BRWMDA:Predicting Microbe-Disease Associations Based on Similarities and Bi-Random Walk on Disease and Microbe Networks

Cheng Yan, Guihua Duan, Fang-Xiang Wu, Yi Pan, and Jianxin Wang

Abstract—Many current studies have evidenced that microbes play important roles in human diseases. Therefore, discovering the associations between microbes and diseases is beneficial to systematically understanding the mechanisms of diseases, diagnosing, and treating complex diseases. It is well known that finding new potential microbe-disease associations via biological experiments is a time-consuming and expensive process. However, the computation methods can provide an opportunity to effectively predict microbe-disease associations. In recent years, efforts toward predicting microbe-disease associations are not in proportional to the importance of microbes to human diseases. In this study, we develop a method (called BRWMDA) to predict new microbe-disease associations based on similarity and improving bi-random walk on the disease and microbe networks. BRWMDA integrates microbe network, disease network, and known microbe-disease associations into a single network. After calculating the Gaussian Interaction Profile (GIP) kernel similarity of microbes based on known microbe-disease associations, the microbe network is obtained by adjusting the similarity with the logistics function. In addition, the disease network is computed by the similarity network fusion (SNF) method with the symptom-based similarity and the GIP kernel similarity based on known microbe-disease associations. Then, these two networks of microbe and disease are connected by known microbe-disease associations. Based on the assumption that similar microbes are normally associated with similar diseases and vice versa, BRWMDA is employed to predict new potential microbe-disease associations via random walk with different steps on microbe and disease networks, which reasonably uses the similarity of microbe network and disease network. The 5-fold cross validation and Leave One Out Cross Validation (LOOCV) are adopted to assess the prediction performance of our BRWMDA algorithm, as well as other competing methods for comparison. 5-fold cross validation experiments show that BRWMDA obtained the maximum AUC value of 0.9087, which is again superior to other methods of 0.9025(NGRHMDA), 0.8797 (LRLSHMDA), 0.8571 (KATZHMDA), 0.7782 (HGBI), and 0.5629 (NBI). In addition, BRWMDA also outperforms other methods in terms of LOOCV, whose AUC value is 0.9397, which is superior to other methods of 0.9111(NGRHMDA), 0.8909 (LRLSHMDA), 0.8644 (KATZHMDA), 0.7866 (HGBI), and 0.5553 (NBI). Case studies also illustrate that BRWMDA is an effective method to predict microbe-disease associations.

Index Terms—Microbe, disease, microbe-disease associations, gaussian interaction profile, bi-random walk

1 Introduction

Many studies have demonstrated that a variety of microbial communities and their genes (microbiome) widely exist throughout the human body and play important roles in human health and diseases [1], [2]. The beneficial associations between the host and microbiota are necessary to maintain the host health [3]. Furthermore, the diversity of microbes within human body associated with several human diseases. For example, low diversity and obesity, as well as high diversity and bacterial vaginosis [4], [5], [6]. So far, we have recognized more than 1000 phenotypes of microbiota of the human gastrointestinal tract (GIT) that can benefit human health via producing indispensable vitamins [7]. A diversity microbial community mainly composes of bacteria, archea, fungi and so on, and the number of bacteria cells in an adult intestine can reach 10^{14}, which is ten times to the number of human cells [3], [8]. The microbes can also be thought of as the important organisms to human body, and even humans are the superorganisms that have microbial and human attributes [9].

With the development of high-throughput sequencing technology and microbiology, more insights can be provided for studying new associations between microbes and diseases [8], [10]. Microorganisms are beneficial to human health mainly by promoting the development of the immune

- *C. Yan is with the School of Computer Science and Engineering, Central South University, Changsha 410083, China, and also with the School of Computer and Information, Qiannan Normal University for Nationalities, Duyun, Guizhou 558000, China. E-mail: yancheng01@mail.csu.edu.cn.*
- *G. Duan and J. Wang are with the School of Computer Science and engineering, Central South University, Changsha 410083, China. E-mail: {duangh, jxwang}@mail.csu.edu.cn.*
- *F. Wu is with the Division of Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK S7N5A9, Canada. E-mail: faw341@mail.usask.ca.*
- *Y. Pan is with the Department of Computer Science, Georgia State University, Atlanta, GA 30302. E-mail: yipan@gsu.edu.*

Manuscript received 21 Mar. 2018; revised 29 Jan. 2019; accepted 12 Mar. 2019. Date of publication 26 Mar. 2019; date of current version 7 Oct. 2020. (Corresponding authors: Guihua Duan and Jianxin Wang.)
Digital Object Identifier no. 10.1109/TCBB.2019.2907626

system [11], [12], altering behavior through regulated metabolites [13], protecting against pathogens via regulating the equilibrium between the large bacterial flora and the immune system [14], regulating body homeostasis and maintaining an ecological balance [15], participating in the process of drug metabolism [16] and so on. In addition, environmental factors also can influence the microbial communities of human body. For instance, the remarkable diversity of the human microbiome is associated with dietary intake habit [17], [18]. Smoking can influence the composition of microbiome within human body and thus increase the risk of diseases [19].

In recent years, there are many microbiome projects committed to study composition profiles of microbiome community, function and diversity of human microbiome, and the associations between microbes and diseases. In order to understand the mechanism of microbes and diseases, Human Microbiome Project (HMP) has generated resources to facilitate characterization of the human microbiota, which encounters an estimated 81-99 percent of the genera, enzyme families and community configurations occupied by the healthy Western microbiome [8]. MetaHIT constructs the catalog of the microbial genes from our intestinal tract that provides the basis for characterization of the gut microbial communities [20]. In addition, Tara Oceans Project has been thought as the milestone for researching in ocean ecology and microbial communities [21]. Human Microbe-Disease Association Database (HMDAD:http://www.cuilab.cn/hmdad) constructs a microbe-disease network based on text mining, which helps study microbe-disease associations [22].

Recently, some computational methods have been developed to study the association mechanism of microbiota and diseases. A tool to identify system-level variations of human microbiome associated with diseases that have been provided by integrating metagenomic data with community-level metabolic network [23]. Taking into account the abundance of each microbial protein, a computational method to predict the impact of microbial proteins in human biological events has been developed [24]. A novel computational framework (called MetaModules) has been developed to identify key functional subnetworks in microbiome-related diseases [25]. In addition, some methods have been developed to predict the associations of microbes and diseases. KATZHMDA is the computational method to predict microbe-disease associations [26], which integrates the Gaussian interaction profile (GIP) kernel similarity of microbes, the mean similarity of symptom-based similarity and GIP similarity of diseases and known microbe-disease associations. It is a network-based method to predict the associations by considering the number of walks between nodes and walk lengths in a heterogeneous network as effective similarity metrics. BMCMDA used binary matrix completion to predict potential microbe-disease associations [27]. LRLSHMDA as a semi-supervised computational model of Laplacian Regularized Least Squares for predicting microbe-disease associations has been developed based on GIP kernel similarity [28]. Huang et al. developed an integrated method (NGRHMDA) to predict microbe-disease associations based on neighbor-based collaborative filtering and a graph-based scoring method [29]. By integrating phylogenetic, metallic, and transporter profiles of microbes, RSNMF (Robust Symmetric Nonnegative Matrix Factorization) analyzed the similarities among microbiome samples [30]. Although some significant progresses have been made in understanding associations between microbes and diseases, more efforts are in strong need to study it according to the important level of microbes to human diseases. The comprehensive understanding of associations of microbes with human diseases remains limited and their mechanisms remain largely unknown. However, modern biomedical studies are in a strong hope in providing computational models for predicting effectively potential microbe-disease associations to understand the mechanisms of diseases and improve the efficiency of biological experiments.

In this study, we develop a method (called BRWMDA) to predict new potential microbe-disease associations based on bi-random walk on the networks of microbes and diseases. The microbe network is constructed from the improved result by logistic function with GIP kernel similarity of microbes based on known microbe-disease associations profiles. The disease network is constructed from two aspects. First, the symptom-based similarity is computed by the co-occurrence between diseases and symptom terms. Then the GIP kernel similarity of diseases is calculated based on known microbe-disease association profiles. Second, The disease network is constructed by SNF method with GIP similarity and symptom-based similarity [31], [32]. Based on the known microbe-disease associations, We then connect two networks of diseases and microbes by known microbe-disease associations. Based on basic assumption that similar microbes are preferring to associate with similar diseases and vice versa, BRWMDA is adopted to predict new potential microbe-disease associations via an improving bi-random walk on disease and microbe network. We adopt 5-fold cross validation (5-fold CV) and Leave One Out Cross Validation (LOOCV) in terms of AUC (area under the receiver operating characteristics curve) value to evaluate the prediction performance of BRWMDA and other competing methods which include NGRHMDA, LRLSHMDA,KATZHMDA,HGBI [33] and NBI [34]. HGBI can be used to predict new microbe-disease associations based on guilt-by-association principle on the heterogeneous network. NBI is a two-step diffusion model to predict new microbe-disease associations on a bipartite graph which only uses known microbe-disease associations. In order to evaluate the prediction performance of BRWMDA method, BRWMDA and other competing methods were assessed by 5-fold cross validation (5-fold CV) and Leave One Out Cross Validation in terms of AUC (area under the receiver operating characteristics curve). In 5-fold CV, BRWMDA obtains the maximum AUC value of 0.9087, and is superior to other methods of 0.9025 (NGRHMDA), 0.8797 (LRLSHMDA), 0.8571 (KATZHMDA), 0.7782 (HGBI), 0.5629 (NBI). Furthermore, BRWMDA also outperforms other methods with LOOCV, whose AUC value is 0.9397, and is again superior to other methods of 0.9111(NGRHMDA), 0.8909 (LRLSHMDA), 0.8644 (KATZHMDA), 0.7866 (HGBI), 0.5553 (NBI). Case studies show that BRWMDA is an effective method to predict microbe-disease associations and can be helpful in understanding mechanisms of diseases, treating human diseases and developing drugs.

2 Materials and Methods

2.1 Materials

The known microbe-disease data are downloaded from HMDAD database (http://www.cuilab.cn/hmdad), which curates microbe-disease associations by using text mining technology. There are 483 microbe-disease associations, 39 human diseases and 292 microbes [22]. The final number of microbe-disease associations is 450 by removing the duplicated data without considering different evidences. This dataset also can download from KATZMDA [26].

In addition, the symptom data of diseases are needed to compute the symptom-based similarity of diseases, which can be downloaded from HSDN (symptom-based human disease network) and are generated from medical bibliographic records and the related Medical Subject Headings (MeSH) metadata [35] from PubMed database [36], [37]. Furthermore, the symptom-based similarity of diseases also can be obtained from KATZMDA [26].

2.2 Construction of Microbe Network

Let $M = \{m_1, m_2, \ldots, m_{N_m}\}$ be the set of N_m microbes and $D = \{d_1, d_2, \ldots, d_{N_d}\}$ be the set of N_d diseases. The matrix $Y \in R^{N_d \times N_m}$ represents known disease-microbe associations, in which the value of y_{ij} is 1 if disease i and microbe j have a known association, otherwise is 0. In addition, the edges of the microbe network is calculated by the GIP kernel similarity of microbes [26], [38], [39], [40]. Specifically, the weight of edge $S_{km}(m_i, m_j)$ between microbe m_i and microbe m_j can be computed as follows:

$$\begin{aligned} S_{km}(m_i, m_j) &= K_{GIP,m}(m_i, m_j) \\ &= exp(-\gamma_m ||y_{m_i} - y_{m_j}||^2) \end{aligned} \tag{1}$$

$$\gamma_m = \gamma'_m / \left(\frac{1}{N_m} \sum_{i=1}^{N_m} ||y_{m_i}||^2 \right), \tag{2}$$

where $y_{m_i} = \{y_{1i}, y_{2i}, \ldots, y_{N_d i}\}^T$ is the association profile of microbe m_i while $y_{m_j} = \{y_{1j}, y_{2j}, \ldots, y_{N_d j}\}^T$ is the association profile of microbe m_j, respectively. γ'_m is set to be 1 in this study [26], [32], [38], [41].

Inspired by successful application to adjusting the similarity by the logistic functions, we also adjust S_m to improve the similarity of microbes [42], [43]. The logistic function is defined as follow:

$$L(x) = \frac{1}{1 + e^{(cx+d)}}, \tag{3}$$

where x is the similarity value of elements of matrix S_{km} and the final microbe similarity matrix is computed by $S_m(m_i, m_j) = L(S_{km}(m_i, m_j)), \forall i, j \in [1, N_m], i \neq j$. Parameters c and d are used to control the adjustment of similarity. Based on previous studies [42], [43], we also set d as $log(9999)$ which represents $L(0) = 0.0001$. In addition, the values of parameter c are set by the 5-fold CV and LOOCV, respectively.

2.3 Construction of Disease Network

For the diseases network, the edge weights are calculated based on the similarity of diseases. Inspired by the SNF (similarity network fusion) method of successful applications in other relative issue [32], we also adopt the SNF method to compute the values of all edges of the disease network based on the symptom-based similarity and GIP kernel similarity of diseases in this study.

First, we compute the GIP kernel similarity matrix $K_{GIP,d} \in R^{N_d \times N_d}$ of diseases from the known microbe-disease associations. We further compute the symptom-based similarity matrix D_{symsim} of diseases according to the symptom profile of diseases [36], [44]. Then we adopt the SNF method to fuse these two similarity kernels to construct a more comprehensive disease network in which $S_d(i, j)$ represents the weight of edge (i, j).

In the SNF process, we first adopt two simple processes to these two similarity matrices of disease: (1) making the matrices D_{symsim} and $K_{GIP,d}$ become symmetrical by $D_{symsim} = (D_{symsim} + D^T_{symsim})/2$ and $K_{GIP,d} = (K_{GIP,d} + K^T_{GIP,d})/2$; (2) adding a small multiple of identity matrix to matrices D_{symsim} and $K_{GIP,d}$ to make sure them positive semi-definite [38]. Then we obtain the row-normalized matrices $P^{(1)}$ and $P^{(2)}$ from the disease matrices D_{symsim} and $K_{GIP,d}$ via dividing by the sum of the rows, respectively. We further calculate the resultant matrices $S^{(1)}$ and $S^{(2)}$ from $P^{(1)}$ and $P^{(2)}$ by the K nearest neighbors (KNN) method, respectively, as follows:

$$S^w(d_i, d_j) = \begin{cases} \frac{P^w(d_i, d_j)}{\sum_{d_k \in N^w(d_i)} P^w(d_i, d_k)}, & d_j \in N^w(d_i) \\ 0, & otherwise, \end{cases} \tag{4}$$

for $w = 1, 2$ and where $N^w(d_i)$ represents the top K similar neighbors of the current diseases d_i, and the number of K is set to be 4 in this study by the cross validation. Then the main process of fusion operation in SNF method is iteratively updating matrices $P^{(1)}$ and $P^{(2)}$ as follows [31]:

$$P^{(1)}_{t+1} = S^{(1)} \times P^{(2)}_t \times (S^{(1)})^T \tag{5}$$

$$P^{(2)}_{t+1} = S^{(2)} \times P^{(1)}_t \times (S^{(2)})^T, \tag{6}$$

where t is the number of iterations, and the maximum iteration number is N_t which is set to be 5 in this study. The initial matrices of $P^{(1)}_{t=1}$ and $P^{(2)}_{t=1}$ are defined as $P^{(1)}_{t=1} = P^{(1)}$ and $P^{(2)}_{t=1} = P^{(2)}$, respectively. After N_t steps, the matrix S_d of disease network is obtained by the average value of matrices $P^{(1)}_{N_t}$ and $P^{(2)}_{N_t}$ ($S_d = \frac{P^{(1)}_{N_t} + P^{(2)}_{N_t}}{2}$).

2.4 Bi-Random Walk on Disease and Microbe Networks

Driven by the successful applications of BRWH (Bi-random walk on a heterogeneous network) in predicting drug-disease associations and discovering miRNA-disease associations [42], [45], [46], we develop a new computational algorithm (called BRWMDA) based on bi-random walk on the disease and microbe networks to predict microbe-disease associations.

The microbe-disease associations can be represented by a bipartite $G(V, E)$, where $V(G)$ is the set of vertices of G and $E(G)$ is the edge set of G. The weight value of edge e_{ij} is 1 if disease d_i and microbe m_j has a known associations, otherwise is 0. Therefore, a network can be constructed properly by connecting the microbe network and disease network via microbe-disease associations.

BRWMDA for predicting microbe-disease associations can be formulated as a random walk on the microbe network and disease network simultaneously based on assumption that the similar microbes tend to associate with similar diseases, and vice versa. The important feature of BRWMDA is the different steps of random walks in microbe network and disease network that can reasonably use the similarity of microbes and diseases.

Let matrix $M_M \in R^{N_m \times N_m}$ be the column-normalized adjacency matrix of the microbe network. The element $M_M(i,j)$ represents the column-normalized weight of the edge between microbe i and microbe j, which can be calculated as follows:

$$M_M(i,j) = \begin{cases} \frac{S_m(i,j)}{\sum_{k=1}^{k=N_m} S_m(k,j)}, & if \sum_{k=1}^{k=N_m} S_m(k,j) > 0 \\ 0, & otherwise, \end{cases} \tag{7}$$

where $S_m(i,j)$ is the weight of the edge between microbe i and microbe j in the microbe network.

Similarly, let matrix $M_D \in R^{N_d \times N_d}$ be the column-normalized adjacency matrix of the disease network. The column-normalized weight $M_D(i,j)$ of the edge between disease i and disease j can be computed as follows:

$$M_D(i,j) = \begin{cases} \frac{S_d(i,j)}{\sum_{k=1}^{k=N_d} S_d(k,j)}, & if \sum_{k=1}^{k=N_d} S_d(k,j) > 0 \\ 0, & otherwise, \end{cases} \tag{8}$$

where $S_d(i,j)$ is the weight of edge between disease i and disease j in the disease network.

After obtaining the matrices of M_D and M_M, we further compute the initial association probability for diseases (microbes) which have no known associations with microbes (diseases). In this study, we used the nearest neighbor of microbes (diseases) to calculate their initial association scores. For disease d_i,the computation process is defined as follows:

$$Y(i,:) = Y(j,:) * M_D(i,j), \tag{9}$$

where disease d_j is the nearest neighbor of disease d_i. Similarly, for microbe m_i,the computation process is also defined as follows:

$$Y(:,i) = Y(:,j) * M_M(i,j), \tag{10}$$

where microbe m_j is the nearest neighbor of microbe m_i.

Let R be a $N_d \times N_m$ matrix, whose element $r(i,j)$ represents the association probability of disease i and microbe j. The iterative random walk process of microbe network and disease network are defined as follows.

In the disease network:

$$L\_R_t = (1-\alpha) * M_D * R_{t-1} + \alpha * Y. \tag{11}$$

In the microbe network:

$$R\_R_t = (1-\alpha) * R_{t-1} * M_M + \alpha * Y, \tag{12}$$

where t represents the iteration steps, α is the decay factor of random walk whose value is from 0 to 1. L_R_t and R_R_t represent the predicted disease-microbe associations via random walk over the disease network and the microbe network, respectively. L_{num} and R_{num} are the number of steps on the disease network and the microbe network that have been iterated, respectively. Considering their different topology and structure characteristics, we set I_l and I_r to be the number of maximal number iterations in the disease network and the microbe network, respectively. In the iteration process, R_{t+1} is the averaged output from the disease network and the microbe network in each step.

Algorithm 1 describes the BRWMDA method for predicting potential microbe-disease associations. First, the adjacency matrix S_{km} of the microbe network is calculated by GIP similarity $K_{GIP,m}$ with Y. Then we further compute the S_m by S_{km} with the logistic function. Second, the adjacency matrix S_d of the disease network is calculated by SNF method with $K_{GIP,d}$ and D_{symsim}. Third, matrices S_d and S_m are column-normalized to matrices M_D and M_M, and the association probability of microbes and diseases which have no known associations is initialized by their nearest neighbors. Finally, with $R_0 = Y/sum(Y(:))$ and via the iteration process, the predicted disease-microbe matrix F is calculated.

Algorithm 1. BRWMDA

Input: Disease-microbe adjacency matrix Y, symptom-based matrix D_{symsim}, and parameter K, N_t, α, I_l and I_r
Output: predicted disease-microbe matrix F
BRWMDI(Y, D_{symsim}, K, N_t, α, I_l and I_r)
1: Calculate matrix S_{km} of microbe network by GIP similarity $K_{GIP,m}$ with Y;
2: Obtain the matrix S_m by logistic function;
3: Calculate GIP similarity matrix $K_{GIP,d}$ of diseases with Y;
4: Calculate matrix S_d of disease network by SNF method with $K_{GIP,d}$ and D_{symsim};
5: Obtain the matrix M_D by col-normalized matrix S_d;
6: Obtain the matrix M_M by col-normalized matrix S_m;
7: Initializing association probability for diseases and microbes which have no associations by their nearest neighbors;
8: $R_0 = Y/sum(Y(:))$;
9: //Iteration process;
10: Max_Iter = max([I_l, I_r]);
11: for $t = 1 : Max\_Iter$
12: $L_{flag} = 0$;
13: $R_{flag} = 0$;
14: //Random walk in disease network;
15: if($t \leq I_l$)
16: $L_{flag} = 1$;
17: $L\_R_t = (1-\alpha) * M_D * R_{t-1} + \alpha * Y$;
18: endif
19: //Random walk in microbe network;
20: if($t \leq I_r$)
21: $R_{flag} = 1$;
22: $R\_R_t = (1-\alpha) * R_{t-1} * M_M + \alpha * Y$;
23: endif
24: $R_{t+1} = (L_{flag} * L\_R_t + R_{flag} * R\_R_t)/(L_{flag} + R_{flag})$;
25: end
26: $F = R_{t+1}$;
27: Return F;

3 Experiments and Results

3.1 Performance Evaluation

In order to systematically evaluate performance of BRWMDA for predicting microbe-disease associations, 5-fold CV was used on HMDAD database. This validation

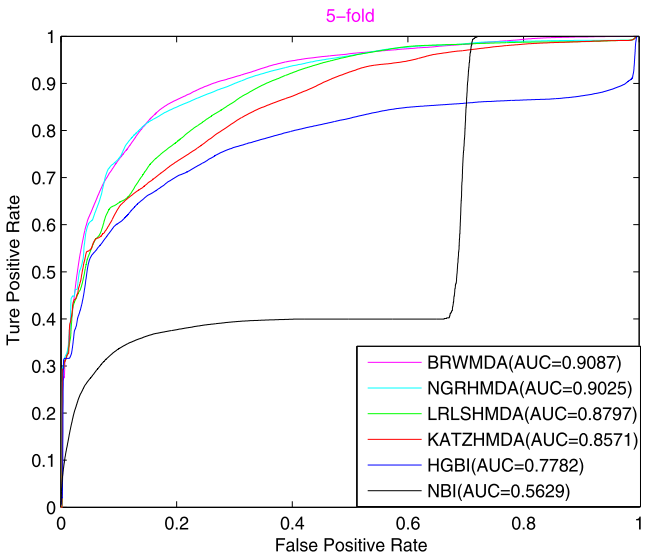

Fig. 1. The AUC curves of six methods on HMDAD in the 5-fold cross validation.

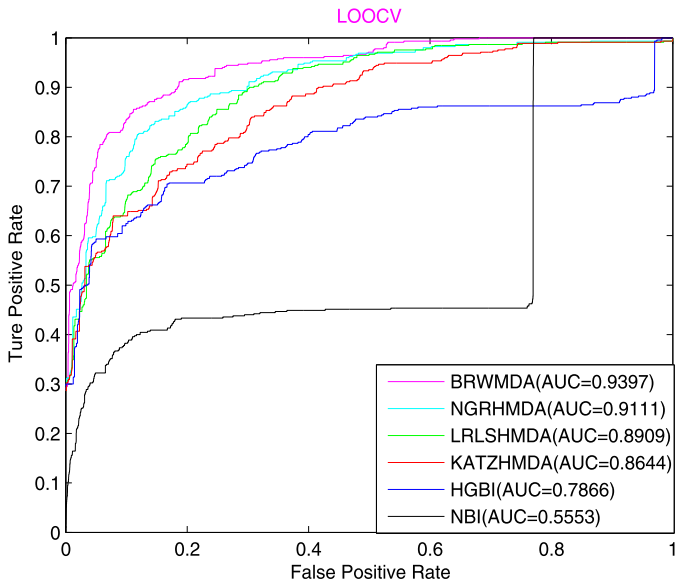

Fig. 2. The AUC curves of six methods on HMDAD in the LOOCV.

method was widely used in inferring miRNA-disease associations and so on [47], [48]. In the 5-fold CV, the known microbe-disease associations are randomly divided into 5 groups. Then one group is in turn chosen as the test samples while the rest 4 groups are regarded as the training samples. The training process which includes similarity computation of microbes and diseases and bi-random walk on disease and microbe networks. In this study, we perform 5-fold cross validation 100 times.

In addition, LOOCV evaluation is also used as another method to evaluate the performance of BRWMDA for predicting microbe-disease associations. In this evaluation, each known microbe-disease association is in turn chosen as the test sample while the remaining known microbe-disease associations are regarded as the training samples in the training process.

The test samples are compared with all candidate samples (unknown microbe-disease associations) in 5-fold CV and LOOCV. The evaluation metric is AUC. Computation model has no prediction ability when the value of its AUC is less than or equal to 0.5. Larger AUC indicates the better prediction ability and 1 represents the perfect prediction approach.

3.2 Comparison with Other Methods

In this study, we compare BRWMDA with other five competing methods, namely, NGRHMDA [29], LRLSHMDA [28], KATZHMDA [26], HGBI [33], NBI [34]. NGRHMDA predicted microbe-disease associations by integrating two single recommendation models (neighbor-based collaborative filtering and graph-based scoring method), which used the GIP kernel similarity and symptom-based disease similarity. LRLSHMDA is a semi-supervised computational method based on Laplacian Regularized Squares classifier, which used only the GIP kernel similarities of microbes and diseases. KATZHMDA is a method to predict microbe-disease associations by considering the number of walks between nodes and walk lengths in a heterogenous. Furthermore, HGBI and NBI are the classic approaches can predict microbe-disease associations. The same data and validation methods are used for BRWMDA and other methods.

In 5-fold CV, Fig. 1 shows the prediction performance of BRWMDA, NGRHMDA, LRLSHMDA, KATZHMDA, HGBI and NBI methods on HMDAD dataset. BRWMDA achieves AUC of 0.9087, which is better than other methods (NGRHMDA: 0.9025, LRLSHMDA: 0.8797, KATZHMDA: 0.8571, HGBI: 0.7782, NBI: 0.5629).

Fig. 2 shows the comparison result between BRWMDA and other five methods (NGRHMDA, LRLSHMDA, KATZHMDA, HGBI and NBI) on HMDAD dataset via the LOOCV. The prediction performance of BRWMDA is also the best in terms of AUC as its AUC value is 0.9397, compared with other methods (NGRHMDA: 0.9111, LRLSHMDA: 0.8909, KATZHMDA: 0.8644, HGBI: 0.7866, NBI: 0.5553).

We can see from Figs. 1 and 2 that the AUC of NBI method is steep when the FPR reaching a certain value. The reason of this result is that NBI method has not used similarity information of diseases and microbes, and then even yields the probability scores of some microbe-disease pairs to have a small difference, which leads to the TPR rapid growth when FPR reaches a certain value.

Because the top-ranked microbe-disease associations are more important, we also measure the top-ranked results predicted by BRWMDA and other five methods in LOOCV. Fig. 3 shows the numbers of correctly identified known microbe-disease associations based on various top portions. We can see from Fig. 3 that except NBI method, the numbers of correctly identified known microbe-disease associations of other five methods are not much different on top 1, 10 and 20. The reason is that 129 associated microbes of Type 1 diabetes are ranked top 1 in LOOCV. Type 1 diabetes associates with 167 microbes, in which 129 microbes only associate with this disease. Among 450 known microbe-disease associations, 129 associations can be correctly identified by BRWMDA at the top one. In addition, BRWMDA can identify more known associations than other five methods at top 50, 100, 200 and 500. At top 50, the numbers of identified associations is 164 and the ratio is 36.44 percent. In addition, the numbers of identified associations is 221 and the ratio is 49.11 percent at top 100.

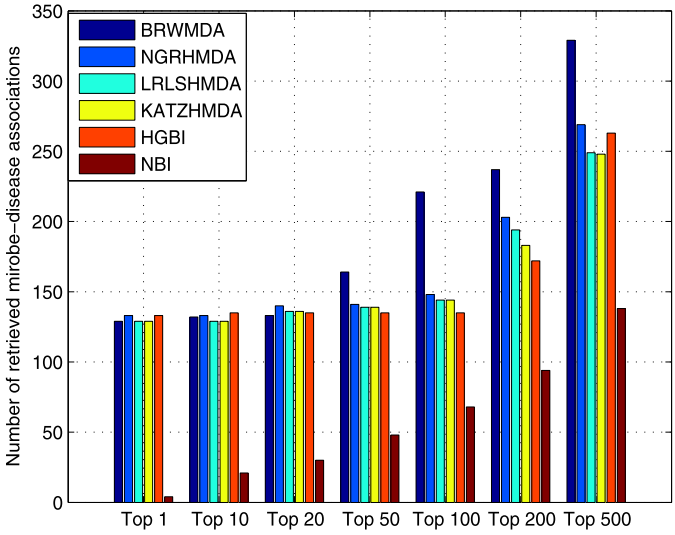

Fig. 3. The number of correctly predicted by six methods on HMDAD in LOOCV.

TABLE 1
The AUC Value of BRWMDA under Different Settings of c in 5-fold CV and LOOCV, the Best Result is in Bold Face

5-fold CV					
c	-3	-5	-7	-9	-11
AUC	0.8888	0.8998	0.8930	0.8748	0.8813
c	-13	-15	-17	-19	-21
AUC	0.8998	0.9068	**0.9087**	0.9054	0.9020
LOOCV					
c	-3	-5	-7	-9	-11
AUC	0.9016	0.9225	0.9378	**0.9397**	0.9386
c	-13	-15	-17	-19	-21
AUC	0.9369	0.9313	0.9253	0.9206	0.9163

3.3 Parameters Analysis for c, K, N_t, α, I_l and I_r

To evaluate the influence of parameters in BRWMDA, we analyze parameter c of logistic function which is used to adjust the similarity of microbes, parameters K and N_t of SNF method and parameters α, I_l and I_r of bi-random walk on disease and microbe networks. we first analyze the parameter c. Then we analyze two parameters: K and N_t of SNF method to fuse two similarity kernels of diseases to construct a more comprehensive similarity kernel, respectively. Finally we further analyze three parameters: decay factor α, maximum numbers I_l and I_r of bi-random walk on a disease and microbe networks, respectively. The parameters K, N_t, α, I_l and I_r are set to be the default values (4, 5, 0.5, 1 and 3) when analyzing parameter c. In addition, because the S_m is calculated from the GIP similarity which relates the known microbe-disease associations, we set different value of c by 5-fold CV and LOOCV, respectively. We set the defaults values (-17, 0.5, 1 and 3) of c, α, I_l and I_r when analyzing parameters K and N_t by 5-fold CV. In addition, we also set the default values (-17, 4, 5) of c, K and N_t when analyzing parameters α, I_l and I_r by 5-fold CV.

According to previous study, we also select the optimal value of regression coefficient parameter c of logistic function by 5-fold CV and LOOCV [43]. By considering the similarity of microbes is computed from the GIP similarity, we set different value of c in 5-fold CV and LOOCV, respectively. Table 1 shows that the prediction performance is best when c is set to be -17 in 5-fold CV, which implies that similarity values below 0.13 are assigned with very low probability (≤ 0.001). However, when c is set to be -9, BRWMDA achieves the best prediction performance in LOOCV, which implies that similarity values below 0.26 are assigned with very low probability (≤ 0.001). Therefore, we set the default value of c to be -17 and -9 in 5-fold CV and LOOCV, respectively.

TABLE 2
The AUC of BRWMDA under Different Settings of K in 5-fold CV, the Best Result is in Bold Face

K	1	2	3	4	5
AUC	0.8975	0.8976	0.9054	**0.9087**	0.9086
K	6	7	8	9	10
AUC	0.9070	0.9067	0.9070	0.9073	0.9071

TABLE 3
The AUC of BRWMDA under Different Settings of N_t in 5-fold CV, the Best Result is in Bold Face

N_t	1	2	3	4	5
AUC	0.9057	0.9084	0.9086	0.9085	**0.9087**
N_t	6	7	8	9	10
AUC	0.9082	0.9078	0.9077	0.9074	0.9071

With parameter N_t of 5, Table 2 shows that the sensitivity of the prediction performances of BRWMDA when K ranges from 1 to 10. We can see from Table 2 that the prediction performance is the best when K is set to be 4. In addition, the prediction performance has little ascending trend when K ranges from 1 to 4, and has little descending trend when K ranges from 4 to 10. So in this study, we set the default value of K to be 4.

Furthermore, Table 3 describes the prediction performances of BRWMDA with different values of N_t when K is set to be 4. we can see from Table 3 that the variation of N_t ranging from 1 to 10 has little effect on prediction performances of BRWMDA and the best prediction performance is obtained when it is set to be 5. Therefore, we set the default value of N_t to be 5.

With $I_l = 1$ and $I_r = 3$, as shown in Fig. 4, the variation of α ranging from 0.1 to 0.5 has little effect via the 5-fold CV and achieves the best prediction performance (AUC: 0.9087) when the value of α is 0.5. In addition, BRWMDA method has little prediction ability when the value of α is set to be 1, which is not suitable to random walk process based on Equations (11) and (12). Therefore, based on the above AUC values, α is fixed to be 0.5 in this study.

We implement a grid searching method to analyze parameters I_l and I_r via 5-fold cross validation when $\alpha = 0.5$. The results in Fig. 5 show that BRWMDA achieves the best performance (AUC: 0.9087) when $I_l = 1$ and $I_r = 3$. In this study we accordingly set the parameters as $\alpha = 0.5$, $I_l = 1$ and $I_r = 3$ in BRWMDA.

3.4 Case Study

To further evaluate the prediction performance of BRWMDA, Type 1 diabetes and Liver cirrhosis are investigated

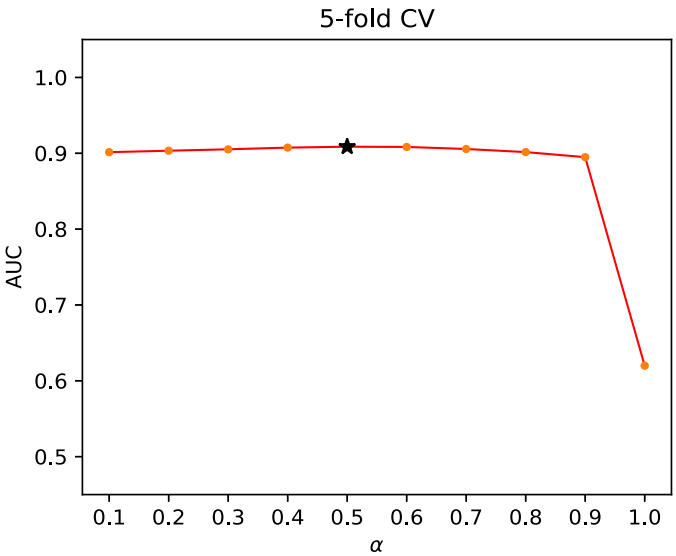

Fig. 4. The 5-fold CV prediction performance of various parameter values of α ranging from 0.1 to 1.0 with 0.1 increments, the sign * represents the default value.

to infer their associated microbes. Due to the lack of other independent microbe-disease associations database, we validate the predicted associations via existing studies. In addition, the prediction results of other five competing methods are also provided.

Type 1 diabetes is a common, multifactorial disease and accounts for only about 5-10 percent of all cases of diabetes [49], [50]. We can see from Table 4 that there are 3 of top 5 predicted microbes of Type 1 diabetes by BRWMDA that are validated by literature. However, HGBI and BNI methods all have 2 of top 5 predicted microbes of Type 1 diabetes are validated. Furthermore, NGRHMDA, LRLSHMDA and KATZHMDA all have same predicted top 5 microbes of Type 1 diabetes, and all have only one validated. Infants born by cesarean delivery are at increased risk of type 1 diabetes and associated with Verrucomicrobiaceae [51]. In recent years, some effective treatments to Autism Spectrum

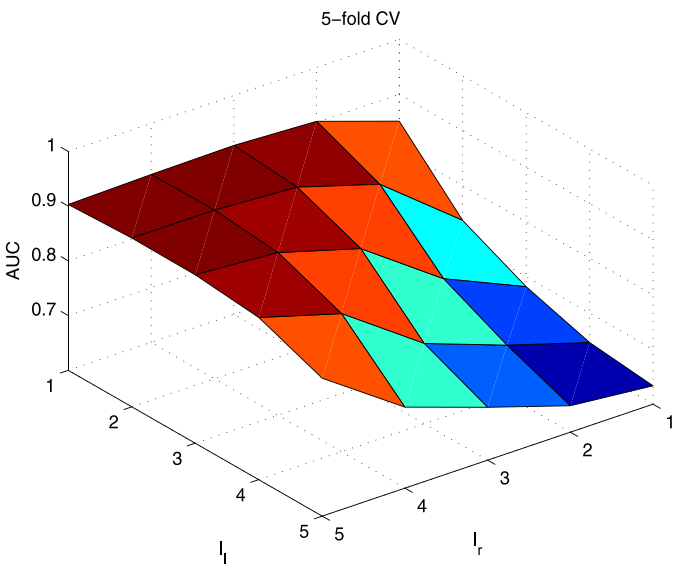

Fig. 5. The 5-fold CV prediction performance of various parameter values of l_l and l_r ranging from 1 to 5 with one increments.

TABLE 4
The Validation Result of Top Five New Microbe-Disease Associations of Type 1 Diabetes by six Methods on Previous Studies

disease	method	microbe	rank	source
Type 1 diabetes	BRWMDA	Verrucomicrobiaceae	1	PMID:23401405
		Desulfovibrio	3	PMID:23273906
		Fusobacterium nucleatum	2	PMID:17092237
		Oxalobacter formigenes	4	Unknown
		Tropheryma whipplei	5	Unknown
	NGRHMDA	Alistipes finegoldii	1	Unknown
		Alistipes putredinis	2	Unknown
		Bacteroides eggerthii	3	PMID:20613793
		Bacteroides finegoldii	4	Unknown
		Bacteroides sp. D20	5	Unknown
	LRLSHMDA	Alistipes finegoldii	1	Unknown
		Alistipes putredinis	2	Unknown
		Bacteroides eggerthii	3	PMID:20613793
		Bacteroides finegoldii	4	Unknown
		Bacteroides sp. D20	5	Unknown
	KATZHMDA	Alistipes finegoldii	1	Unknown
		Alistipes putredinis	2	Unknown
		Bacteroides eggerthii	3	PMID:20613793
		Bacteroides finegoldii	4	Unknown
		Bacteroides sp. D20	5	Unknown
	HGBI	Clostridium leptum	1	Unknown
		Enterococcus	2	PMID:24475780
		Verrucomicrobiaceae	3	PMID:23401405
		Alistipes finegoldii	4	Unknown
		Alistipes putredinis	5	Unknown
	NBI	Clostridium coccoides	1	Unknown
		Enterobacteriaceae	2	PMID:24475780
		Clostridium leptum	3	Unknown
		Faecalibacterium prausnitzii	4	Unknown
		Verrucomicrobiaceae	5	PMID:23401405

Disorders have been found, including treating Desulfovibrio overgrowth with aztreonam and preventing onset of type 1 diabetes in diabetes-prone rats [52]. The putative pathogens and health-related specie, such as Fusobacterium nucleatum and so on, are the main component of the periodontal bacteria that associate with Type 1 diabetes [53]. The autoimmune microbiome (Bacteroides) for Type 1 diabetes may be distinctly different from that found in healthy children [54]. Enterobacteriaceae colonization has statistically significant difference in the type 1 diabetes group [55]. A decrease in Enterococcus colonization was identified in type 1 diabetes patients [55].

Liver cirrhosis is often indolent, asymptomatic, and unsuspected until complications of liver disease are present and even leads to end-stage liver disease [56]. From Table 5, there are 2 of top 5 predicted microbes by BRWMDA that are validated in literature,while other methods all have only 1 of top 5 predicted microbes that are confirmed in literature. There is a higher representation of Clostridiales members, families belonging to Verrucomicrobiaceae, which is a significantly lower median cirrhosis dysbiosis ratio in liver cirrhosis patients [57]. Lack of the bacterium Oxalobacter formigenes and recurrent episodes of d-lactic acidosis affect the development of enteric hyperoxaluria which can be associated liver cirrhosis [58]. Compared with healthy controls, patients with liver cirrhosis had fewer Bacteroidetes and Firmicutes which dominated the faecal microbial communities [59]. Aeromonas sepsis is associated with liver cirrhosis, which belongs to the family Aeromonadaceae of the

TABLE 5
The Validation Result of Top Five New Microbe-Disease Associations of Liver Cirrhosis by Six Methods on Previous Studies

disease	method	microbe	rank	source
Liver cirrhosis	BRWMDA	Verrucomicrobiaceae	1	PMID:29515036
		Fusobacterium nucleatum	2	Unknown
		Oxalobacter formigenes	3	PMID:19664961
		Tropheryma whipplei	4	Unknown
		Prevotella copri	5	Unknown
	NGRHMDA	Firmicutes	1	PMID:25079328
		Actinobacteria	2	Unknown
		Acetonema	3	Unknown
		Actinobacillus	4	Unknown
		Aggregatibacter	5	Unknown
	LRLSHMDA	Acidobacteriaceae	1	Unknown
		Aeromonadaceae	2	PMID:18696148
		Alteromonadaceae	3	Unknown
		Anaerovorax	4	Unknown
		Bacillaceae	5	Unknown
	KATZHMDA	Acetonema	1	Unknown
		Actinobacillus	2	PMID:15295960
		Aggregatibacter	3	Unknown
		Akkermansia	4	Unknown
		Anaerofilum	5	Unknown
	HGBI	Bacteroides ovatus	1	Unknown
		Porphyromonadaceae	2	Unknown
		Acidobacteriaceae	3	Unknown
		Aeromonadaceae	4	PMID:17063071
		Alteromonadaceae	5	Unknown
	NBI	Clostridium coccoides	1	Unknown
		Bacteroides vulgatus	2	Unknown
		Actinobacteria	3	PMID:21574172
		Clostridia	4	Unknown
		Porphyromonadaceae	5	Unknown

order Aeromonadales and consists of 14 phenospecies and 17 genomospecies [60]. Actinobacillus is a consequence of end-stage liver disease with cirrhosis and hepatocellular [61]. Aeromonas species is Aeromonadaceae ubiquitous gram-negative bacilli that causes bacteremia mostly in patients with malignancy or other immunocompromising conditions such as serious liver disease (mostly Liver cirrhosis) [62]. Despite the highly diverse bacterial communities and interindividual differences, liver cirrhosis clearly affects the intestinal microbial community (Actinobacteria) [63]. In addition, other microbes have not been validated yet while deserving to be studied via biological experiments.

4 Conclusion

With the development of biotechnology, various studies have indicated that microbes have key impacts on health body and human diseases. Therefore, discovering the potential associations of microbes and diseases is beneficial to comprehensively understanding the association mechanisms of diseases, developing drugs, diagnosing and treating diseases. However, the systematic understanding of mechanisms of microbes and human diseases remains limited. Furthermore, based on the important level to human body, it is very necessary to develop effective methods to identify the associations between them.

In this study, we have developed a BRWMDA approach to predict potential associations of microbes and diseases based on an improving bi-random walk on the microbe network and the disease network. The main contributions of BRWMDA includes three aspects: (1) To reasonable use the GIP similarity of microbes, we adjust it by logistic function which can improve the prediction performance. Based on the feature of GIP similarity, we select two optimal values of regression coefficient parameter c of logistic function in 5-fold CV and LOOCV, respectively. Table 1 shows the prediction sensitivity under different settings of parameter c; (2) In addition, in order to obtain the more reasonable similarity of diseases, we integrate the disease GIP similarity and symptom-based similarity by the SNF method. We also analyze the influence of parameters. Tables 2 and 3 show that the prediction performance of BRWMDA is stable when the parameters K and N_t are changed from 1 to 10, respectively; (3) Finally, in the bi-random walk on disease and microbe networks, we add an improving process which computes the initial association probability for diseases (microbes) that have no known associations with microbes (diseases).

However, there are still some limitations of BRWMDA. First, BRWMDA can not predict potential microbe-disease associations for new microbes because the GIP similarity is the only similarity of microbes in constructing the microbe network. Second, other biological information should be considered, such as functional, metabolic and phylogenetic profiles of microbes, as well as disease functional information and phenotype infromation [64], [65]. In the future, we would like to develop a more effective method for predicting microbe-disease associations by addressing the above limitations.

Acknowledgments

The authors would like to express their gratitude for the support from the National Natural Science Foundation of China under Grant No.61772552, No.61420106009, and No. 61622213.

References

[1] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. Fraser-Liggett, R. Knight, and J. I. Gordon, "The human microbiome project: exploring the microbial part of ourselves in a changing world," *Nature*, vol. 449, no. 7164, 2007, Art. no. 804.
[2] H. M. P. Consortium, et al., "A framework for human microbiome research," *Nature*, vol. 486, no. 7402, pp. 215–221, 2012.
[3] F. Sommer and F. Bäckhed, "The gut microbiotamasters of host development and physiology," *Nature Rev. Microbiology*, vol. 11, no. 4, pp. 227–238, 2013.
[4] P. J. Turnbaugh, M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, et al., "A core gut microbiome in obese and lean twins," *Nature*, vol. 457, no. 7228, pp. 480–484, 2009.
[5] J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, et al., "A human gut microbial gene catalogue established by metagenomic sequencing," *Nature*, vol. 464, no. 7285, pp. 59–65, 2010.
[6] D. N. Fredricks, T. L. Fiedler, and J. M. Marrazzo, "Molecular identification of bacteria associated with bacterial vaginosis," *New England J. Med.*, vol. 353, no. 18, pp. 1899–1911, 2005.
[7] M. Ventura, S. O'flaherty, M. J. Claesson, F. Turroni, T. R. Klaenhammer, D. Van Sinderen, and P. W. O'toole, "Genome-scale analyses of health-promoting bacteria: probiogenomics," *Nature Rev. Microbiology*, vol. 7, no. 1, pp. 61–71, 2009.
[8] H. M. P. Consortium, et al., "Structure, function and diversity of the healthy human microbiome," *Nature*, vol. 486, no. 7402, pp. 207–214, 2012.

[9] S. R. Gill, M. Pop, R. T. DeBoy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett, and K. E. Nelson, "Metagenomic analysis of the human distal gut microbiome," *Sci.*, vol. 312, no. 5778, pp. 1355–1359, 2006.
[10] H. M. J. R. S. Consortium, et al., "A catalog of reference genomes from the human microbiome," *Sci.*, vol. 328, no. 5981, pp. 994–999, 2010.
[11] E. S. Gollwitzer, S. Saglani, A. Trompette, K. Yadava, R. Sherburn, K. D. McCoy, L. P. Nicod, C. M. Lloyd, and B. J. Marsland, "Lung microbiota promotes tolerance to allergens in neonates via pd-l1," *Nature Med.*, vol. 20, no. 6, pp. 642–647, 2014.
[12] S. K. Mazmanian, C. H. Liu, A. O. Tzianabos, and D. L. Kasper, "An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system," *Cell*, vol. 122, no. 1, pp. 107–118, 2005.
[13] E. Y. Hsiao, S. W. McBride, S. Hsien, G. Sharon, E. R. Hyde, T. McCue, J. A. Codelli, J. Chow, S. E. Reisman, J. F. Petrosino, et al., "Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders," *Cell*, vol. 155, no. 7, pp. 1451–1463, 2013.
[14] D. Bouskra, C. Brézillon, M. Bérard, C. Werts, R. Varona, I. G. Boneca, and G. Eberl, "Lymphoid tissue genesis induced by commensals through nod1 regulates intestinal homeostasis," *Nature*, vol. 456, no. 7221, pp. 507–510, 2008.
[15] J. Kreth, Y. Zhang, and M. C. Herzberg, "Streptococcal antagonism in oral biofilms: Streptococcus sanguinis and streptococcus gordonii interference with streptococcus mutans," *J. Bacteriology*, vol. 190, no. 13, pp. 4632–4640, 2008.
[16] W. Jia, H. Li, L. Zhao, and J. K. Nicholson, "Gut microbiota: A potential new territory for drug targeting," *Nature Rev. Drug Discovery*, vol. 7, no. 2, pp. 123–129, 2008.
[17] A. W. Walker, J. Ince, S. H. Duncan, L. M. Webster, G. Holtrop, X. Ze, D. Brown, M. D. Stares, P. Scott, A. Bergerat, et al., "Dominant and diet-responsive groups of bacteria within the human colonic microbiota," *ISME J.*, vol. 5, no. 2, pp. 220–230, 2011.
[18] G. D. Wu, J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters, R. Knight, et al., "Linking long-term dietary patterns with gut microbial enterotypes," *Sci.*, vol. 334, no. 6052, pp. 105–108, 2011.
[19] M. R. Mason, P. M. Preshaw, H. N. Nagaraja, S. M. Dabdoub, A. Rahman, and P. S. Kumar, "The subgingival microbiome of clinically healthy current and never smokers," *ISME J.*, vol. 9, no. 1, pp. 268–272, 2015.
[20] S. D. Ehrlich, M. Consortium, et al., "Metahit: The european union project on metagenomics of the human intestinal tract," in *Metagenomics of the Human Body*. New York, NY, USA: Springer, 2011, pp. 307–316.
[21] H. Zhang and K. Ning, "The tara oceans project: New opportunities and greater challenges ahead," *Genomics Proteomics Bioinf.*, vol. 13, no. 5, 2015, Art. no. 275.
[22] W. Ma, L. Zhang, P. Zeng, C. Huang, J. Li, B. Geng, J. Yang, W. Kong, X. Zhou, and Q. Cui, "An analysis of human microbe–disease associations," *Briefings Bioinf.*, vol. 18, no. 1, pp. 85–97, 2016.
[23] Y. Cao, X. Zheng, F. Li, and X. Bo, "mmnet: An r package for metagenomics systems biology analysis," *BioMed Res. Int.*, vol. 2015, 2015, Art. no. 167249.
[24] E. D. Coelho, A. M. Santiago, J. P. Arrais, and J. L. Oliveira, "Computational methodology for predicting the landscape of the human–microbial interactome region level influence," *J. Bioinf. Comput. Biol.*, vol. 13, no. 5, 2015, Art. no. 1550023.
[25] A. May, B. W. Brandt, M. El-Kebir, G. W. Klau, E. Zaura, W. Crielaard, J. Heringa, and S. Abeln, "metamodules identifies key functional subnetworks in microbiome-related disease," *Bioinf.*, vol. 32, no. 11, pp. 1678–1685, 2015.
[26] X. Chen, Y.-A. Huang, Z.-H. You, G.-Y. Yan, and X.-S. Wang, "A novel approach based on katz measure to predict associations of human microbiota with non-infectious diseases," *Bioinf.*, vol. 33, no. 5, pp. 733–739, 2016.
[27] J.-Y. Shi, H. Huang, Y.-N. Zhang, J.-B. Cao, and S.-M. Yiu, "Bmcmda: a novel model for predicting human microbe-disease associations via binary matrix completion," *BMC Bioinf.*, vol. 19, no. 9, p. 169, 2018.
[28] F. Wang, Z.-A. Huang, X. Chen, Z. Zhu, Z. Wen, J. Zhao, and G.-Y. Yan, "Lrlshmda: Laplacian regularized least squares for human microbe–disease association prediction," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 7601.
[29] Y.-A. Huang, Z.-H. You, X. Chen, Z.-A. Huang, S. Zhang, and G.-Y. Yan, "Prediction of microbe–disease association from the integration of neighbor and graph with collaborative recommendation model," *J. Translational Med.*, vol. 15, no. 1, 2017, Art. no. 209.
[30] Y. Ma, X. Hu, T. He, and X. Jiang, "A robust symmetric nonnegative matrix factorization framework for clustering multiple heterogeneous microbiome data," *Preprints 2017*, 2017040105, doi: 10.20944/preprints201704.0105.v1.
[31] B. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, and A. Goldenberg, "Similarity network fusion for aggregating data types on a genomic scale," *Nature Methods*, vol. 11, no. 3, pp. 333–337, 2014.
[32] C. Yan, J. Wang, W. Lan, F.-X. Wu, and Y. Pan, "Sdtrls: Predicting drug-target interactions for complex diseases based on chemical substructures," *Complexity*, vol. 2017, 2017, Art. no. 2713280.
[33] W. Wang, S. Yang, and J. Li, "Drug target predictions based on heterogeneous graph inference," in *Pacific Symp. Biocomputing*, 2013, pp. 53–64.
[34] F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang, and Y. Tang, "Prediction of drug-target interactions and drug repositioning via network-based inference," *PLoS Comput. Biol.*, vol. 8, no. 5, 2012, Art. no. e1002503.
[35] H. J. Lowe and G. O. Barnett, "Understanding and using the medical subject headings (mesh) vocabulary to perform literature searches," *JAMA*, vol. 271, no. 14, pp. 1103–1108, 1994.
[36] X. Zhou, J. Menche, A.-L. Barabási, and A. Sharma, "Human symptoms–disease network," *Nature Commun.*, vol. 5, 2014, Art. no. 4212.
[37] D. L. Wheeler, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, et al., "Database resources of the national center for biotechnology information," *Nucleic Acids Res.*, vol. 36, no. suppl_1, pp. D13–D21, 2007.
[38] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug–target interaction," *Bioinf.*, vol. 27, no. 21, pp. 3036–3043, 2011.
[39] W. Lan, M. Li, K. Zhao, J. Liu, F.-X. Wu, Y. Pan, and J. Wang, "Ldap: A web server for lncrna-disease association prediction," *Bioinf.*, vol. 33, no. 3, pp. 458–460, 2016.
[40] C. Lu, M. Yang, F. Luo, F.-X. Wu, M. Li, Y. Pan, Y. Li, and J. Wang, "Prediction of lncrna-disease associations based on inductive matrix completion," *Bioinf.*, vol. 1, 2018, Art. no. 8.
[41] C. Yan, J. Wang, P. Ni, W. Lan, F. X. Wu, and Y. Pan, "DNRLMF-MDA:predicting microrna-disease associations based on similarities of micrornas and diseases," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 1, pp. 233–243, Jan./Feb. 2019, doi: 10.1109/TCBB.2017.2776101.
[42] H. Luo, J. Wang, M. Li, J. Luo, X. Peng, F.-X. Wu, and Y. Pan, "Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm," *Bioinf.*, vol. 32, no. 17, pp. 2664–2671, 2016.
[43] O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, and R. Sharan, "Associating genes and protein complexes with disease via network propagation," *PLoS Computat. Biol.*, vol. 6, no. 1, 2010, Art. no. e1000641.
[44] G. Salton, A. Wong, and C.-S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975.
[45] J. Luo and Q. Xiao, "A novel approach for predicting microrna-disease associations by unbalanced bi-random walk on heterogeneous network," *J. Biomed. Inform.*, vol. 66, pp. 194–203, 2017.
[46] W. Peng, W. Lan, Z. Yu, J. Wang, and Y. Pan, "A framework for integrating multiple biological networks to predict microrna-disease associations,," *IEEE transactions on nanobioscience*, vol. 16, no. 2, pp. 100–107, 2017.
[47] W. Peng, W. Lan, J. Zhong, J. Wang, and Y. Pan, "A novel method of predicting microrna-disease associations based on microrna, disease, gene and environment factor networks," *Methods*, vol. 124, pp. 69–77, 2017.
[48] W. Lan, J. Wang, M. Li, J. Liu, F.-X. Wu, and Y. Pan, "Predicting microrna-disease associations based on improved microrna and disease similarities," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 6, pp. 1774–1782, Nov./Dec. 2018, doi:10.1109/TCBB.2016.2586190.
[49] P. Concannon, H. A. Erlich, C. Julier, G. Morahan, J. Nerup, F. Pociot, J. A. Todd, S. S. Rich, et al., "Type 1 diabetes," *Diabetes*, vol. 54, no. 10, pp. 2995–3001, 2005.
[50] D. Daneman, "Type 1 diabetes," *Lancet*, vol. 367, no. 9513, pp. 847–858, 2006.

[51] M. B. Azad, T. Konya, H. Maughan, D. S. Guttman, C. J. Field, R. S. Chari, M. R. Sears, A. B. Becker, J. A. Scott, and A. L. Kozyrskyj, "Gut microbiota of healthy canadian infants: profiles by mode of delivery and infant diet at 4 months," *Can. Med. Assoc. J.*, vol. 185, no. 5, pp. 385–394, 2013.
[52] C. A. Heberling, P. S. Dhurjati, and M. Sasser, "Hypothesis for a systems connectivity model of autism spectrum disorder pathogenesis: Links to gut bacteria, oxidative stress, and intestinal permeability," *Med. Hypotheses*, vol. 80, no. 3, pp. 264–270, 2013.
[53] E. Lalla, S. Kaplan, S.-m. J. Chang, G. A. Roth, R. Celenti, K. Hinckley, E. Greenberg, and P. N. Papapanou, "Periodontal infection profiles in type 1 diabetes," *J. Clinical Periodontology*, vol. 33, no. 12, pp. 855–862, 2006.
[54] A. Giongo, K. A. Gano, D. B. Crabb, N. Mukherjee, L. L. Novelo, G. Casella, J. C. Drew, J. Ilonen, M. Knip, H. Hyöty, et al., "Toward defining the autoimmune microbiome for type 1 diabetes," *ISME J.*, vol. 5, no. 1, 2011, Art. no. 82.
[55] E. Soyucen, A. Gulcan, A. C. Aktuglu-Zeybek, H. Onal, E. Kiykim, and A. Aydin, "Differences in the gut microbiota of healthy children and those with type 1 diabetes," *Pediatrics Int.*, vol. 56, no. 3, pp. 336–343, 2014.
[56] D. Schuppan and N. H. Afdhal, "Liver cirrhosis," *Lancet*, vol. 371, no. 9615, pp. 838–851, 2008.
[57] J. S. Bajaj, L. R. Thacker, A. Fagan, M. B. White, E. A. Gavis, P. B. Hylemon, R. Brown, C. Acharya, D. M. Heuman, M. Fuchs, et al., "Gut microbial rna and dna analysis predicts hospitalizations in cirrhosis," *JCI Insight*, vol. 3, no. 5, 2018, Art. no. e98019.
[58] N. Rahman and R. Hitchcock, "Case report of paediatric oxalate urolithiasis and a review of enteric hyperoxaluria," *J. Pediatric Urology*, vol. 6, no. 2, pp. 112–116, 2010.
[59] N. Qin, F. Yang, A. Li, E. Prifti, Y. Chen, L. Shao, J. Guo, E. Le Chatelier, J. Yao, L. Wu, et al., "Alterations of the human gut microbiome in liver cirrhosis," *Nature*, vol. 513, no. 7516, 2014, Art. no. 59.
[60] R. Pund and D. Theegarten, "The importance of aeromonads as a human pathogen," *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, vol. 51, no. 5, pp. 569–576, 2008.
[61] S. M. Heckmann, J. G. Heckmann, J. J. Linke, W. Hohenberger, and A. Mombelli, "Implant therapy following liver transplantation: clinical and microbiological results after 10 years," *J. Periodontology*, vol. 75, no. 6, pp. 909–913, 2004.
[62] P. Sebo, K. Sakbani, P. Rohner, and G. Gavazzi, "Aeromonas bacteremia in an elderly immunocompetent patient," *Aging Clinical Exp. Res.*, vol. 18, no. 4, pp. 344–346, 2006.
[63] Y. Chen, F. Yang, H. Lu, B. Wang, Y. Chen, D. Lei, Y. Wang, B. Zhu, and L. Li, "Characterization of fecal microbial communities in patients with liver cirrhosis," *Hepatology*, vol. 54, no. 2, pp. 562–572, 2011.
[64] L. Cheng, J. Li, P. Ju, J. Peng, and Y. Wang, "Semfunsim: A new method for measuring disease similarity by integrating semantic and gene functional association," *PLoS One*, vol. 9, no. 6, 2014, Art. no. e99415.
[65] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, and J. Wang, "Computational drug repositioning using low-rank matrix approximation and randomized algorithms," *Bioinf.*, vol. 34, no. 11, pp. 1904–1912, 2018.

Cheng Yan is working toward the PhD degree in the School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include bioinformatics and machine learning.

Guihua Duan received the PhD degree in computer science and technology from Central South University, in 2010. She is currently an associate professor with the School of Computer Science and engineering, Central South University. Her research interests include network security, bioinformatics, and computer network.

Fang-Xiang Wu (M'06-SM'11) received the BSc degree and the MSc degree in applied mathematics, both from the Dalian University of Technology, Dalian, China, in 1990 and 1993, respectively, the first PhD degree in control theory and its applications from Northwestern Polytechnical University, Xi'an, China, in 1998, and the second PhD degree in biomedical engineering from University of Saskatchewan (U of S), Saskatoon, Canada, in 2004. During 2004-2005, he worked as a postdoctoral fellow with the Laval University Medical Research Center (CHUL), Quebec City, Canada. He is currently a professor of the Division of Biomedical Engineering and the Department of Mechanical Engineering, U of S. His current research interests include computational and systems biology, genomic and proteomic data analysis, biological system identification and parameter estimation, and applications of control theory to biological systems. He has published more than 300 technical papers in refereed journals and conference proceedings. He is serving as on the editorial board member of six international journals, the guest editor of various international journals, and as the program committee chair or member of many international conferences. He has also reviewed papers for many international journals. He is a senior member of the IEEE.

Yi Pan received the BEng and MEng degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and the PhD degree in computer science from the University of Pittsburgh, in 1991. He is a regents' professor of computer science and an interim associate dean and chair of biology with Georgia State University. He joined Georgia State University, in 2000 and was promoted to full professor in 2004, named a distinguished university professor in 2013, and designated a regents' professor (the highest recognition given to a faculty member by the University System of Georgia) in 2015. He served as the chair of the Computer Science Department from 2005-2013. He is also a visiting Changjiang chair professor with Central South University, China. His profile has been featured as a distinguished alumnus in both the *Tsinghua Alumni Newsletter* and *University of Pittsburgh CS Alumni Newsletter*. His research interests include parallel and cloud computing, wireless networks, and bioinformatics. He has published more than 330 papers including over 180 SCI journal papers and 60 IEEE/ACM Transactions papers. In addition, he has edited/authored 40 books. His work has been cited more than 6,500 times. He has served as an editor-in-chief or editorial board member for 15 journals including seven IEEE Transactions. He is the recipient of many awards including IEEE transactions Best Paper Award, four other international conference or journal Best Paper Awards, four IBM Faculty Awards, two JSPS Senior Invitation Fellowships, IEEE BIBE Outstanding Achievement Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at over 50 international conferences around the world.

Jianxin Wang received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the chair of and a professor with the Department of Computer Science, Central South University, Changsha, Hunan, P.R. China. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics, and computer network. He is a senior member of the IEEE.