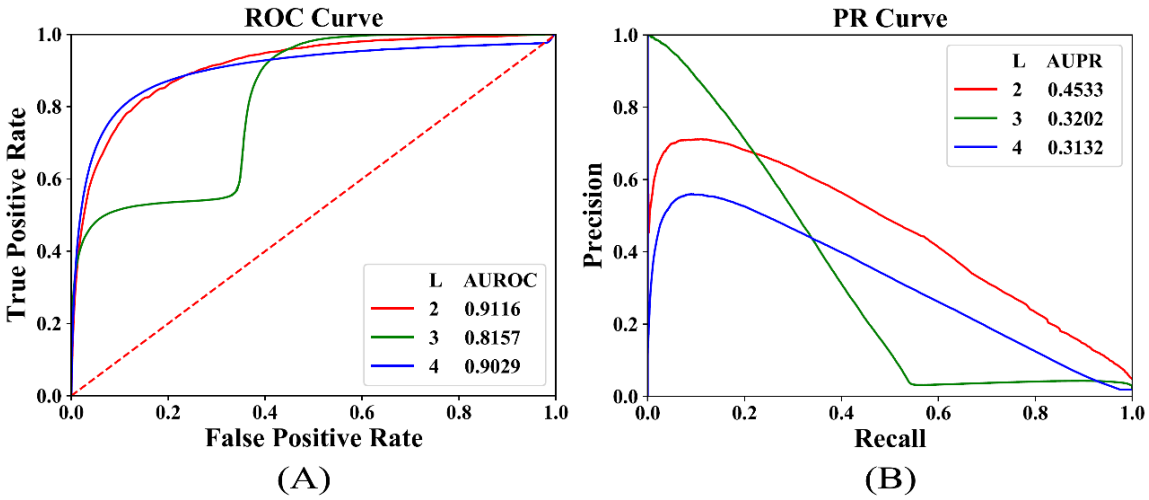


**Figure 2.** ROC and PR curves of path-based method with different parameters ($L$) on path length limitation under local jackknife tests. (A) ROC curves; (B) PR curves.

**Table 1.** Performance of the path-based method under local and global jackknife tests.

| Path length limitation | Local jackknife test | | | Global jackknife test | | |
|---|---|---|---|---|---|---|
| ($L$) | AUROC | AUPR | Average | AUROC | AUPR | Average |
| 2 | 0.9116 | 0.4533 | 0.6825 | 0.8662 | 0.3353 | 0.6008 |
| 3 | 0.8157 | 0.3202 | 0.5680 | 0.9178 | 0.2224 | 0.5701 |
| 4 | 0.9029 | 0.3132 | 0.6081 | 0.8668 | 0.1883 | 0.5276 |

### *3.2. Effect of the parameter on path length limitation*

In the proposed path-based method, the parameter ($L$) on path length limitation was set to 2. Other values such as $L = 3$ and $L = 4$ were also investigated. For each method with different values of such parameter, two types of jackknife tests were used to assess the performance. The corresponding ROC and PR curves are illustrated in Figures 2 and 3. The AUROC and AUPR values are listed in Table 1. The AUPR value decreased with increasing $L$. However, the values of AUROC were quite strange. $L = 2$ yielded the highest AUROC for local jackknife test, while $L = 3$ produced the highest AUROC for global jackknife test. For a full comparison, we further calculated the average AUROC and AUPR values for methods with different values of $L$, as listed in Table 1. $L = 2$ always yielded the highest averages regardless of the local or global jackknife tests. Thus, we set $L = 2$ in the proposed path-based method.

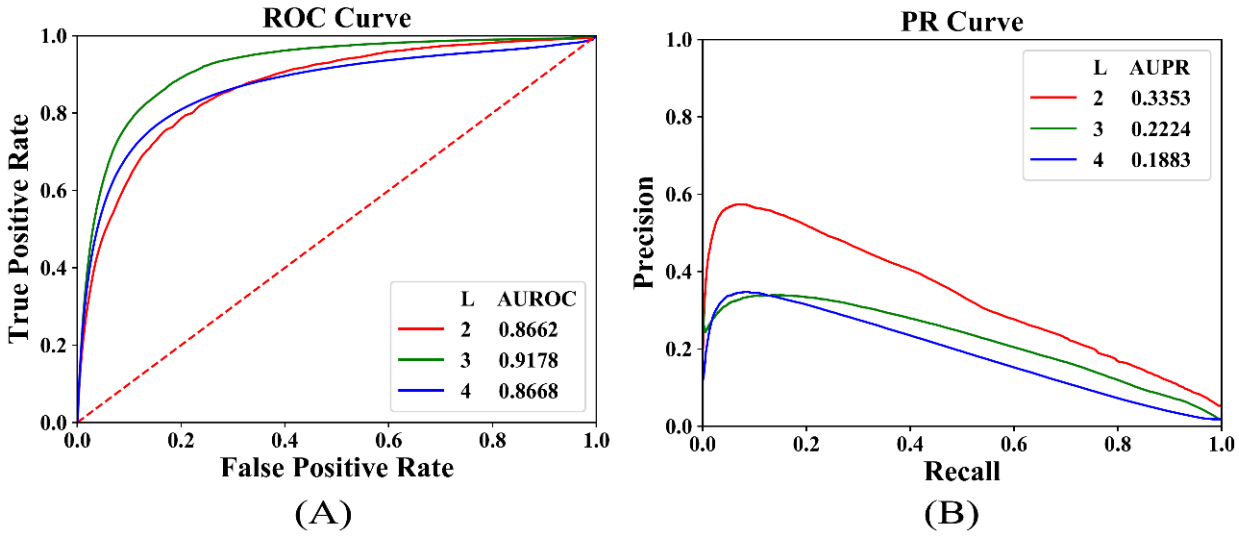


**Figure 3.** ROC and PR curves of path-based method with different parameters ($L$) on path length limitation under global jackknife tests. (A) ROC curves; (B) PR curves.

### *3.3. Effect of the networks*

The proposed path-based method for the identification of drug side effects was implemented on a heterogeneous network $N_H$. The accuracy of the network may affect the performance of the method. This section provides several permutation tests to confirm this finding.

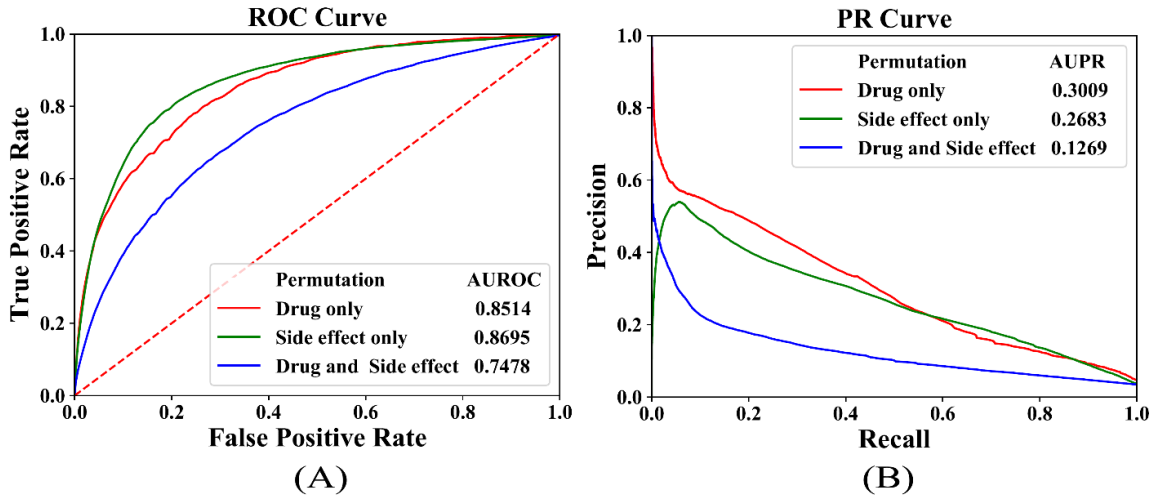


**Figure 4.** ROC and PR curves of path-based method ($L$ = 2) under different permutation tests and local jackknife test. (A) ROC curves; (B) PR curves.

Three permutation tests were conducted. Each test resulted in a permutation on part or all nodes in $N_H$. The first test permutated the drug nodes in $N_H$, whereas the side effects nodes did not change. The second test was conducted for the side effect nodes, in which a permutation was done on side effects nodes, and the drug nodes were not changed. The last test resulted in a permutation on both

drug and side effect nodes. Notably, the relationships between drugs and side effects, such as the edges in drug-side effect network, did not change in the three permutation tests. Under each permutation test, the path-based method ($L$ = 2) was assessed using local and global jackknife tests. Based on the outcomes, ROC and PR curves were plotted, as shown in Figures 4 and 5. Corresponding AUROC and AUPR values are listed in Table 2, showing that the performance of the path-based method descended compared with its original performance (Table 1). For local jackknife test, the performance of the method was almost at the same level when permutation test was applied on drug or side effect nodes. However, this finding was not observed case for global jackknife test. The permutation test on drug nodes provided much more influence on the method than the test on side effect nodes. Therefore, the heterogeneous network is important for the path-based method.

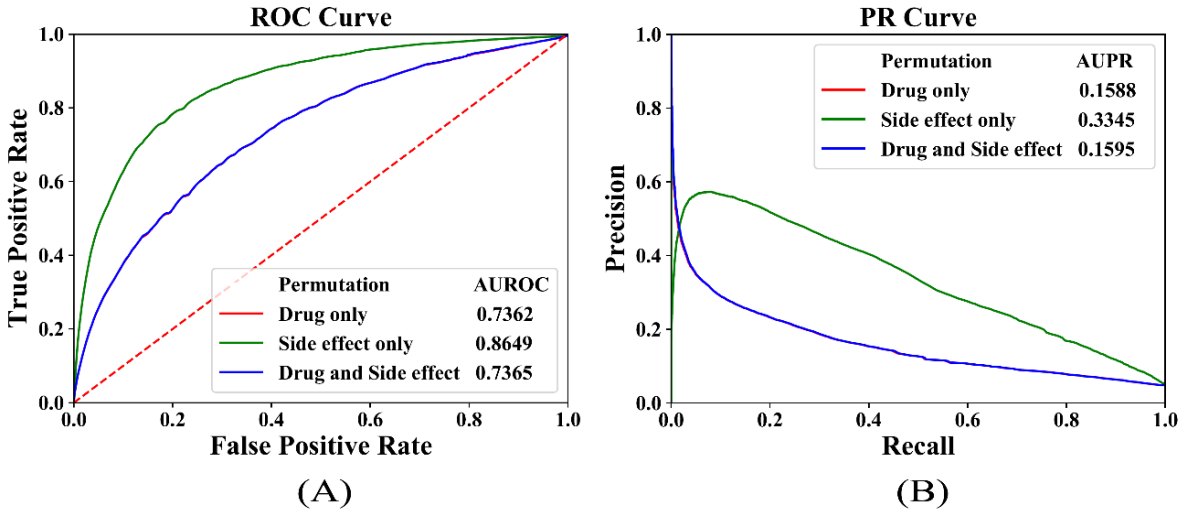


**Figure 5.** ROC and PR curves of path-based method ($L$ = 2) under different permutation tests and global jackknife test. (A) ROC curves; (B) PR curves.

**Table 2.** Performance of path-based method ($L$ = 2) under different permutation tests.

| Cross-validation | Permutation test | AUROC | AUPR | Average |
|---|---|---|---|---|
| Local jackknife test | Drug only | 0.8514 | 0.3009 | 0.5762 |
| | Side effect only | 0.8695 | 0.2683 | 0.5689 |
| | Drug and side effect | 0.7478 | 0.1269 | 0.4374 |
| Global jackknife test | Drug only | 0.7362 | 0.1588 | 0.4475 |
| | Side effect only | 0.8649 | 0.3345 | 0.5997 |
| | Drug and side effect | 0.7365 | 0.1595 | 0.4480 |

### *3.4. Comparison of other network-based methods*

In this study, a path-based method was proposed to identify drug side effects. This method provided satisfactory performance as mentioned in Section 3.1. To further confirm its superiority, we employed two network-based methods for comparison.

The first network-based method employed the RWR algorithm [20], which has wide applications in the field of bioinformatics [29–33]. This algorithm simulates a walker starting from one or more

seed nodes and delivers the probabilities on seed nodes to other nodes in a network. Based on $n$ seed nodes, a probability vector $P_0$ was constructed with a length that is the same as the node number in the network. One component corresponds to one node. The component corresponding to a seed node was set to $1/n$, whereas others were set to zero. Such vector was repeatedly updated in the following manner

$$P_{t+1} = (1-r)W^T P_t + rP_0, \tag{4}$$

where $W$ represents the column-wise normalized adjacency matrix of network, $r$ denotes the restarting probability, which was set to 0.15 in this study. When $P_{t+1}$ and $P_t$ are close enough, as measured by $\|P_{t+1} - P_t\|_{L_1} < 10^{-6}$, the updating procedures stop. $P_{t+1}$ is outputted as the outcome of the RWR algorithm. From such outcome, a probability can be obtained for each node other than seed nodes. This probability can reflect the association between the corresponding node and seed nodes. For the identification of drug side effects, the RWR algorithm was executed on the drug network (Section 2.2.1) for each side effect. Given a side effect $s$, we first extracted all drugs with such side effect, comprising a drug set denoted by $D(s)$. Jackknife test was performed by first singling out each drug in $D(s)$ individually and feeding the remaining drugs in $D(s)$ into the RWR algorithm as the seed nodes. When the RWR algorithm stopped, the probability on the singled out drug was picked up to measure its association to the side effect, that is, the association between the drug and the side effect. For drugs not in $D(s)$, all drugs in $D(s)$ were set as seed nodes and fed into the RWR algorithm to access their associations to the side effect. When all side effects had been considered, the pairs of drugs and side effects were ranked in decreasing order of associations. This further induced the ROC and PR curves, as shown in Figure 6. Their AUROC and AUPR are listed in Table 3. The AUROC was 0.5432 and AUPR was 0.0193. The average was only 0.2813. They were all much smaller than those of the path-based method ($L$=2, Table 1), indicating the superiority of the path-based method.

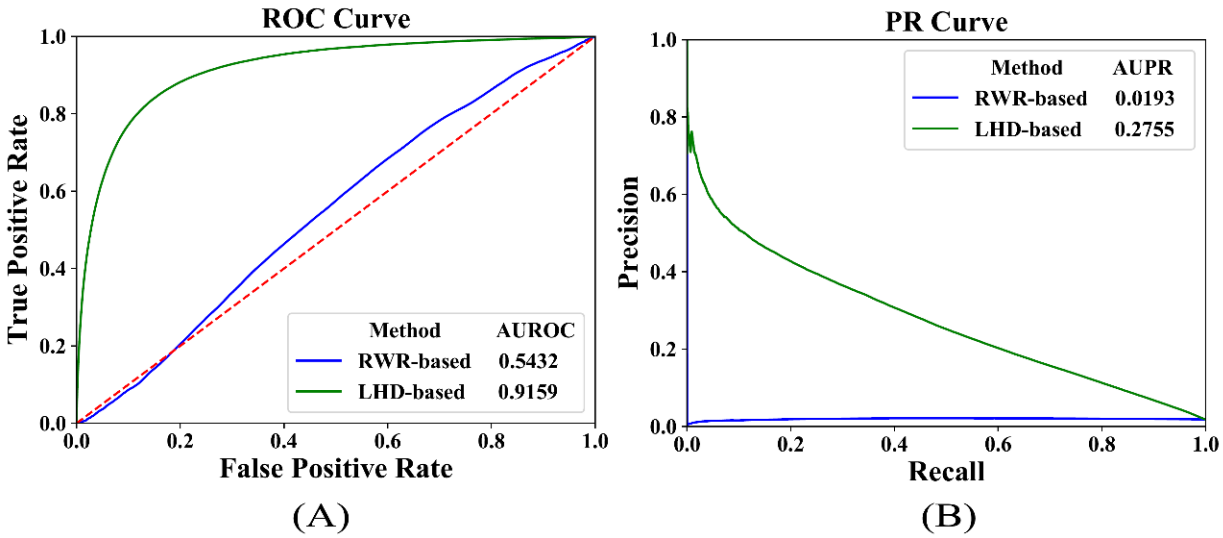


**Figure 6.** ROC and PR curves of RWR-based and LHD-based methods. (A) ROC curve of RWR- and LHD-based methods; (B) PR curve of RWR- and LHD-based methods.

The second method was based on another powerful network algorithm, namely, the LHD algorithm [21]. This algorithm has also been used to tackle various bioinformatics problems [33–35]. Given a network, let $A$ be its adjacent matrix and $D$ be the diagonal matrix, storing the degree of each node. $L$ is the Laplacian matrix of the network, defined as $D$-$A$. According to the seed nodes, a heat distribution vector $H(t_0)$ at time $t_0$ was constructed, which is the same as $P_0$ used in the RWR algorithm. This vector was updated as follows:

$$H(t) = H(t_0) \cdot e^{-Lt}, \tag{5}$$

where $H(t)$ denotes the heat distribution vector at time $t$. When the heat distribution vectors at two consecutive time are close enough, the procedures stop. The final vector was used as output as the outcome of the LHD algorithm. Then, the heat value of each node can be obtained to indicate the association between such node and seed nodes. Similar to the RWR algorithm, the LHD algorithm was applied into the drug network for each side effect. Jackknife test was executed to generate ROC and PR curves, as illustrated in Figure 6. The corresponding AUROC and AUPR values are listed in Table 3. The AUROC and AUPR values were 0.9159 and 0.2755, respectively, and the average value was 0.5957. The AUROC was even higher than the value obtained using the path-based method ($L$ = 2), but AUPR was much lower. The average value was lower than that generated by the path-based method ($L$ = 2) obtained using local or global jackknife test. These results also prove the superiority of the path-based method.

**Table 3.** Comparison of other network-based methods.

| Method | Cross-validation | AUROC | AUPR | Average |
|---|---|---|---|---|
| Path-based method ($L$=2) | Local jackknife test | 0.9116 | 0.4533 | 0.6825 |
| | Global jackknife test | 0.8662 | 0.3353 | 0.6008 |
| RWR-based method | Jackknife test | 0.5432 | 0.0193 | 0.2813 |
| LHD-based method | Jackknife test | 0.9159 | 0.2755 | 0.5957 |

Based on the above arguments, the path-based method was superior to the RWR and LHD-based methods. In the path-based method, we constructed a heterogeneous network containing the three following types of associations: (1) drug associations; (2) side effect associations; and (3) associations between drugs and side effects. For the RWR and LHD-based methods, only drug associations were adopted. This finding indicates that these methods do not consider side effect associations and associations between drugs and side effects. Less known information was involved in these two methods, inducing their lower performance.

### *3.5. Case study*

The path-based method provided good performance for the identification of drug side effects. Thus, we analyzed some drug-side effect pairs that were assigned with high measurements (cf. Eq (2)). The results of local jackknife test were adopted to identify novel side effects for known drugs instead of candidate or completely new drugs. The top 1,000 pairs of drugs and side effects are summarized in Table S1.

Zinc is classified under the mineral and supplemental class of drugs. It plays a central role in physiological function, including gene expression, protein synthesis, and immune function, and it is

essential for the sense of smell and taste [36–38]. Zinc supplementation on infectious diseases is beneficial in clinical and experimental settings, such as diarrhea [39], acute lower respiratory tract infections [40], the common cold [41], respiratory syncytial virus infections [42], cytomegalovirus infections [43] and herpes labialis [44]. In our prediction results, for zinc, the side effects "nausea" and "vomiting" ranked highest, with values reaching 64.04 and 58.67, respectively, which are consistent with the common adverse events of zinc actually observed in many clinical trials. Singh et al. [45] extensively reviewed 18 randomized controlled trials, enrolling 1781 participants of all age groups, by using zinc for at least five consecutive days to treat, or for at least five months to prevent the common cold. The incidence of any adverse event was higher in the zinc group (OR 1.58, 95% CI 1.19 to 2.09, $p = 0.002$) than in the placebo group. The participants who received zinc experienced significantly worse nausea (OR: 2.15, 95% CI: 1.44–3.23, $p = 0.002$) than those in the placebo group. Lazzerini and Wanzira [46] combined 31 trial data, including 10,841 children, and compared oral zinc supplementation with placebo in children aged 1 month to 5 years with acute or persistent diarrhea data in meta-analyses. Considering the metallic taste of zinc, its supplementation increases the risk of "vomiting" in those given zinc across all age groups (children greater than six months of age: RR: 1.57, 95% CI: 1.32–1.86; 2,605 children, six trials, moderate certainty evidence; children less than six months of age: RR: 1.54, 95% CI: 1.05–2.24; 1,334 children, two trials, moderate certainty evidence) [46]. When zinc was administered to patients with taste disorders, compared with placebo, Sakai et al. [47] reported adverse events such as "nausea", "abdominal pain" and "diarrhea" in 16% (6/37) of patients. Watson et al. [48] reported "nausea" and "vomiting" in one patient after zinc intervention only. In a review by Cervantes et al., 10 studies using zinc in the treatment of acne vulgaris mainly reported gastrointestinal adverse reactions, the most common adverse reaction was "nausea", which even led to 3 patients withdrawing from two studies [49].

Glycerol is a naturally occurring chemical substance. Glycerol is commonly used to treat constipation, improve hydration and athletic performance of athletes, and treat certain skin diseases (e.g., ichthyosis and xerosis), meningitis, stroke, obesity, ear infections, and other conditions. For either oral, rectal, topical or intravenous administration, glycerol is a well-tolerated and safe agent. "Nausea" (54.44), "vomiting" (51.64), "diarrhea" (45.96), and "headache" (40.83) are the most common adverse reactions that may be caused by glycerol, as confirmed in clinical applications [50]. In addition to the above-mentioned more likely adverse effects, "hematuria" (18.03) and "renal failure" (15.44) are the moderately possible adverse effects of glycerol that we speculate, and they have also appeared in clinical use [51–54]. Interestingly, different reports have inconsistent explanations of the correlation between hemolysis and glycerol. Theoretically, glycerol is expected to be related to hemolysis because it can enter erythrocytes, causing water to follow and swell cells [55,56]. Therefore, some scholars believe that hemolysis is an important adverse reaction caused by high concentrations of glycerol [55,57,58]. However, no hemolysis event was found in Wang et al.'s review analysis with administration of glycerol supplemented with 5% fructose in saline [52]. Hemolysis induced by glycerol can be avoided by the addition of fructose or glucose solution, which may be attributed to the weak acidifying effect of fructose or glucose solution [56]. Hemolysis did not appear in the top of our prediction results.

Dopamine is a neurotransmitter naturally found in the body. It acts by stimulating dopamine receptors, α- and β-adrenergic receptors. When administered therapeutically, dopamine belongs to a class of drugs called inotropic drugs, which are used for the treatment of low blood pressure, low cardiac output, and reduced perfusion of body organs caused by shock, trauma, and sepsis. Generally,

low doses of dopamine (1–2 μg/kg/min) can activate dopamine receptors and cause vasodilation. The common side effects (1–10%) of dopamine that we predicted are consistent with clinical observations, including "gastrointestinal disorders" (20.50) in digestive system with symptoms of nausea and vomiting [59]; "hypotension" (23.96) in cardiovascular system (https://www.drugs.com/sfx/intropin-side-effects.html#refs). With the increase in dose (2–10 μg/kg/min), β1-adrenoceptors were activated, which may lead to "arrhythmias", including supraventricular tachycardia, atrial fibrillation, tachyarrhythmia, and even ventricular tachycardia [60]. This uncommon adverse effect (0.1 to 1%) has been reported in many studies. A further increase in the dose (10 μg/kg/min) stimulates α-adrenergic receptors resulting in a potent vasoconstrictor effect. This condition may cause "skin disorders" (15.36), such as gangrene, which is a very rare adverse effect (less than 0.01%). Therefore, common, uncommon, and rare adverse reactions related to dopamine are reflected in the top prediction results. Unfortunately, the predictive value of various adverse reactions has no corresponding score distinction, which can be further improved in subsequent studies. In addition, dopamine does not easily penetrate the blood–brain barrier. In theory, dopamine has fewer central adverse reactions. However, we also predicted the possible adverse reactions in the central nervous system by dopamine, such as "hallucination" (13.28). Although these conditions have not been observed in the clinical application of dopamine, they have appeared in the clinical application of dopamine receptor agonists [61].

Therefore, the path-based method can discover latent side effects of some drugs.

### *3.6. Limitations of the path-based method*

The path-based method also has some limitations. As a network method, its utility is highly related to the accuracy of the network. For a novel drug, its associations to other drugs were not easy to evaluate because the limited information about this drug was known. In this case, its associations to side effects cannot be fully measured, influencing the performance for assigning side effects to novel drugs. We also measured the associations between side effects based on drugs annotated by them. This scheme is not perfect. The path-based method can be improved if more accurate associations between side effects can be evaluated. In the future, this study will be continued to improve the path-based method in these aspects.

## 4. Conclusions

This study developed a path-based method for identification of drug side effects. It was executed on a heterogeneous network, which can fully indicate the associations of drug or side effects. The good performance of such method was proved using two types of jackknife test and the comparisons of other network-based methods. We also proposed some latent side effects for some drugs according to the results of this method. The path-based method has a potential use for studying drug side effects.

## Conflict of interest

The authors declare no conflict of interest.

## References


1. S. Shabani-Mashcool, S. A. Marashi, S. Gharaghani, NDDSA: A network- and domain-based method for predicting drug-side effect associations, *Inform. Process. Manag.*, **57** (2020), 102357. https://doi.org/10.1016/j.ipm.2020.102357
2. Y. J. Ding, J. J. Tang, F. Guo, Identification of drug-side effect association via multiple information integration with centered kernel alignment, *Neurocomputing*, **325** (2019), 211–224. https://doi.org/10.1016/j.neucom.2018.10.028
3. A. Lakizadeh, S. M. H. Mir-Ashrafi, Drug repurposing improvement using a novel data integration framework based on the drug side effect, *Inform. Med. Unlocked*, **23** (2021), 100523. https://doi.org/10.1016/j.imu.2021.100523
4. E. Pauwels, V. Stoven, Y. Yamanishi, Predicting drug side-effect profiles: a chemical fragment-based approach, *BMC Bioinformatics*, **12** (2011), 169. https://doi.org/10.1186/1471-2105-12-169
5. S. Jamal, S. Goyal, A. Shanker, A. Grover, Predicting neurological adverse drug reactions based on biological, chemical and phenotypic properties of drugs using machine learning models, *Sci. Rep.*, **7** (2017), 872. https://doi.org/10.1038/s41598-017-00908-z
6. Y. Zheng, H. Peng, S. Ghosh, C. Lan, J. Li, Inverse similarity and reliable negative samples for drug side-effect prediction, *BMC Bioinformatics*, **19** (2019), 554. https://doi.org/10.1186/s12859-018-2563-x
7. M. Liu, Y. Wu, Y. Chen, J. Sun, Z. Zhao, X. W. Chen, et al., Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs, *J. Am. Med. Inform. Assoc.*, **19** (2012), e28–35. https://doi.org/10.1136/amiajnl-2011-000699
8. S. Dey, H. Luo, A. Fokoue, J. Hu, P. Zhang, Predicting adverse drug reactions through interpretable deep learning framework, *BMC Bioinformatics*, **19** (2018), 476. https://doi.org/10.1186/s12859-018-2544-0
9. L. Chen, T. Huang, J. Zhang, M. Y. Zheng, K. Y. Feng, Y. D. Cai, et al., Predicting drugs side effects based on chemical-chemical interactions and protein-chemical interactions, *BioMed Res. Int.*, **2013** (2013), 485034. https://doi.org/10.1155/2013/485034
10. W. Zhang, F. Liu, L. Luo, J. Zhang, Predicting drug side effects by multi-label learning and ensemble learning, *BMC Bioinformatics*, **16** (2015), 365. https://doi.org/10.1186/s12859-015-0774-y
11. N. Atias, R. Sharan, An algorithmic framework for predicting side effects of drugs, *J. Comput. Biol.*, **18** (2011), 207–218. https://doi.org/10.1089/cmb.2010.0255
12. E. Muñoz, V. Novácek, P. Y. Vandenbussche, Facilitating prediction of adverse drug reactions by using knowledge graphs and multi-label learning models, *Brief. Bioinform.*, **20** (2017), 190–202. https://doi.org/10.1093/bib/bbx099
13. W. Zhang, Y. Chen, S. Tu, F. Liu, Q. Qu, Drug side effect prediction through linear neighborhoods and multiple data source integration, in *IEEE International Conference on Bioinformatics and Biomedicine*, (2016), 427–434. https://doi.org/10.1109/BIBM.2016.7822555
14. E. Munoz, V. Novacek, P. Y. Vandenbussche, Using drug similarities for discovery of possible adverse reactions, *AMIA Annu. Symp. Proc.*, **2016** (2016), 924–933.
15. X. Zhao, L. Chen, J. Lu, A similarity-based method for prediction of drug side effects with heterogeneous information, *Math. Biosci.*, **306** (2018), 136–144. https://doi.org/10.1016/j.mbs.2018.09.010

16. H. Liang, L. Chen, X. Zhao, X. Zhang, Prediction of drug side effects with a refined negative sample selection strategy, *Comput. Math. Method. M.*, **2020** (2020), 1573543. https://doi.org/10.1155/2020/1573543
17. X. Zhao, L. Chen, Z. H. Guo, T. Liu, Predicting drug side effects with compact integration of heterogeneous networks, *Curr. Bioinform.*, **14** (2019), 709–720. https://doi.org/10.2174/1574893614666190220114644
18. X. Guo, W. Zhou, Y. Yu, Y. Ding, J. Tang, F. Guo, A novel triple matrix factorization method for detecting drug-side effect association based on kernel target alignment, *BioMed Res. Int.*, **2020** (2020), 4675395. https://doi.org/10.1155/2020/4675395
19. Y. Ding, J. Tang, F. Guo, Identification of drug-side effect association via semi-supervised model and multiple kernel learning, *IEEE J. Biomed. Health*, **23** (2019), 2619–2632. https://doi.org/10.1109/JBHI.2018.2883834
20. H. Tong, C. Faloutsos, J. Pan, Fast random walk with restart and its applications, in *Sixth International Conference on Data Mining*, (2006), 613–622. https://doi.org/10.1109/ICDM.2006.70
21. D. E. Carlin, B. Demchak, D. Pratt, E. Sage, T. Ideker, Network propagation in the cytoscape cyberinfrastructure, *PLoS Comput. Biol.*, **13** (2017), e1005598. https://doi.org/10.1371/journal.pcbi.1005598
22. M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, P. Bork, A side effect resource to capture phenotypic effects of drugs, *Mol. Syst. Biol.*, **6** (2010), 343. https://doi.org/10.1038/msb.2009.98
23. M. Kuhn, D. Szklarczyk, S. Pletscher-Frankild, T. H. Blicher, C. von Mering, L. J. Jensen, et al., STITCH 4: integration of protein–chemical interactions with user data, *Nucleic Acids Res.*, **42** (2014), D401–D407. https://doi.org/10.1093/nar/gkt1207
24. D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.*, **28** (1988), 31–36. https://doi.org/10.1021/ci00057a005
25. X. Xiao, W. Zhu, B. Liao, J. Xu, C. Gu, B. Ji, et al., BPLLDA: predicting lncRNA-Disease associations based on simple paths with limited lengths in a heterogeneous network, *Front. Genet.*, **9** (2018), 411. https://doi.org/10.3389/fgene.2018.00411
26. W. Ba-Alawi, O. Soufan, M. Essack, P. Kalnis, V. B. Bajic, DASPfind: new efficient method to predict drug-target interactions, *J. Cheminformatics*, **8** (2016), 15. https://doi.org/10.1186/s13321-016-0128-4
27. Z. H. You, Z. A. Huang, Z. Zhu, G. Y. Yan, Z. W. Li, Z. Wen, et al., PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction, *PLoS Comput. Biol.*, **13** (2017), e1005455. https://doi.org/10.1371/journal.pcbi.1005455
28. J. Gao, B. Hu, L. Chen, A path-based method for identification of protein phenotypic annotations, *Curr. Bioinform.*, **16** (2021), 1214–1222. https://doi.org/10.2174/1574893616666210531100035
29. S. Kohler, S. Bauer, D. Horn, P. N. Robinson, Walking the interactome for prioritization of candidate disease genes, *Am. J. Hum. Genet.*, **82** (2008), 949–958. https://doi.org/10.1016/j.ajhg.2008.02.013
30. Y.J. Li, J. C. Patra, Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network, *Bioinformatics*, **26** (2010), 1219–1224. https://doi.org/10.1093/bioinformatics/btq108

31. X. Chen, M. X. Liu, G. Y. Yan, Drug-target interaction prediction by random walk on the heterogeneous network, *Mol. BioSyst.*, **8** (2012), 1970–1978. https://doi.org/10.1039/C2MB00002D
32. L. Chen, T. Liu, X. Zhao, Inferring anatomical therapeutic chemical (ATC) class of drugs using shortest path and random walk with restart algorithms, *BBA Mol. Basis Dis.*, **1864** (2017), 2228–2240. https://doi.org/10.1016/j.bbadis.2017.12.019
33. L. Chen, Y. H. Zhang, Z. Zhang, T. Huang, Y. D. Cai, Inferring novel tumor suppressor genes with a protein-protein interaction network and network diffusion algorithms, *Mol. Ther. Methods Clin. Dev.*, **10** (2018), 57–67. https://doi.org/10.1016/j.omtm.2018.06.007
34. S. Lu, K. Zhao, X. Wang, H. Liu, X. Ainiwaer, Y. Xu, et al., Use of laplacian heat diffusion algorithm to infer novel genes with functions related to uveitis, *Front. Genet.*, **9** (2018), 425. https://doi.org/10.3389/fgene.2018.00425
35. H. Y. Liang, B. Hu, L. Chen, S. Q. Wang, Aorigele, Recognizing novel chemicals/drugs for anatomical therapeutic chemical classes with a heat diffusion algorithm, *BBA Mol. Basis. Dis.*, **1866** (2020), 165910. https://doi.org/10.1016/j.bbadis.2020.165910
36. M. Imanishi, Y. Hori, M. Nagaoka, Y. Sugiura, Design of novel zinc finger proteins: towards artificial control of specific gene expression, *Eur. J. Pharm. Sci.*, **13** (2001), 91–97. https://doi.org/10.1016/S0928-0987(00)00212-8
37. M. Alirezaei, E. Mordelet, N. Rouach, A. C. Nairn, J. Glowinski, J. Premont, Zinc-induced inhibition of protein synthesis and reduction of connexin-43 expression and intercellular communication in mouse cortical astrocytes, *Eur. J. Neurosci.*, **16** (2002), 1037–1044. https://doi.org/10.1046/j.1460-9568.2002.02180.x
38. K. H. Ibs, L. Rink, Zinc-altered immune function, *J. Nutr.*, **133** (2003), 1452s–1456s. https://doi.org/10.1093/jn/133.5.1452S
39. Z. A. Bhutta, R. E. Black, K. H. Brown, J. M. Gardner, S. Gore, A. Hidayat, et al., Prevention of diarrhea and pneumonia by zinc supplementation in children in developing countries: Pooled analysis of randomized controlled trials, *J. Pediatr.*, **135** (1999), 689–697. https://doi.org/10.1016/S0022-3476(99)70086-7
40. D. E. Roth, S. A. Richard, R. E. Black, Zinc supplementation for the prevention of acute lower respiratory infection in children in developing countries: meta-analysis and meta-regression of randomized trials, *Int. J. Epidemiol.*, **39** (2010), 795–808. https://doi.org/10.1093/ije/dyp391
41. D. Hulisz, Efficacy of zinc against common cold viruses: an overview, *J. Am. Pharm. Assoc.*, **44** (2004), 594–603. https://doi.org/10.1331/1544-3191.44.5.594.Hulisz
42. R. O. Suara, J. E. Crowe, Effect of zinc salts on respiratory syncytial virus replication, *Antimicrob. Agents Ch.*, **48** (2004), 783–790. https://doi.org/10.1128/AAC.48.3.783-790.2004
43. D. Li, L. Z. Wen, H. Yuan, Observation on clinical efficacy of combined therapy of zinc supplement and jinye baidu granule in treating human cytomegalovirus infection, *Zhongguo Zhong xi yi jie he za zhi*, **25** (2005), 449–451.
44. F. Femiano, F. Gombos, C. Scully, Recurrent herpes labialis: a pilot study of the efficacy of zinc therapy, *J. Oral Pathol. Med.*, **34** (2005), 423–425. https://doi.org/10.1111/j.1600-0714.2005.00327.x
45. M. Singh, R. R. Das, Zinc for the common cold, *Cochrane Database Syst. Rev.*, **6** (2013), CD001364. https://doi.org/10.1002/14651858.CD001364.pub4

46. M. Lazzerini, H. Wanzira, Oral zinc for treating diarrhoea in children, *Cochrane Database Syst. Rev.*, **12** (2016), CD005436. https://doi.org/10.1002/14651858.CD005436.pub5
47. F. Sakai, S. Yoshida, S. Endo, H. Tomita, Double-blind, placebo-controlled trial of zinc picolinate for taste disorders, *Acta oto-laryngol.*, **122** (2002), 129–133. https://doi.org/10.1080/00016480260046517
48. A. R. Watson, A. Stuart, F. E. Wells, I. B. Houston, G. M. Addison, Zinc supplementation and its effect on taste acuity in children with chronic renal failure, *Hum. Nutr. Clin. Nutr.*, **37** (1983), 219–225.
49. J. Cervantes, A. E. Eber, M. Perper, V. M. Nascimento, K. Nouri, J. E. Keri, The role of zinc in the treatment of acne: A review of the literature, *Dermatol. Ther.*, **31** (2018), e12576. https://doi.org/10.1111/dth.12576
50. A. Y. Bedikian, M. Valdivieso, L. K. Heilbrun, R. H. Withers, G. P. Bodey, E. J. Freireich, Glycerol: an alternative to dexamethasone for patients receiving brain irradiation for metastatic disease, *South. Med. J.*, **73** (1980), 1210–1214.
51. M. S. Frank, M. C. Nahata, M. D. Hilty, Glycerol: a review of its pharmacology, pharmacokinetics, adverse reactions, and clinical use, *Pharmacotherapy*, **1** (1981), 147–160. https://doi.org/10.1002/j.1875-9114.1981.tb03562.x
52. J. Wang, Y. Ren, S. F. Wang, L. D. Kan, L. J. Zhou, H. M. Fang, et al., Comparative efficacy and safety of glycerol versus mannitol in patients with cerebral oedema and elevated intracranial pressure: A systematic review and meta-analysis, *J. Clin. Pharm. Ther.*, **46** (2021), 504–514. https://doi.org/10.1111/jcpt.13314
53. J. Wang, Y. Ren, L. J. Zhou, L. D. Kan, H. Fan, H. M. Fang, Glycerol Infusion Versus Mannitol for Cerebral Edema: A Systematic Review and Meta-analysis, *Clin. Ther.*, **43** (2021), 637–649. https://doi.org/10.1016/j.clinthera.2021.01.010
54. E. Righetti, M. G. Celani, T. A. Cantisani, R. Sterzi, G. Boysen, S. Ricci, Glycerol for acute stroke, *Cochrane Database Syst. Rev.*, **2** (2004), CD000096. https://doi.org/10.1002/14651858.CD000096.pub2
55. A. Frei, C. Cottier, P. Wunderlich, E. Lüdin, Glycerol and dextran combined in the therapy of acute stroke. A placebo-controlled, double-blind trial with a planned interim analysis, *Stroke*, **18** (1987), 373–379. https://doi.org/10.1161/01.STR.18.2.373
56. E. Lin, Glycerol utilization and its regulation in mammals, *Annu. Rev. Biochem.*, **46** (1977), 765–795. https://doi.org/10.1146/annurev.bi.46.070177.004001
57. Y. Yu, C. Kumana, I. Lauder, Y. Cheung, F. Chan, M. Kou, et al., Treatment of acute cortical infarct with intravenous glycerol. A double-blind, placebo-controlled randomized trial, *Stroke*, **24** (1993), 1119–1124. https://doi.org/10.1161/01.STR.24.8.1119
58. B. á Rogvi-Hansen, G. Boysen, Intravenous Glycerol Treatment of Acute Stroke – A Statistical Review, *Cerebrovasc. Dis.*, **2** (1992), 11–13. https://doi.org/10.1159/000108981
59. H. L. Philpott, S. Nandurkar, J. Lubel, P. R. Gibson, Drug-induced gastrointestinal disorders, *Frontline Gastroente.*, **5** (2014), 49–57. http://dx.doi.org/10.1136/flgastro-2013-100316
60. S. Saleem, How to induce arrhythmias with dopamine, in *Arrhythmia Induction in the EP Lab*, Springer, (2019), 81–89. https://doi.org/10.1007/978-3-319-92729-9_9
61. R. Ceravolo, C. Rossi, E. Del Prete, U. Bonuccelli, A review of adverse events linked to dopamine agonists in the treatment of Parkinson's disease, *Expert Opin. Drug Saf.*, **15** (2016), 181–198. https://doi.org/10.1517/14740338.2016.1130128

## Supplementary

**Table S1.** Top 1000 non-determined drug-side effect pairs yielded by the path-based method under local jackknife test.