BMC Bioinformatics

RESEARCH   **Open Access**

Drug repositioning based on individual bi-random walks on a heterogeneous network

Yuehui Wang[1], Maozu Guo[2]*, Yazhou Ren[3], Lianyin Jia[4] and Guoxian Yu[1,5]*

Abstract

Background: Traditional drug research and development is high cost, time-consuming and risky. Computationally identifying new indications for existing drugs, referred as *drug repositioning*, greatly reduces the cost and attracts ever-increasing research interests. Many network-based methods have been proposed for drug repositioning and most of them apply random walk on a heterogeneous network consisted with disease and drug nodes. However, these methods generally adopt the same walk-length for all nodes, and ignore the different contributions of different nodes.

Results: In this study, we propose a drug repositioning approach based on individual bi-random walks (DR-IBRW) on the heterogeneous network. DR-IBRW firstly quantifies the individual work-length of random walks for each node based on the network topology and knowledge that similar drugs tend to be associated with similar diseases. To account for the inner structural difference of the heterogeneous network, it performs bi-random walks with the quantified walk-lengths, and thus to identify new indications for approved drugs. Empirical study on public datasets shows that DR-IBRW achieves a much better drug repositioning performance than other related competitive methods.

Conclusions: Using individual random walk-lengths for different nodes of heterogeneous network indeed boosts the repositioning performance. DR-IBRW can be easily generalized to prioritize links between nodes of a network.

Keywords: Drug repositioning, Drug-disease heterogeneous network, Individual walk-length, Bi-random walks

Background

Traditional drug research and development depends on cell-based or target-based screening of chemical compounds to identify a small subset of 'hits'. The identification process aims to further increase their affinity, efficacy and selectivity, before moving forward to animal tests and clinical trials [1]. Drug development in general is complicated, time-consuming and expensive with high-risk [2]. In light of these difficulties in traditional drug discovery, identifying new indications for existing drugs, also known as *drug repositioning*, has attracted increasing interests from both the pharmaceutical industry and research community [3]. Drug repositioning is much more economic compared with traditional approaches, it offers a promising alternative to reduce the cost and time, since the repositioned drug has already passed the required safety tests.

However, most successfully repositioned drugs up to date have been the consequence of incidental observations of unexpected efficacy and side effects in the development or on the market [4]. For example, Sildenafil was originally tested for angina, now is indicated for erectile dysfunction and pulmonary hypertension [2]; Minoxidil was originally tested for hypertension; now is indicated for hair loss [5]. With the influx of big biochemical and phenotypic data, drug repositioning holds great potential for

*Correspondence: guomaozu@bucea.edu.cn; gxyu@swu.edu.cn
[1]College of Computer and Information Sciences, Southwest University, 400715 Beibei, Chongqing, China
[2]School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, 100044 Beijing, China
[5]Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, 430074 Wuhan, China
Full list of author information is available at the end of the article

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 2 of 13

precise medicine. It is profitable and promising to develop computational methods to predict new indications for approved drugs on large scale.

Some computational drug repositioning methods have been proposed and they can be roughly divided into two categories: focusing on the interactions between drugs and the targets; and focusing on exploiting the knowledge of diseases and drugs [6]. To name a few, Bleakley and Yamanishi [7] developed a bipartite local model (BLM) to predict target proteins of a given drug and target drugs of a give protein, and then combine these two predictions to give a final prediction for each candidate drug-target interaction. Cheng et al. [8] used a drug-target bipartite network topology similarity and a network based inference algorithm (NBI) to infer new targets for known drugs. Wang et al. [9] used known drug-target interactions as well as drug-drug and target-target similarities to construct a heterogeneous network, and then introduced a Heterogeneous Graph Based Inference (HGBI) method to iteratively update the strength between unlinked drug-target pairs based on all the paths in the network connecting them. These drug-target prediction methods can be readily adopted for drug repositioning.

Chiang et al. [10] attempted to predict novel associations between drugs and diseases based on the widely-adopted 'guilt-by-association' principle. This principle assumes that if a drug can treat one of two similar diseases, then it might treat the other also; alternatively a disease can be treated by two similar drugs. Following this principle, Gottlieb et al. [11] measured the similarity between the pertaining drug and disease of drug-disease pairs that are known to be associated based on multiple drug-drug sources and disease-disease similarity metrics, and then ranked the accumulative evidence for association using a logistic regression scheme to predict novel drug indications. Wang et al. [1] integrated omics data about diseases, drugs and drug targets to construct a heterogeneous network and then applied random walks on the network to replenish missing associations between drugs and diseases. Martinez et al. [6] integrated information on diseases, drugs and targets (proteins) to construct a heterogeneous network and then performed propagation flow on the network to prioritize candidate associations between diseases and drugs according to their interconnections in the network. Luo et al. [12] proposed MBiRW to predict drug-disease associations. MBiRW employs known drug-disease associations to improve the drug-drug and disease-disease similarity measures; and then integrates the similarity networks and drug-disease associations to build a drug-disease heterogeneous network; after that, it performs bi-random walk with restart on the network to predict novel potential drug-disease associations. Liu et al. [13] performed a drug-centric random walk and a disease-centric random walk to obtain the

association confidence between the disease nodes and drug nodes of a heterogeneous network.

Most of these aforementioned methods in essence are random walk based solutions. Although they make use of the network topology from different perspectives, they ignore the different contributions of different nodes on transferring the information on the network and almost all adopt a *fixed walk-length* for all nodes. To overcome this issue, we propose a novel drug repositioning approach (called DR-IBRW) that performs bi-random walk with restart on a heterogeneous network with quantified individual walk-length for each node. DR-IBRW uses disease symptom information [14] and drug chemical fingerprints [15] to construct a composite disease-disease similarity network, drug-drug similarity network. It then quantifies the individual walk-length for each node based on the topology of known drug-disease association network. Next, it constructs a heterogeneous network based on these three networks. After that, it performs bi-random walks with the quantified walk-lengths to account for the structural differences of these networks and contribution differences of different nodes (including diseases and drugs), and to predict new associations between drugs and diseases, and thus to accomplish the drug repositioning. We evaluate and compare the performance of DR-IBRW on several public datasets. DR-IBRW obtains much better performance than other related comparing methods [7–9, 12, 16] in identifying new indications for existing drugs, and the quantified individual walk-length indeed contributes to an improved prediction performance. We want to remark that the proposed individual bi-random walk solution is *different* from existing personalized random walk solutions [17, 18] that mainly focus on setting different restart probabilities for different nodes.

Materials and methods
Dataset
The datasets used in this work include drug-disease associations, drug fingerprints and disease symptoms. We collected 4219 diseases from MeSH [19] and 322 symptoms for each disease from the supplementary material of [14]. The drug-disease association dataset was obtained from [20], it includes 3250 known drug-disease associations involving 799 drugs and 719 diseases. We also collected 881 fingerprints for each drug from PubChem [15]. Since only 525 diseases can find their relevant symptom information from the supplementary material of [14], the final processed dataset includes 525 diseases, 718 drugs and 2177 drug-disease associations. All these data were collected on November 1st, 2017.

Similarity measures
We separately apply a four-step measurement to quantify the inner-similarity between diseases and between drugs.

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 3 of 13

The first three steps are based on the comprehensive similarity measurement used by Luo et al. [12]. In the fourth step, we use Gaussian interaction profile kernel similarity [21] to measure the similarity between drugs and diseases. Finally, we combine these similarities to form the composite similarity between diseases and between drugs. The four-step procedure of measuring the similarity between drugs is briefly introduced as follows.

Step 1: Based on the chemical fingerprints of the drug molecules, we can initially measure the similarity $\mathbf{S}_r^1 \in \mathbb{R}^{n_r \times n_r}$ between n_r drugs via the widely used Cosine similarity metric [22]. Let $\mathbf{r}_i$ and $\mathbf{r}_j$ be the vector forms of the chemical fingerprints of drug r_i and r_j, the chemical similarity $\mathbf{S}_r^1(r_i, r_j)$ between drug r_i and r_j is defined as:

$$\mathbf{S}_r^1(r_i, r_j) = \frac{\mathbf{r}_i^T \mathbf{r}_j}{\|\mathbf{r}_i\| \, \|\mathbf{r}_j\|}, \tag{1}$$

Step 2: Too small similarity provides little information for drug repositioning and can be transformed into zeros for accurate prediction [9, 12]. We partition $\mathbf{S}_r^1$ into ten subranges $((0, 0.1], (0.1, 0.2],$ etc.) and calculate the average similarity of drug pairs with shared diseases for each subrange. We also randomly shuffle $\mathbf{S}_r^1$ and repeat the partition and calculation process again. If the average of the non-shuffled subrange is smaller than that of the respective shuffled subrange, the drug similarities divided into this subrange are viewed as not informative; otherwise, they are informative. We then adopt a logistic function [23] to shrink these non-informative similarities to zero and to enlarge these informative similarities. The logistic function is defined as follows:

$$\mathbf{S}_r^2(r_i, r_j) = \frac{1}{1 + e^{c \mathbf{S}_r^1(r_i, r_j) + d}}, \tag{2}$$

where c and d are the parameters can be tuned to control the adjustment of $\mathbf{S}_r^1$. c is the upper bound of the first subrange whose average similarity is smaller than that of the respective shuffled subrange, $d = log(999)$. After that, we obtain a updated drug similarity matrix $\mathbf{S}_r^2$.

Step 3: Two drugs are more similar if they are grouped into the same cluster. To make use of this assumption, we first construct a new weighted drug sharing network with drugs as nodes and edge weight reflecting the number of common diseases by respective pair nodes. After that, we adopt a graph clustering method, ClusterONE [24], to identify potential drug clusters on the network. We then add the clustering cohesiveness of a cluster with $\mathbf{S}_r^2$ if and only if the two drugs belong to that cluster.

$$f(\mathcal{C}) = \frac{W_{in}(\mathcal{C})}{(W_{in}(\mathcal{C}) + W_{bound}(\mathcal{C}) + P(\mathcal{C}))}, \tag{3}$$

where $W_{in}(\mathcal{C})$ denotes the total weight of edges within a cluster of vertices, $W_{bound}(\mathcal{C})$ represents the total weight of edges connecting nodes of this cluster to nodes of other clusters, and $P(\mathcal{C})$ is the penalty term. Suppose that drug r_i and drug r_j locating in the same cluster $\mathcal{C}$, the comprehensive drug similarity $\mathbf{S}_r^3(r_i, r_j)$ between drug r_i and r_j is defined as $(1 + f(\mathcal{C})) * \mathbf{S}_r^2(r_i, r_j)$. In this way, we obtain an improved drug similarity matrix $\mathbf{S}_r^3$.

Step 4: Based on the assumption that similar drugs tend to show similar interaction and non-interaction profiles with the diseases, we further use Gaussian interaction profile kernel similarity to measure the similarity between drugs [21, 25, 26]. The interaction profile $IP(r_i)$ of drug r_i is defined as a binary vector encoding the presence or absence of the known associations between the drug and n_d diseases. The Gaussian interaction profile kernel similarity between two drugs (r_i and r_j) is computed as follows:

$$\mathbf{S}_r^{KR}(r_i, r_j) = exp\left(-\Upsilon_r \|IP(r_i) - IP(r_j)\|^2\right), \tag{4}$$

$$\Upsilon_r = \tilde{\Upsilon}_r / \left(\frac{1}{n_r} \sum_{i=1}^{n_r} \|IP(r_i)\|^2\right), \tag{5}$$

where Υ_r is the kernel bandwidth, $\tilde{\Upsilon}_r$ is the average number of associated diseases per drug.

To this end, we combine $\mathbf{S}_r^3$ and $\mathbf{S}_r^{KR}$ into the composite similarity matrix $\mathbf{S}_r$ between n_r drugs as follows:

$$\mathbf{S}_r = \left(\mathbf{S}_r^3 + \mathbf{S}_r^{KR}\right) / 2 \tag{6}$$

Following the above four-step, we can also compute the composite similarity $\mathbf{S}_d \in \mathbb{R}^{n_d \times n_d}$ between n_d diseases based on the symptom information of these diseases and drug-disease associations.

Quantifying individual walk-length

Network-based drug repositioning methods generally apply random walk on a network with a fixed walk-length for all nodes to explore the network topology [12, 27, 28]. They ignore the different contributions of different nodes to some extent. Given that, we introduce an individual walk-length measure and try to make better use of the topology of known drug-disease association bipartite network $\mathbf{W}_{rd} \in \mathbb{R}^{n_r \times n_d}$ of n_r drugs and n_d diseases. $\mathbf{W}_{rd}(r_i, d_j) = 1$ if the association between the drug r_i and disease d_j is known; and 0 otherwise.

The walk-length of a node generally depends on its influence in the network [29]. We extend the Jaccard index measure introduced by Lu et al. [16] to quantify the individual walk-length of nodes. Suppose $\mathcal{N}^r(r_i)$ denote the set of neighbours of drug r_i and $\mathcal{N}^d(d_j)$ denote the set of neighbours of disease d_j, if r_i and d_j share many common neighbours, they will be more probably influenced with each other. For a randomly selected feature f of either r_i or d_j, traditional Jaccard index measures the probability that both r_i and d_j have that feature as follows [30]:

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 4 of 13

$$\mathbf{JI}'(r_i, d_j) = \frac{|\mathcal{N}^r(r_i) \cap \mathcal{N}^d(d_j)|}{|\mathcal{N}^r(r_i) \cup \mathcal{N}^d(d_j)|}, \tag{7}$$

Since there is no relationship between diseases or between drugs in the drug-disease bipartite network, $\mathcal{N}^r(r_i) \cap \mathcal{N}^d(d_j)$ is an empty set. For this reason, we have to modify the definition of Jaccard index for a bipartite graph. Particularly, we define $\widehat{\mathcal{N}}^r(r_i) = \cup_{c \in \mathcal{N}^r(r_i)} \mathcal{N}^d(c)$ as the set of drugs associated with r_i's neighbours. Then the Jaccard index on the bipartite network is defined as follows:

$$\mathbf{JI}(r_i, d_j) = \frac{|\mathcal{N}^r(r_i) \cap \widehat{\mathcal{N}}^d(d_j)|}{|\mathcal{N}^r(r_i) \cup \widehat{\mathcal{N}}^d(d_j)|}, \tag{8}$$

$\mathbf{JI}(r_i, d_i)$ represents the influence between drug-disease pair (r_i, d_j). We assume that a node with *high* quantified influence has *more* probability to interact with others during the random walk process, and this node should have *larger* walk-length. Based on this assumption, we can measure the walk-length of each node as follows:

$$\mathbf{L}_r(r_i) = \sum_{j=1}^{n_d} \mathbf{JI}(r_i, d_j), \quad \mathbf{L}_d(d_j) = \sum_{i=1}^{n_r} \mathbf{JI}(r_i, d_j), \tag{9}$$

where $\mathbf{L}_r \in \mathbb{R}^{n_r}$ and $\mathbf{L}_d \in \mathbb{R}^{n_d}$ store the individual walk-lengths of n_r drugs and n_d diseases, respectively.

Individual bi-random walk

Based on the inner similarity network (defined by $\mathbf{S}^r$) of drugs, the inner similarity network (defined by $\mathbf{S}^d$) of diseases, and the drug-disease bipartite network initialized by known drug-disease associations, we can construct a heterogeneous network of drugs and diseases (see Fig. 1 for example). We adopt a bi-random walk with restart procedure [27] on the heterogeneous network. Compared with traditional random walk with restart, the bi-random walk with restart can separately propagate information in different subnetworks, instead of the global network [28]. For this reason, bi-random walk can separately account for the inner structure of disease similarity network and of drug similarity network, and also make use of associations between drugs and diseases.

A random walker can take a drug as the starting node, its associated diseases as intermediate nodes, and then traverse to other disease nodes. In this way, we can get probabilistic associations between the drug and new diseases, and thus identify potential new indications of the drug. To mimic this process, we perform random walk with restart starting from drug nodes and then traversing to disease nodes based on the quantified individual walk-length and the heterogeneous network topology as follows:

$$\mathbf{F}_r^t(r_i, d_j) = \begin{cases} \alpha \sum_{k=1}^{n_d} \tilde{\mathbf{S}}_d(d_k, d_j)\mathbf{F}_r^{t-1}(r_i, d_k) + \\ (1 - \alpha)\mathbf{W}_{rd}(r_i, d_j), \text{if } t \leq \lfloor \mathbf{L}_r(r_i) \rfloor \\ \mathbf{F}_r^{t-1}(r_i, d_j), \quad \text{otherwise} \end{cases} \tag{10}$$

where $\mathbf{F}_r^t(r_i, d_j)$ is the predicted relevance between drug r_i and disease d_j in the t-th iteration, $\mathbf{F}_r^0 = \mathbf{W}_{rd}$, $\alpha > 0$ controls the probability for a walker staying at the starting point, $\tilde{\mathbf{S}}_d = \mathbf{D}_d^{-\frac{1}{2}} * \mathbf{S}_d * \mathbf{D}_d^{-\frac{1}{2}}$ is the Laplacian normalized result of $\mathbf{S}_d$ and $\mathbf{D}_d$ is a diagonal matrix with $\mathbf{D}_d(d_j, d_j) = \Sigma_{k=1}^{n_d} \mathbf{S}_d(d_j, d_k)$. If $t > \mathbf{L}_r(r_i)$, the random walker starting
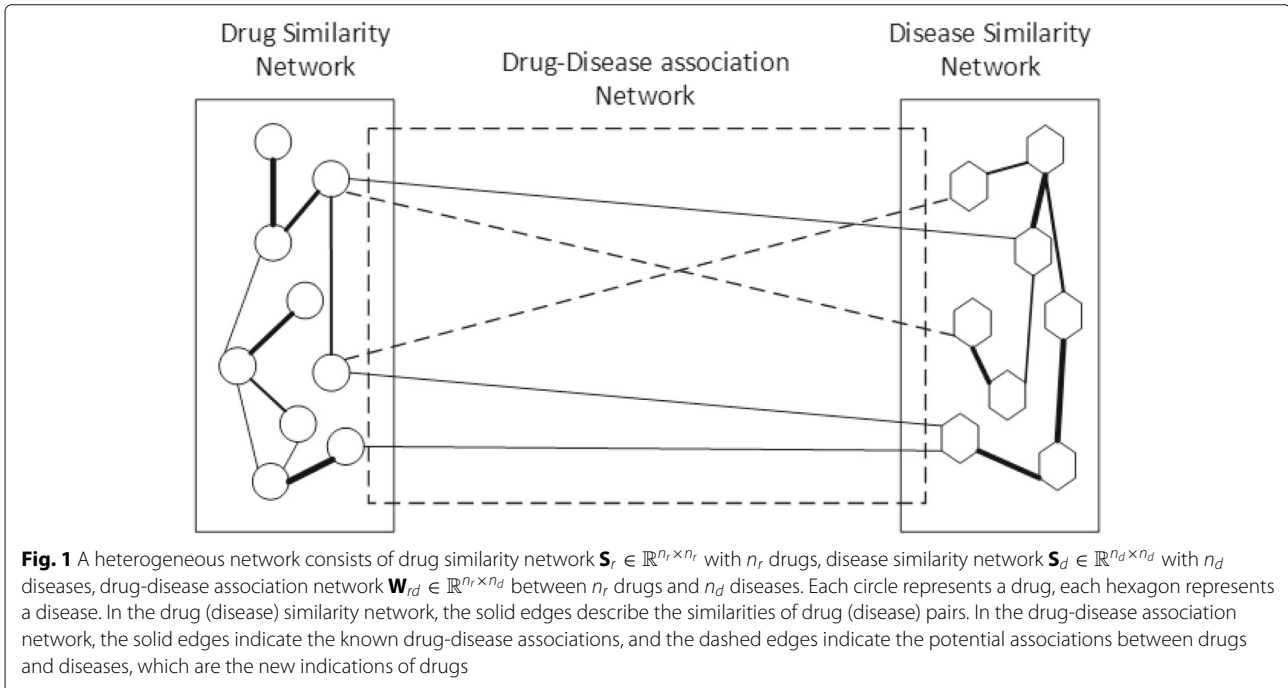

Fig. 1 A heterogeneous network consists of drug similarity network $\mathbf{S}_r \in \mathbb{R}^{n_r \times n_r}$ with n_r drugs, disease similarity network $\mathbf{S}_d \in \mathbb{R}^{n_d \times n_d}$ with n_d diseases, drug-disease association network $\mathbf{W}_{rd} \in \mathbb{R}^{n_r \times n_d}$ between n_r drugs and n_d diseases. Each circle represents a drug, each hexagon represents a disease. In the drug (disease) similarity network, the solid edges describe the similarities of drug (disease) pairs. In the drug-disease association network, the solid edges indicate the known drug-disease associations, and the dashed edges indicate the potential associations between drugs and diseases, which are the new indications of drugs

Wang et al. BMC Bioinformatics 2019, **20**(Suppl 15):547

Page 5 of 13

from r_i will not jump any more. We want to recomment that unlike traditional random walks and bi-random walks that adopt the same walk-length for all the nodes, the walk-length of a node in Eq. (10) is adaptively set based on its topology relationship with other nodes and is different from the walk-lengths of other nodes.

Similarly, a random walker can also start from a disease node and then traverse to drug nodes based on known drug-disease relationships and drug similarity network. In this way, we can obtain another probability between the disease and drug. To simulate this process, we perform random walk with restart from the disease node (d_j) as follows:

$$\mathbf{F}_d^t(r_i, d_j) = \begin{cases} \alpha \sum_{k=1}^{n_r} \tilde{\mathbf{S}}_r(r_i, r_k) \mathbf{F}_d^{t-1}(r_k, d_j) + \\ (1-\alpha) \mathbf{W}_{rd}(r_i, d_j), \text{if } t \leq \lfloor \mathbf{L}_d(d_j) \rfloor \\ \mathbf{F}_d^{t-1}(r_i, d_j), \quad \text{otherwise} \end{cases} \quad (11)$$

where $\mathbf{F}_d^t(r_i, d_j)$ is the predicted relevance between drug r_i and disease d_j in the t-th iteration, and the same normalization procedure is applied to $\mathbf{S}_r$ to construct the normalization matrix $\tilde{\mathbf{S}}_r = \mathbf{D}_r^{-\frac{1}{2}} * \mathbf{S}_r * \mathbf{D}_r^{-\frac{1}{2}}$, $\mathbf{D}_r$ is a diagonal matrix with $\mathbf{D}_r(r_i, r_i) = \Sigma_{j=1}^{n_r} \mathbf{S}_r(r_i, r_j)$.

After iteratively applying Eqs. (10-11) with individual walk-lengths, we can obtain $\mathbf{F}_r$ and $\mathbf{F}_d$, which separately reflect the association confidences between n_r drugs and n_d diseases from the perspective of the disease similarity network, and from the drug similarity network, along with the known drug-disease associations. To this end, we integrate them as follows:

$$\mathbf{F} = \frac{\mathbf{F}_r + \mathbf{F}_d}{2} \quad (12)$$

Obviously, the larger the value of $\mathbf{F}(r_i, d_j)$, the larger the probability that drug r_i associated with disease d_j is. In this way, we can finally identify new indications for existing drugs. The whole procedure of DR-IBRW is described in Algorithm 1.

Results and discussion
Performance comparison with other methods
DR-IBRW is compared with five related and recent methods (MBiRW [12], BLM [7], JI (Jaccard Index) [16], HGBI [9] and NBI [8]) on the processed dataset. MBiRW, BLM, HGBI and NBI were introduced in the Introduction, the last four methods are originally developed for predicting drug-target interactions and can be directly adopted to predict drug-disease associations. Parameters of these comparing methods are set (or optimized) as the authors suggested (or provided) in their respective papers or codes. As to DR-IBRW, α for random walk restart probability is set to 0.1. To reach a comprehensive evaluation, we use six widely used metrics, namely *AUROC, AUPR, Macro-F1, Micro-F1, Precision, Recall*. These metrics are

---

Algorithm 1: DR-IBRW

Input: Drug set $\mathcal{R}$, disease set $\mathcal{D}$, drug-disease association matrix $\mathbf{W}_{rd}$ and parameter α

Output: predicted drug-disease association matrix $\mathbf{F}$

1 Calculate drug-drug (disease-disease) similarity matrix $\mathbf{S}_r$ based on Eqs. (1-6);

2 Quantifying individual walk-length $\mathbf{L}_r$ and $\mathbf{L}_d$ for drugs and diseases using Eqs. (7-9);

3 $\mathbf{W}_{rd} = \mathbf{W}_{rd}/diag(\sum \mathbf{W}_{rd}(:, j))$, $\mathbf{F}_r^0 = \mathbf{F}_d^0 = \mathbf{W}_{rd}$;

4 **for** $t = 1$ *to* $\max \mathbf{L}_r$ **do**

5   **for** $i = 1$ *to* n_r **do**

6   Update $\mathbf{F}_r^t(r_i, :)$ using Eq. (10);

7   **end**

8 **end**

9 **for** $t = 1$ *to* $\max \mathbf{L}_d$ **do**

10   **for** $j = 1$ *to* n_d **do**

11   Update $\mathbf{F}_d^t(:, d_j)$ using Eq. (11);

12   **end**

13 **end**

14 $\mathbf{F} = (\mathbf{F}_r + \mathbf{F}_d)/2$;

15 **return F**

---

also used by those comparing methods [7–9, 12, 16]. The formal definitions of these metrics are omitted here, but interested readers than can find the formal definitions of these metrics in these references and references therein. All these methods follow ten fold cross-validation experimental protocol, and then report the average results and standard deviation in Table 1. In addition, we also plot the receiver operating characteristic (ROC) curve and precision recall (PR) curve, and the value of area under perspective curve in Fig. 2.

We can easily find that DR-IBRW achieves better performance than these comparing methods. Although both DR-IBRW and MBiRW utilize the drug similarity network, disease similarity network and drug-disease association network to construct a heterogeneous network, and then apply bi-random walks with restart to account for the structural difference of this network, DR-IBRW still performs significantly better than MBiRW. That is because DR-IBRW takes into account the different contributions of different nodes and applies individual walk-lengths for them, whereas MBiRW equally treats all nodes and applies the same walk-length. In addition, DR-IBRW uses the Gaussian interaction profile kernel similarity to strengthen the effect of known drug-disease associations.

HGBI also applies random walks with restart on the heterogeneous network, but it does not take into account structural difference between drug similarity network and disease similarity network. BLM tries to build a separate classifier for each drug and each drug, but it is still suffered

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 6 of 13

Table 1 The results of DR-IBRW and five comparing methods

Methods	Micro-F1	Macro-F1	Precision	Recall
DR-IBRW	0.395±0.002	**0.328 ± 0.001**	0.212±0.000	**0.766 ± 0.007**
MBiRW	0.294±0.004	0.245±0.003	0.158±0.001	0.572±0.016
BLM	**0.413 ± 0.006**	0.304±0.002	**0.224 ± 0.003**	0.740±0.002
JI	0.229±0.001	0.188±0.001	0.123±0.000	0.427±0.002
HGBI	0.013±0.000	0.010±0.000	0.007±0.000	0.021±0.000
NBI	0.009±0.000	0.007±0.000	0.004±0.000	0.016±0.000

The entry in boldface represent the method perform best in this evaluation metric

from biased training data, since there are more negative samples than positive samples (known associations). In fact, a number of negative samples should be positive ones. For this reason, BLM has a high Precision and Recall but with a low AUPR value. JI takes into account the influence of a node in the bipartite network and uses common neighbours to predict drug-disease associations. NBI only utilizes known drug-disease associations to run a two-step diffusion model on the bipartite graph and it can not predict new associations for a drug without known associations. For these reasons, both JI and NBI are outperformed by DR-IBRW.
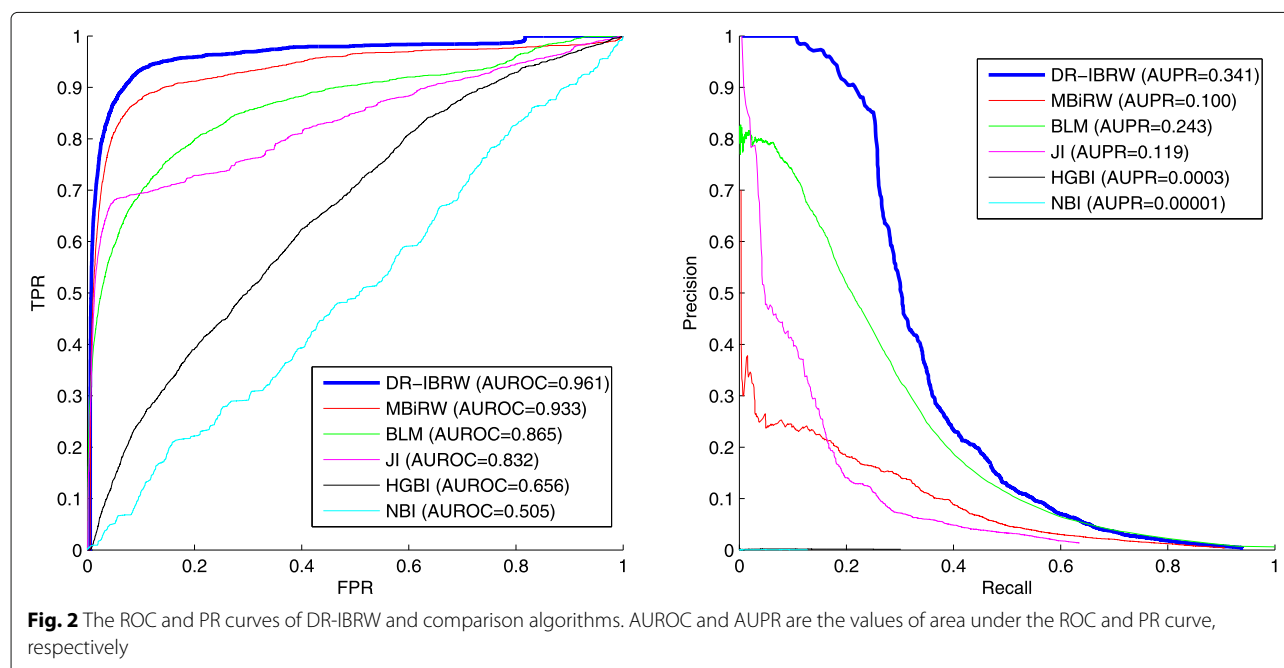
Individual walk-length analysis

To study the contribution of our proposed individual walk-lengths, we also test the performance of DR-IBRW with fixed walk-lengths for all the nodes by varying walk-length in the disease network and drug network from 0 to 10, respectively. Fig. 3 reveals the AUROC and AUPR of DR-IBRW under different combined configurations of L_r and L_d. From this figure, we can clearly see that the

AUROC stops increasing when L_r and L_d are larger than 2, and the AUROC and AUPR values with a fixed walk-length are smaller than those of DR-IBRW with individual walk-lengths. This comparison further corroborates the effectiveness and rationality of individual walk-lengths.

Drug and disease similarity analysis

As introduced in Section 5, we measure the composite inner similarity between diseases and drugs in four steps. To investigate the impact of these four steps and the contribution of Gaussian interaction kernel profile similarity, we introduce three variants (DR-IBRW123, DR-IBRW124, DR-IBRW134) of DR-IBRW. Particularly, DR-IBRW123 only uses the first three steps (as done by Luo et al. [12]), or excludes the Gaussian interaction kernel profile similarity, to measure the inner similarity between diseases and between drugs. Similarly, DR-IBRW134 excludes the second step without shrinking low similarity and enlarging high similarity. DR-IBRW124 follows the same naming rule. The AUROC and AUPR values of DR-IBRW and its variants by ten fold cross-validations are shown in

Fig. 2 The ROC and PR curves of DR-IBRW and comparison algorithms. AUROC and AUPR are the values of area under the ROC and PR curve, respectively

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 7 of 13

Fig. 3 The AUROC and AUPR values of DR-IBRW with different fixed walk-lengths. The blue star is the value of DR-IBRW with individual walk-lengths
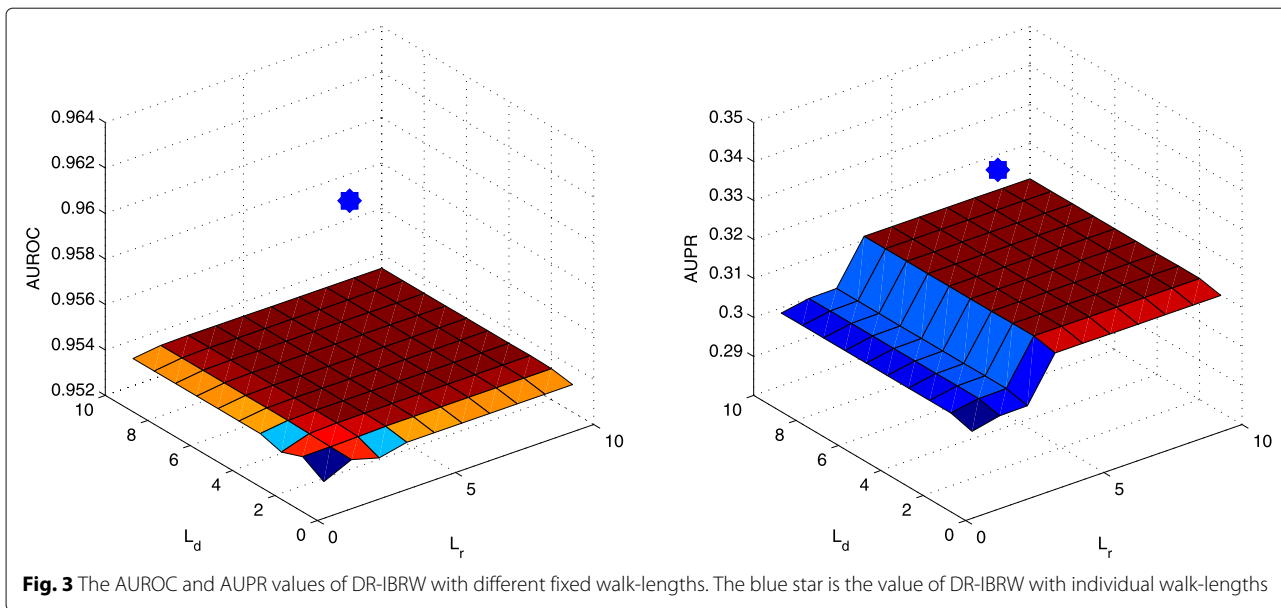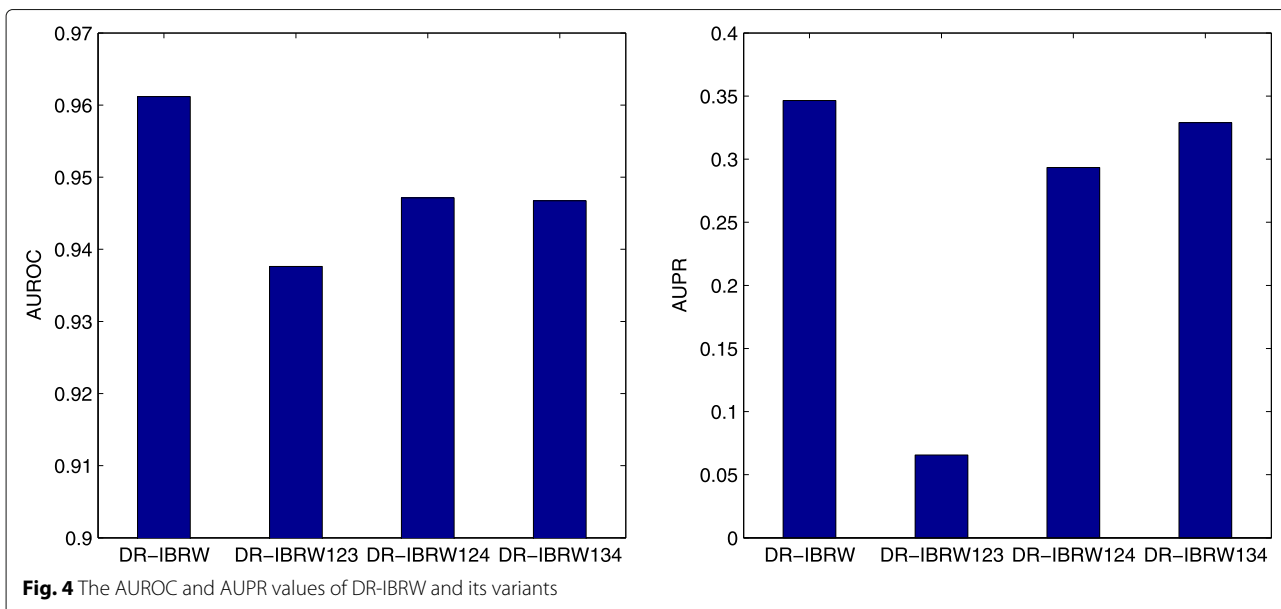
Fig. 4. Obviously, the AUROC and AUPR values of DR-IBRW123 are lower than those of other methods, which show the contribution of Gaussian interaction profile kernel similarity for drug repositioning. Another interesting observation is that DR-IBRW134 has a higher AUPR value than other variants and DR-IBRW. The cause is that AUPR and AUROC measure the performance from different perspectives and under varying thresholds. The second step may wrongly shrink low similarity and enlarge high similarity, and thus compromise the performance.

Experiments on another two datasets

We collected another two datasets to further study the performance of DR-IBRW. The first dataset (named 'Gottlieb's Dataset'), was obtained from [11]. This dataset contains 1933 known drug-disease associations involving 593 drugs registered in DrugBank and 313 diseases listed in the Online Mendelian Inheritance in Man (OMIM). The another dataset ('Luo's Dataset') is obtained from [12], it includes 663 drugs registered in DrugBank, 409 diseases listed in OMIM database and 2352 known drug-disease

Fig. 4 The AUROC and AUPR values of DR-IBRW and its variants

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 8 of 13

associations. Table 2 reports the results of 10 fold cross-validation of DR-IBRW and comparing methods on these two datasets. The experimental setups are kept the same as in previous experiments. From these tables, we can also find that DR-IBRW again obtains much better performance than these comparing methods across different evaluation metrics.

Case study

To further demonstrate that the drug-disease associations predicted by DR-IBRW can be confirmed by biological experiments, we apply DR-IBRW to prioritize potential drug-disease pairs. Here, we use all the collected drug-disease associations as training samples, and then select the top 10 drug-disease pairs with the largest association probabilities as the predicted drug-disease associations. After that, we manually check these associations by referring to the associations stored in Comparative Toxicogenomics Database(CTD) [31]. Particularly, we use the data of chemical-disease associations labeled with therapeutic downloaded from CTD. The label therapeutic represents a chemical that has a known or potential therapeutic role in a disease. For the predicted associations cannot find in the CTD, we further manually check them on PubMed and list the supportive PubMed IDs. We highlight the drug-disease associations supported by recent papers in PubMed but not included in CTD in **boldface**. The currently supported and un-supported associations are listed in Table 3.

From Table 3, 6 out of top 10 predicted associations are supported by associations in CTD, the other two drug-disease pairs are supported by recent papers in PubMed but not included in CTD. For instance, Labetalol is an effective agent in essential hypertension as documented in open studies and controlled studies [32]. For another instance, Greminger et al. confirmed the high efficacy of captopril in treatment of severe hypertension refractory to conventional drugs [33]. Meanwhile, ranolazine therapy is safe and well tolerated in a pilot study involving pulmonary arterial hypertension [34]. Although we can not find the direct evidence for the associations of flurandrenolide and scalp dermatoses, flurandrenolide topical is used to treat the itching, redness, dryness, crusting, scaling, inflammation, and discomfort of various skin conditions [35].

These predicted results confirm the capability of DR-IBRW in identifying novel drug-disease associations with high confidence. We want to remark that the 2 unsupported associations should not be viewed as incorrect associations. As more experimental evidence becomes available, they maybe further supported.

We also report the top 10 repositioned examples made by other comparing methods, and then manually check these examples by referring to the associations stored in CTD. We further check the associations that cannot find in the CTD on PubMed and list the supportive PubMed IDs. We highlight the drug-disease associations supported by recent papers in PubMed but not included in CTD in **boldface**. Tables 4, 5, 6, 7 and 8 list the currently supported and un-supported associations for MBiRW, BLM, JI, HGBI and NBI, respectively.

From Table 4, 5 out of top 10 predicted associations are supported by associations in CTD, the other two drug-disease pairs are supported by recent papers in PubMed but not include in CTD. From Table 5, we can clearly see that 1 out of top 10 predicted associations is supported

Table 2 The performance results of DR-IBRW and comparing methods on Gottlieb's dataset [11] and Luo's dataset [12]

	AUROC	AUPR	Micro-F1	Macro-F1	Precision	Recall
	Gottlieb's dataset					
DR-IBRW	**0.955 ± 0.000**	**0.499 ± 0.174**	**0.613 ± 0.006**	**0.513 ± 0.005**	**0.332 ± 0.002**	0.880±0.000
MBiRW	0.933±0.000	0.213±0.028	0.294±0.004	0.244±0.003	0.256±0.001	**0.906 ± 0.000**
BLM	0.865±0.000	0.298±0.003	0.583±0.001	0.479±0.001	0.315±0.000	0.891±0.000
JI	0.845±0.001	0.247±0.043	0.385±0.003	0.462±0.004	0.250±0.001	0.894±0.181
HGBI	0.811±0.000	0.016±0.000	0.187±0.001	0.157±0.001	0.101±0.000	0.367±0.007
NBI	0.503±0.000	0.000±0.000	0.022±0.000	0.018±0.000	0.012±0.000	0.039±0.001
	Luo's dataset					
DR-IBRW	**0.964 ± 0.000**	**0.529 ± 0.167**	**0.537 ± 0.006**	0.452±0.004	**0.294 ± 0.002**	**0.895 ± 0.002**
MBiRW	0.945±0.000	0.285±0.042	0.431±0.004	0.363±0.003	0.236±0.001	0.835±0.013
BLM	0.892±0.000	0.424±0.017	0.527±0.003	**0.463 ± 0.004**	0.278±0.001	0.843±0.000
JI	0.865±0.000	0.287±0.041	0.537±0.004	0.447±0.003	0.294±0.001	0.783±0.000
HGBI	0.848±0.000	0.037±0.001	0.170±0.001	0.141±0.001	0.093±0.000	0.318±0.005
NBI	0.479±0.000	0.000±0.000	0.020±0.000	0.016±0.000	0.011±0.000	0.032±0.000

The entry in boldface represent the method perform best in this evaluation metric

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 9 of 13

Table 3 DR-IBRW predicted drug-disease associations (top 10 in ranking list), and the corresponding evidence

Drug	Disease	Evidence (PMID)	Rank
Labetalol	Hypertension	6124264; 25692529	1
Irbesartan	Heart Failure	19001508	2
Enalapril	Hypertension	2994986	3
Flurandrenolide	Scalp Dermatoses	without evidence	4
Hydralazine	Hypertension	20687078; 22071816	5
Fenoldopam	Hypertension	8105829	6
Captopril	Hypertension	**6754186; 3520132; 1747216; 23161035**	7
Erythrityl Tetranitrate	Hypertension	without evidence	8
Nitroprusside	Hypertension	21272230; 9796241	9
Ranolazine	Hypertension	**24464752; 26401256**	10

The entries in boldface represent the drug-disease associations supported by recent papers in PubMed but not included in CTD

Table 4 MBiRW predicted drug-disease associations (top 10 in ranking list), and the corresponding evidence.

Drug	Disease	Evidence (PMID)	Rank
Echothiophate	Esotropia	**13907355; 7166393**	1
Cysteamine	Cystinosis	22532830;23651769	2
Ethanol	Complex Regional Pain Syndromes	without evidence	3
Ethanol	Warts	**9557098**	4
Foscarnet	Herpes Genitalis	without evidence	5
Ribavirin	Respiratory Syncytial Virus Infections	11781627	6
Foscarnet	Cytomegalovirus Infecti	11362300;11050094;10795660	7
Nitisinone	Tyrosinemias	11488774	8
Hydroxocobalamin	Alcoholic Neuropathy	without evidence	9
Methimazole	Goiter	14723259	10

The entries in boldface represent the drug-disease associations supported by recent papers in PubMed but not included in CTD

Table 5 BLM predicted drug-disease associations (top 10 in ranking list), and the corresponding evidence

Drug	Disease	Evidence (PMID)	Rank
Cefixime	Streptococcal Infections	2041146	1
Cefdinir	Urinary Tract Infections	**20573040**	2
Ceftibuten	Soft Tissue Infections	without evidence	3
Ceftibuten	Klebsiella Infections	**25813819**	4
Ceftibuten	Urinary Tract Infections	**11605809;2391749**	5
Cefdinir	Soft Tissue Infections	**15313534;16765555**	6
Cefditoren	Urinary Tract Infections	**20542206;8455334**	7
Cefditoren	Escherichia coli Infections	without evidence	8
Cefprozil	Urinary Tract Infections	**26391612**	9
Aztreonam	Soft Tissue Infections	without evidence	10

The entries in boldface represent the drug-disease associations supported by recent papers in PubMed but not included in CTD

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 10 of 13

Table 6 JI predicted drug-disease associations (top 10 in ranking list), and the corresponding evidence

Drug	Disease	Evidence (PMID)	Rank
Alprostadil	Tetralogy of Fallot	3543871	1
Alprostadil	Tricuspid Atresia	without evidence	2
Alprostadil	Hypoplastic Left Heart Syndrome	without evidence	3
Cefepime	Escherichia coli Infections	**26815433**	4
Cefepime	Urinary Tract Infections	**26243291;1804010**	5
Atorvastatin	Hypercholesterolemia	24593216;20946910;20135644	6
Cefotaxime	Escherichia coli Infections	without evidence	7
Clofibrate	Hypercholesterolemi	1175893;7080553;7157849	8
Fenofibrate	Hypercholesterolemi	2492189;24593216;2045526	9
Levofloxacin	Escherichia coli Infections	without evidence	10

The entries in boldface represent the drug-disease associations supported by recent papers in PubMed but not included in CTD

Table 7 HGBI predicted drug-disease associations (top 10 in ranking list), and the corresponding evidence

Drug	Disease	Evidence (PMID)	Rank
Lithium	Conduct Disorder	7491395;7691178;7751258	1
Lithium	Depressive Disorder	27752079;2723135;21252007	2
Ertapenem	Pyelonephritis	**22563210**	3
Moxifloxacin	Pyelonephritis	without evidence	4
Gatifloxacin	Pyelonephritis	**11911553;15037328**	5
Methotrexate	Psoriasis	20178709;19626273;19323665	6
Levofloxacin	Pyelonephritis	without evidence	7
Cefamandole	Staphylococcal Infections	9419181	8
Cefprozil	Staphylococcal Infections	without evidence	9
Vinblastine	Kidney Neoplasms	8602639;11194540	10

The entries in boldface represent the drug-disease associations supported by recent papers in PubMed but not included in CTD

Table 8 NBI predicted drug-disease associations (top 10 in ranking list), and the corresponding evidence

Drug	Disease	Evidence (PMID)	Rank
Flurandrenolide	Facial Dermatoses	without evidence	1
Levofloxacin	Urinary Tract Infections	**25931244**	2
Cefoperazone	Escherichia coli Infections	without evidence	3
Flurandrenolide	Scalp Dermatoses	without evidence	4
Ceftazidime	Proteus Infections	without evidence	5
Ceftizoxime	Escherichia coli Infections	**24755996**	6
Moxifloxacin	Streptococcal Infections	19188393;18818055;17562794	7
Cefpodoxime	Escherichia coli Infections	**23537823**	8
Ofloxacin	Streptococcal Infections	19856068	9
Ampicillin	Streptococcal Infections	2306432	10

The entries in boldface represent the drug-disease associations supported by recent papers in PubMed but not included in CTD

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 11 of 13

by CTD and the other six associations are supported by recent papers in PubMed. From Table 6, JI totally finds 6 drug-disease pairs with evidence among the top 10 predicted associations. From Table 7, 5 out of top 10 predicted associations are supported by associations in CTD, the other two drug-disease pairs are supported by recent papers in PubMed but not include in CTD. From Table 8, NBI can find 6 associations with evidence. In summary, DR-IBRW can make more confident drug-disease repositioning than these comparing methods.

Quantified individual walk length is reasonable

The drug-disease association prediction task is frequently modeled as a link prediction problem in a heterogeneous graph [36–38]. The link prediction relies on calculating the similarity between nodes. The number of paths between nodes and walk lengths are regarded as effective similarity metrics in the social network and biological network [36, 39, 40]. The similarities between drugs and diseases can be measured based on the number of walks that connect drug nodes and disease nodes in the network. Integrating the number of walks and their lengths can more comprehensively quantify the potential association probability of the drug-disease pair. In addition, the contribution of different nodes in the heterogeneous network is different. In other words, the information carried by each node in the heterogeneous work is imbalanced. Therefore, it is an issue to adopt a fixed walk-length for all nodes in link prediction.
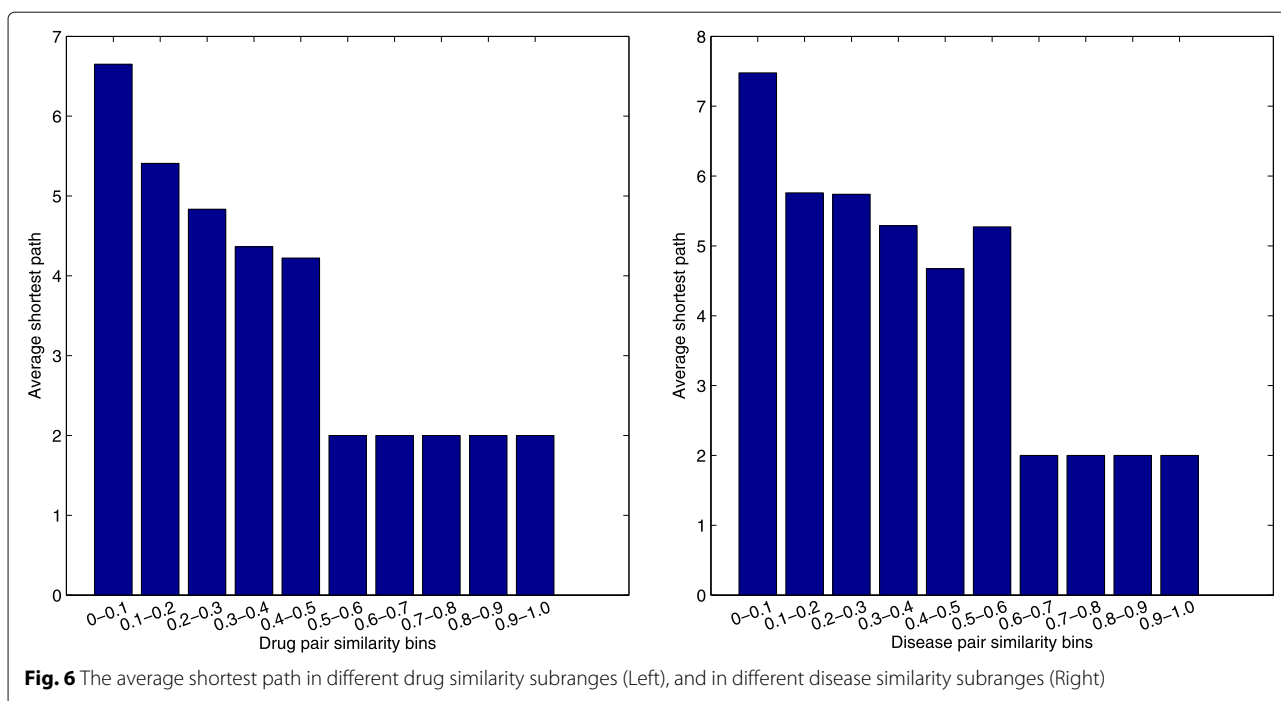
In order to answer why the choice of quantified individual walk length is reasonable, we calculate the shortest path for each drug and disease node, and measure the difference between shortest path and quantified individual walk length. We use the matrix $\mathbf{SP}(r_i, d_j)$ to represent the shortest path from the $i-th$ drug to $j-th$ disease, $\mathbf{SP} \in$

$\mathbb{R}^{(n_r+n_d)\times(n_r+n_d)}$. To calculate $\mathbf{SP}$, we firstly construct an adjacency matrix $\mathbf{W}$:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{rr} & \mathbf{W}_{rd} \\ \mathbf{W}_{dr} & \mathbf{W}_{dd} \end{bmatrix}$$

where $\mathbf{W}_{rr} \in \mathbb{R}^{n_r \times n_r}$ contains the shortest path between each two drug nodes, $\mathbf{W}_{dd} \in \mathbb{R}^{n_d \times n_d}$ contains the shortest path between each two disease nodes. $\mathbf{W}_{rd}$ is the drug-disease association matrix and $\mathbf{W}_{dr}$ is the transpose of $\mathbf{W}_{rd}$. Then, we adopt the Dijkstra algorithm to compute the shortest path between two nodes in matrix $\mathbf{W}$. $\mathbf{P}_r = (rp_1, rp_2, \ldots, rp_i, \ldots, rp_{n_r})$ where rp_i represents the longest path in the shortest path between i-th drug and all the diseases. $\mathbf{P}_d = (dp_1, dp_2, \ldots, dp_j, \ldots, dp_{n_d})$ where dp_j represents the longest path in the shortest path between j-th disease and all the drugs. In other words, rp_i is the maximum shortest path for drug i, which can include nearly all the path information with diseases. dp_j is the maximum shortest path for disease j and it can approximately represent the path between disease j and all the drugs. $\mathbf{L}_r$ and $\mathbf{L}_d$ store the quantified individual walk-lengths of n_r drugs and n_d diseases. After that, we calculate the margin between $\mathbf{P}_r$ and $\mathbf{L}_r$ for drugs, and that between $\mathbf{P}_d$ and $\mathbf{L}_d$ for diseases. The statistical results are shown in Fig. 5. We can find that nearly 60% nodes' differences are no larger than one. It can explain that the quantified individual walk lengths of most nodes are inline with the shortest path between the respective nodes. However, the maximum shortest path can only partially represent the path information from a drug node to a disease node. $\mathbf{L}_r$ can give more emphasize on shorter path between diseases and drugs than maximum shortest path, and it generally has a smaller value than $\mathbf{P}_r$. It is recognized that the shorter the distance between two nodes, the larger the

Fig. 5 The margin between $\mathbf{P}_r$ and $\mathbf{L}_r$ for drugs (Left), and the margin between $\mathbf{P}_d$ and $\mathbf{L}_d$ for diseases (Right)

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 12 of 13

Fig. 6 The average shortest path in different drug similarity subranges (Left), and in different disease similarity subranges (Right)
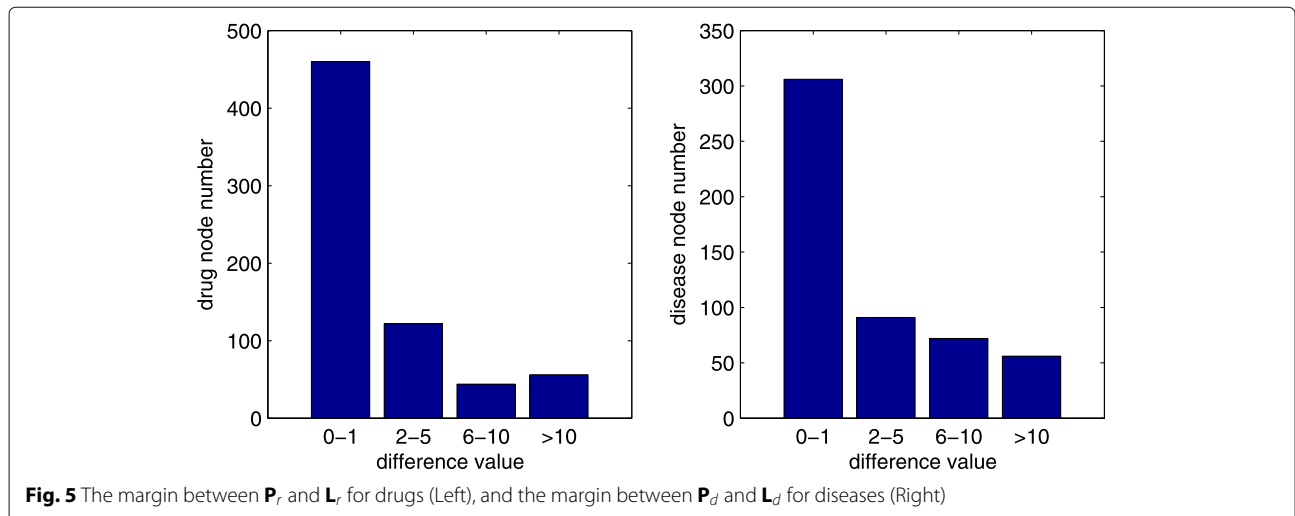
similarity between them is. For these reasons, our random walk with individual walk achieves more prominent performance than random walk fixed walk length (as shown in Fig. 3)

We also perform the correlation analysis on drug similarity matrix S_r and drug shortest path matrix W_{rr}. We firstly partition S_r into ten subranges ((0, 0.1], (0.1, 0.2], etc.) and then partition W_{rr} into ten subranges to ensure that all the drug pairs in each subrange of S_r falling into the corresponding subrange of W_{rr}. Next, we calculate the average shortest path of each subrange for W_{rr}, and compute the correlation of average shortest paths and drug similarities between W_{rr} and S_r in each subrange. Similarly, we conduct the correlation analysis on disease similarity S_d and disease shortest path matrix W_{dd} in the same way and report the results in Fig. 6. We can clearly observe that the average shortest paths between drug pairs or disease pairs decrease as the increases of their similarities. This observation also differentiates the contribution of different walk lengths based on the assumption that nodes with shorter walk lengths contribute more to the similarity between two nodes.

Conclusion

In this paper, we proposed a computational drug repositioning approach that encodes the drug chemical structure information, disease symptom information and known drug-disease interactions information into a heterogeneous network. Our approach accounts for structural difference of subnetworks of the heterogeneous network by bi-random walk, and for the contribution differences of different nodes via specifying quantified individual walk-lengths to them. Experimental study demonstrates that our approach performs better than other related competitive methods and the individual walk lengths contribute to an improved performance. We want to remark that our proposed approach can be easily generalized to predict links between nodes of a heterogeneous network.

Wang *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):547

Page 13 of 13

Author details
[1]College of Computer and Information Sciences, Southwest University, 400715 Beibei, Chongqing, China. [2]School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, 100044 Beijing, China. [3]School of Computer Science and Engineering, University of Electronic Science and Technology of China, 611731 Chengdu, China. [4]College of Information Engineering and Automation, Kunming University of Science and Technology, 650504 Kunming, China. [5]Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, 430074 Wuhan, China.

References
1. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. Bioinformatics. 2014;30(20):2923–30.
2. Sardana D., Zhu C., Zhang M., Gudivada R. C., Yang L., Jegga A. G. Drug repositioning for orphan diseases. Brief Bioinform. 2011;12(4):346–56.
3. Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational drug repositioning: From data to therapeutics. Clin Pharmacol Ther. 2013;93(4):335–41.
4. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3(4):673–683.
5. Graul AI, Sorbera L, Pina P, Tell M, Cruces E, Rosa E. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3(4):673–83.
6. Martínez V, Navarro C, Cano C, Fajardo W, Blanco A. Drugnet: Network-based drug–disease prioritization by integrating heterogeneous data. Artif Intell Med. 2015;63(1):41–9.
7. Bleakley K., Yamanishi Y. Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics. 2009;25(18):2397–403.
8. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, Zhou W, Huang J, Tang Y. Prediction of drug-target interactions and drug repositioning via network-based inference. PLOS Computat Biol. 2012;8(5):1–12.
9. Wang W, Yang S, Li J. Drug target predictions based on heterogeneous graph inference. In: Pacific Symposium on Biocomputing; 2013. p. 53–64. https://doi.org/10.1142/9789814447973_0006.
10. Chiang AP, Butte AJ. Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. Clin Pharmacol Ther. 2009;86(5):507–10.
11. Gottlieb A, Stein GY, Ruppin E, Sharan R. Predict: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol. 2011;7(1):.
12. Luo H, Wang J, Li M, Luo J, Peng X, Wu F-X, Pan Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. Bioinformatics. 2016;32(17):2664–71.
13. Liu H, Song Y, Guan J, Luo L, Zhuang Z. Inferring new indications for approved drugs via random walk on drug-disease heterogenous networks. BMC Bioinformatics. 2016;17(17):539.
14. Zhou X., Menche J., Barabási A-L, Sharma A. Human symptoms-?disease network. Nature Communications. 2014;5:4212.
15. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. Pubchem substance and compound databases. Nucleic Acids Res. 2016;44(D1):1202–13.
16. Lu Y, Guo Y, Korhonen A. Link prediction in drug-target interactions network using similarity indices. BMC Bioinformatics. 2017;18(1):39.
17. Tong H, Faloutsos C, Pan J-Y. Random walk with restart: fast solutions and applications. Knowl Inf Syst. 2008;14(3):327–46.
18. Lofgren P, Banerjee S, Goel A. Personalized pagerank estimation and search: A bidirectional approach. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining; 2016. p. 163–172. https://doi.org/10.1145/2835776.2835823.
19. Lipscomb CE. Medical subject headings (mesh). Bull Med Libr Assoc. 2000;88(3):265–6.
20. Li J, Lu Z. A new method for computational drug repositioning using drug pairwise similarity. In: IEEE International Conference on Bioinformatics and Biomedicine; 2012. p. 1–4. https://doi.org/10.1109/bibm.2012.6392722.
21. van der Vaart A, van Zanten H. Information rates of nonparametric gaussian process methods. J Mach Learn Res. 2011;12(6):2095–119.
22. Zhang F, Gong T, Lee VE, Zhao G, Rong C, Qu G. Fast algorithms to evaluate collaborative filtering recommender systems. Knowl-Based Syst. 2016;96:96–103.
23. Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. PLoS Comput Biol. 2010;6(1):1000641.
24. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods. 2012;9:471.
25. Wang J, Zuo R. Identification of geochemical anomalies through combined sequential gaussian simulation and grid-based local singularity analysis. Comput Geosci. 2018;118:52–64.
26. Yu S, Zhang A, Li H. A review of estimating the shape parameter of generalized gaussian distribution. J Comput Inf Syst. 2012;8(21):9055–64.
27. Xie M, Hwang T, Kuang R. Prioritizing disease genes by bi-random walk. Lect Notes Comput Sci (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2012;7301 LNAI(PART 2):292–303.
28. Yu G, Fu G, Wang J, Zhao Y. Newgoa: Predicting new go annotations of proteins by bi-random walks on a hybrid graph. IEEE/ACM Trans Comput Biol Bioinforma. 2018;15(4):1390–402.
29. Estrada E. Generalized walks-based centrality measures for complex biological networks. J Theor Biol. 2010;263(4):556–65.
30. Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. J Am Soc Inf Sci Technol. 2007;58(7):1019–31.
31. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ. The comparative toxicogenomics database: update 2017. Nucleic Acids Res. 2016;45(D1):972–8.
32. Breckenridge A, Orme M, Serlin M, Maciver M. Labetalol in essential hypertension. Br J Clin Pharmacol. 1982;13(S1):37–9.
33. Greminger P, Foerster E, Vetter H, Baumgart P, Vetter W. Minoxidil and captopril in severe hypertension. Klin Wochenschr. 1986;64(7):327–32.
34. Khan SS, Cuttica MJ, Beussink-Nelson L, Kozyleva A, Sanchez C, Mkrdichian H, Selvaraj S, Dematte JE, Lee DC, Shah SJ. Effects of ranolazine on exercise capacity, right ventricular indices, and hemodynamic characteristics in pulmonary arterial hypertension: A pilot study. Pulm Circ. 2015;5(3):547–56.
35. Setaluri V, Clark AR, Feldman SR. Transmittance properties of flurandrenolide tape for psoriasis: Helpful adjunct to phototherapy. J Cutan Med Surg. 2000;4(4):196–8.
36. Chen X. Katzlda: Katz measure for the lncrna-disease association prediction. Sci Rep. 2015;5:16840.
37. Yang X, Gao L, Guo X, Shi X, Wu H, Song F, Wang B. A network based method for analysis of lncrna-disease associations and prediction of lncrnas implicated in diseases. PLOS ONE. 2014;9(1):1–10.
38. Katz L. A new status index derived from sociometric analysis. Psychometrika. 1953;18(1):39–43.
39. Krauthammer M, Kaufmann CA, Gilliam TC, Rzhetsky A. Molecular triangulation: Bridging linkage and molecular-network information for identifying candidate genes in alzheimer's disease. Proc Nat Acad Sci. 2004;101(42):15148–53.
40. Radivojac P, Peng K, Clark WT, Peters BJ, Mohan A, Boyle SM, Mooney SD. An integrated approach to inferring gene-disease associations in humans. Proteins. 2008;72(3):1030–7.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.