



A Novel Approach Based on Bipartite Network Recommendation and KATZ Model to Predict Potential Micro-Disease Associations

Shiru Li¹, Minzhu Xie^{1*} and Xinqiu Liu²

¹ College of Information Science and Engineering, Hunan Normal University, Changsha, China, ² Hunan Vocational College of Engineering, Changsha, China

OPEN ACCESS

Edited by:

Quan Zou,
University of Electronic Science and
Technology of China, China

Reviewed by:

Xiangxiang Zeng,
Xiamen University, China
Wen Zhang,
Huazhong Agricultural University,
China

*Correspondence:

Minzhu Xie
xieminzhu@hunnu.edu.cn

Specialty section:

This article was submitted to
Bioinformatics and
Computational Biology,
a section of the journal
Frontiers in Genetics

Received: 19 July 2019

Accepted: 21 October 2019

Published: 15 November 2019

Citation:

Li S, Xie M and Liu X (2019) A
Novel Approach Based on Bipartite
Network Recommendation and KATZ
Model to Predict Potential Micro-
Disease Associations.
Front. Genet. 10:1147.
doi: 10.3389/fgene.2019.01147

Accumulating evidence indicates that the microbes colonizing human bodies have crucial effects on human health and the discovery of disease-related microbes will promote the discovery of biomarkers and drugs for the prevention, diagnosis, treatment, and prognosis of diseases. However clinical experiments of disease-microbe associations are time-consuming, laborious and expensive, and there are few methods for predicting potential microbe-disease association. Therefore, developing effective computational models utilizing the accumulated public data of clinically validated microbe-disease associations to identify novel disease-microbe associations is of practical importance. We propose a novel method based on the KATZ model and Bipartite Network Recommendation Algorithm (KATZBNRA) to discover potential associations between microbes and diseases. We calculate the Gaussian interaction profile kernel similarity of diseases and microbes based on validated disease-microbe associations. Then, we construct a bipartite graph and execute a bipartite network recommendation algorithm. Finally, we integrate the disease similarity, microbe similarity and bipartite network recommendation score to obtain the final score, which is used to infer whether there are some novel disease-microbe interactions. To evaluate the predictive power of KATZBNRA, we tested it with the walk length 2 using global leave-one-out cross validation (LOOV), two-fold and five-fold cross validations, with AUCs of 0.9098, 0.8463 and 0.8969, respectively. The test results also show that KATZBNRA is more accurate than two recent similar methods KATZHMDA and BNPMDA.

Keywords: microbe, disease, KATZ model, bipartite network recommendation, Gaussian interaction profile kernel similarity

INTRODUCTION

A microbe is a microscopic organism, including bacteria, eukaryotes, archaea, and viruses (Wu et al., 2018). Various types of microbes live on or in different parts of a human body such as the skin, mouth, hair, stomach, and gastrointestinal tract. An adult human body contains a large number of bacterial cells, which is estimated to reach 10^{14} and much more than the total number of human cells, with more than 5 million microbe genes, outnumbering the human genes by more than 100 fold (Sommer and Backhed, 2013). Most microbes are harmless and some are beneficial to humans

(Grice and Segre, 2011). Recently, accumulated experimental evidence shows that microbes have an important impact on human health, nutrient absorption, immune response, cancer control, and the prevention of pathogen colonization (Wu et al., 2018). For example, the gut microbiota could significantly contribute nutrition absorption by producing indispensable vitamins and decomposing indigestible polysaccharides, and it also has an important impact on the mucus layer, the balance of antimicrobial peptides, and immunoglobulin A, and the differentiation and activation of some lymphocyte populations (Sommer and Backhed, 2013). Therefore, the gut microbiota is thought to be an extra 'organ' of humans (Gill et al., 2006). But some microbes may contribute to disease, such as psoriasis and inflammatory bowel disease (IBD). There have been reports that psoriasis occurs after strep throat and could worsen due to the colonization of *Candida albicans*, *Malassezia*, and *Staphylococcus aureus* on the skin or in the gut (Fry and Baker, 2007). Aroniadis et al. (Aroniadis and Brandt, 2013) indicated that the biodiversity of bacteria, such as Bacteroidetes and Firmicutes, colonizing in individuals affected by IBD has been found to be reduced by 30 to 50%. Wang et al. (Wang and Jia, 2016) showed that the gut microbiota's dysbiosis might be a key environmental risk factor of many human diseases, though it's difficult to reveal the true causality.

To explore the relationship between microbes and their human hosts, scientists from many countries collaborated and launched the Human Microbiome Project (HMP) (Human Microbiome Project, 2012a). Recently, high-throughput sequencing techniques and corresponding software packages have been developed rapidly, and a growing number of research analyses have been carried out on the microbiome, such as whole-genome shotgun (WGS), 16S, and the taxonomic profiling (Human Microbiome Project, 2012b), and have demonstrated significant associations between microbes and complex human diseases such as rheumatoid arthritis, colorectal cancer, obesity, and type 2 diabetes (Wang and Jia, 2016). However, these studies involve time-consuming and expensive biological experiments. Therefore, it is necessary to utilize the known information to predict the unknown microbe-disease interactions. Identifying microbe-disease interactions could promote discovering biomarkers and drugs for the prevention, diagnosis, treatment, and prognosis of diseases. Now, more and more computer algorithms (Chen and Zhang, 2013; Yang et al., 2014; Zhang et al., 2017; Zeng et al., 2018; Zhang et al., 2018a; Zhang et al., 2018b; Zhang et al., 2018c; Zhang et al., 2018d; Zeng et al., 2019) have been proposed for interaction prediction of miRNA-disease, lncRNA-disease, and drug-drug, and it is feasible to apply these methods to the microbe-disease association prediction field.

Recently, Ma et al. (2017) collected microbe-disease association data from previous published studies and constructed the Human Microbe-Disease Association Database (HMDAD). Based on the data from HMDAD, some microbe-disease association prediction methods have been proposed. Chen et al. (2017) used a KATZ measure to predict human microbe-disease association, and proposed an algorithm named KATZHMDA. KATZHMDA can predict new microbe-disease associations at a

large scale. Bao et al. (2017) used network consistency projection and introduced an algorithm NCPHMD to predict human microbe-disease association. NCPHMD deals with unknown diseases or microbes that are not present in the disease-microbe databases. He et al. (He et al., 2018) presented an algorithm GRNMFHMDA. GRNMFHMDA assigns likelihood scores to unknown microbe-disease pairs by calculating weighted K nearest neighbor profiles of microbes and diseases, and then adapts the standard non-negative matrix factorization by integrating graph Laplacian and Tikhonov (L2) regularization to obtain a microbe-disease association prediction score matrix. Zou et al. (2017) designed an approach BiRWHMDA. BiRWHMDA constructs a heterogeneous network by connecting the microbe similarity network and the disease similarity network based on known microbe-disease associations, and then uses a bi-random walk to predict microbe-disease association.

In the paper, we propose a novel approach to predict potential micro-disease association based on the KATZ measure and bipartite network recommendation algorithm (KATZBNRA), which is an improvement on KATZHMDA (Chen et al., 2017). Similar to KATZHMDA, KATZBNRA uses the KATZ measure and the similarity of diseases and microbes according to the Gaussian interaction profile kernel to predict novel microbe-disease associations based on the known microbe-disease associations. Furthermore, in order to improve the predicting accuracy, KATZBNRA uses a bipartite network recommendation algorithm.

MATERIALS AND METHODS

Known Disease-Microbe Associations

HMDAD (Human Microbe-Disease Association Database, <http://www.cuilab.cn/hmdad>) collected the curated human microbe-disease association data from microbiota studies where the microbes were determined by 16s RNA sequencing on the genus level (Ma et al., 2017). HMDAD provides public access to the data, and our known microbe-disease association data were downloaded from HMDAD. The data contains 450 distinct confirmed associations between 39 diseases and 292 microbes and is coded in an adjacency matrix $A \in R^{n_d \times n_m}$, where n_d (or n_m) is the number of diseases (or microbes). If there has been an experiment confirming that microbe m_j relates to disease d_i , $A(i,j)$ is set to 1, otherwise $A(i,j)$ is set to 0.

Disease Gaussian Interaction Profile Kernel Similarity

According to (Chen et al., 2017), there is a generally accepted assumption that similar diseases show an interaction tendency to similar microbes. Similar to (Chen and Yan, 2013) and (Chen et al., 2017), we compute the disease network topologic similarity based on the Gaussian interaction profile kernel. For a disease-microbe association adjacent matrix A , the binary element $A(i,j)$ at row i and column j encodes whether there is a confirmed association between disease $d(i)$ and microbe $m(j)$. The i th row of A is denoted by $IP(d(i))$. $IP(d(i))$ can be regarded a binary vector and is called the interaction profile of $d(i)$ since it provides

the association information of disease $d(i)$ with all microbes. For two diseases, their similarity $KD(d_p, d_i)$, based on the Gaussian interaction profile kernel, is calculated from their interaction profiles according to the following equations.

$$KD(d_i, d_j) = \exp(-\gamma_d \|IP(d_i) - IP(d_j)\|^2) \tag{1}$$

$$\gamma_d = \frac{\gamma'_d}{\left(\frac{1}{n_d} \sum_{k=1}^{n_d} \|IP(d_k)\|^2\right)} \tag{2}$$

$KD(d_p, d_i)$ is adjusted by the norm kernel bandwidth γ_d , which is controlled by the bandwidth parameter γ'_d . It is obvious that $KD(d_p, d_i) = 1$ and $0 < KD(d_p, d_i) \leq 1$. According to (Vanunu et al., 2010), KD values in $(0, 0.3]$ may be not informative, while KD values in $[0.6, 1]$ may show significant similarity. Therefore, a logistic function transformation from $KD(x, y)$ to $KD'(x, y)$ in Equation (3) is utilized in order to measure the similarity of diseases x and y more appropriately.

$$KD'(d_i, d_j) = \frac{1}{1 + e^{c \cdot KD(d_i, d_j) + d}} \tag{3}$$

The parameters γ'_d and c could be set with cross-validation, but to simplify the calculation, we set $\gamma'_d = 1$ as in van Laarhoven et al., 2011, $c = -15$ as in (Vanunu et al., 2010). According to (Vanunu et al., 2010), we set $d = \log(9999)$ such that $KD'(d_p, d_j) = 0.0001$ when $KD(d_p, d_j) = 0$.

Microbe Gaussian Interaction Profile Kernel Similarity

As mentioned before, similar diseases show an association tendency with similar microbes. To measure the similarity of microbes, we also used the Gaussian interaction profile kernel as before. It could be calculated in a similar way as follows.

$$KM(m_i, m_j) = \exp(-\gamma_m \|IP(m_i) - IP(m_j)\|^2) \tag{4}$$

$$\gamma_m = \frac{\gamma'_m}{\left(\frac{1}{n_m} \sum_{k=1}^{n_m} \|IP(m_k)\|^2\right)} \tag{5}$$

where γ'_m is also set to 1, and $IP(m_i)$ is the i th column of matrix A . Similarly, $KM'(m_p, m_j)$ could be calculated as Equation (3). It should be noted that in each cross-validation experiment, the similarities of diseases and microbes will be recalculated (Sun et al., 2018).

Bipartite Network Recommendation

The bipartite network recommendation is a two-step resource allocation process (Chen et al., 2018b), which is based on a bipartite graph $G(D, M, E)$, where D represents disease nodes,

M microbe nodes, E the edges corresponding to the known microbe-disease associations. Let $f_{0,i}(m_j)$ denote the initial resource allocated to a microbe node m_j when considering disease d_p , $k(m_j)$ be the number of adjacent disease nodes of microbe m_j , and let $k(d_i)$ be the number of adjacent microbe nodes of disease d_i in graph G .

When focusing on disease d_p , each disease d_i related microbe node is initially allocated with a resource value of 1, i.e. if there is an edge between the disease node d_i and a microbe node m_j in G , allocate an initial resource of 1 to m_j . The first step of the bipartite network recommendation is to transfer the resource from microbe nodes to disease nodes according to Equation (6), and the second step is to transfer the resource of the disease nodes back to microbe nodes according to Equation (8).

$$f_{1,i}(d_i) = \sum_{j=1}^{n_m} \frac{a_{ij} f_{0,i}(m_j)}{k(m_j)} \tag{6}$$

where a_{ij} is an element of matrix A , i.e. $a_{ij} = A(i, j)$ and

$$a_{ij} = \begin{cases} 1, & \text{if disease } d_i \text{ is related to microbe } m_j \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

In fact, $f_{0,i}(m_j)$ is also equal to $A(i, j)$.

$$f_{2,i}(m_j) = \sum_{l=1}^{n_d} \frac{a_{lj} f_{1,i}(d_l)}{k(d_l)} \tag{8}$$

Equations (6) and (8) are integrated into Equation (9).

$$f_{2,i}(m_i) = \sum_{j=1}^{n_m} w_{ij} f_{0,i}(m_j) \tag{9}$$

$$w_{ij} = \frac{1}{k(m_j)} \sum_{q=1}^{n_d} \frac{a_{iq} a_{qj}}{k(d_q)} \tag{10}$$

Please see an example of the process of the bipartite network recommendation focusing on disease d_1 in **Figure 1**. After the process, we obtain the recommendation scores $(1, 0, 1/4, 3/4, 0)$ of the microbes for disease d_1 , which suggests that besides m_1 and m_4 , m_3 may also be related to the disease.

The matrix form of Equation (9) is as follows.

$$B = W \times A^T \tag{11}$$

where $W = \{w_{ij}\}_{n_m \times n_m}$, and B is a matrix with n_m rows and n_d columns. The i th column of B is the recommend scores of bipartite network recommendation regarding disease d_i

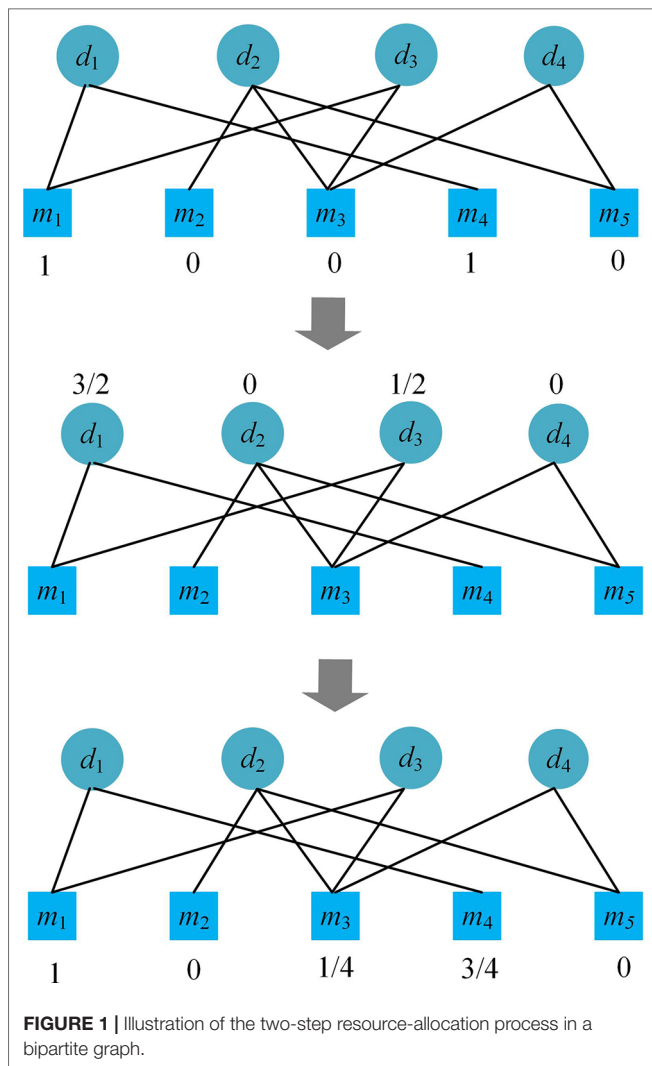


FIGURE 1 | Illustration of the two-step resource-allocation process in a bipartite graph.

KATZBNRA

KATZBNRA uses the KATZ model to compute the associations between diseases and microbes and is illustrated in **Figure 2**. As a network-based computation method, the KATZ model (Chen, 2015) had been used in the problem of link prediction in the heterogenous network to calculate the similarity of nodes. There are two factors that have been regarded as effective similarity metrics in the KATZ model, the walk steps (length, i.e. the number of edges of the walk) and the number of walks from one node to another. We use the KATZ model to calculate similarities between the nodes of the microbe and disease by counting the number of walks between them. Here $A^l(i, j)$, the element of the l -th power of A , is the number of l -length walks between disease node d_i and microbe node m_j . Due to the limited data from HMDAD, matrix A is sparse. In order to use more information, we integrated the matrices KM , KD , B into a matrix B^* as Equation (12) and replace A by B^*

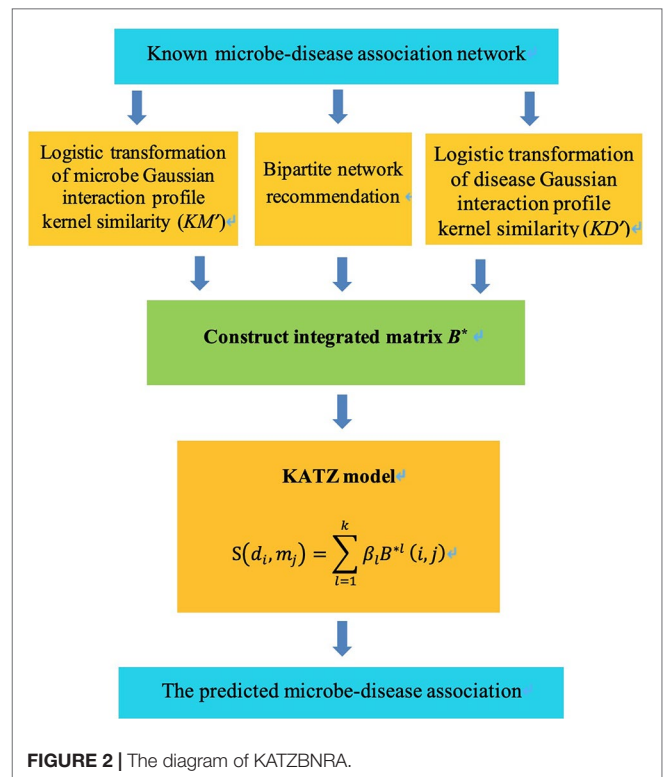


FIGURE 2 | The diagram of KATZBNRA.

in the KATZ model to calculate similarities between microbes and diseases.

$$B^* = \begin{bmatrix} KD' & B \\ B^T & KM' \end{bmatrix} \tag{12}$$

Since walks between nodes of microbe and disease with different lengths have different contributions to similarities of node pairs, in order to dampen longer walks' contribution, we introduced a parameter β_l which is no smaller than 0, and if $l_1 > l_2$, then $\beta_{l_1} < \beta_{l_2}$. The potential association between diseases d_i and microbe m_j can be calculated as follows.

$$S(d_i, m_j) = \sum_{l=1}^k \beta_l B^{*l}(i, j) \tag{13}$$

If $k \rightarrow \infty$, replace β_l with β^l ($0 < \beta < 1$) (Qu et al., 2018) and the matrix form of Equation (13) is as follows.

$$S = \sum_{l \geq 1} \beta^l B^{*l} = (I - \beta B^*)^{-1} - I \tag{14}$$

S is a matrix of size $(n_d + n_m) \times (n_d + n_m)$, and could be partitioned into four sub-matrices as shown in Equation (12).

$$S = \begin{bmatrix} S_{1,1} & S_{1,2} \\ S_{2,1} & S_{2,2} \end{bmatrix} \quad (15)$$

where the rows of $S_{1,1}$ and $S_{1,2}$ are n_d , the rows of $S_{2,1}$ and $S_{2,2}$ are n_m , the columns of $S_{1,1}$ and $S_{2,1}$ are n_d , and the columns of $S_{1,2}$ and $S_{2,2}$ are n_m . The element $S_{1,2}(i, j)$ of $S_{1,2}$ provides the possibility that an association between disease d_i and the microbe m_j exists, and our prediction result can be obtained from $S_{1,2}$.

Considering that the walks of long lengths may be meaningless, we limit k in Equation (13) to be 2, 3 and 4, and the expression can be as follows.

$$S_{k=2} = \beta \cdot B + \beta^2 \cdot (KM' \cdot B + B \cdot KD') \quad (16)$$

$$S_{k=3} = S_{k=2} + \beta^3 \cdot (B \cdot B^T \cdot B + KM'^2 \cdot B + KM' \cdot B \cdot KD' + B \cdot KD'^2) \quad (17)$$

$$S_{k=4} = S_{k=3} + \beta^4 \cdot (KM'^3 \cdot B + B \cdot B^T \cdot KM' \cdot B + KM' \cdot B \cdot B^T \cdot B + B \cdot KD' \cdot B^T \cdot B) + \beta^4 \cdot (B \cdot B^T \cdot B \cdot KD' + KM'^2 \cdot B \cdot KD' + KM' \cdot B \cdot KD'^2 + B \cdot KD'^3) \quad (18)$$

RESULTS

Performance Evaluation

The test dataset of microbe-disease association was downloaded from HMDAD. We used LOOCV (leave-one-out cross validation), two-fold cross validation and five-fold cross validation to test the prediction performance of KATZBNRA on the HMDAD data.

In LOOCV, each known microbe-disease association takes turns to be picked out as the testing case and the other known associations are regarded as training data. We then obtained the prediction score of the test case output by KATZBNRA and ranked of the test case in the sorted list of all predicted microbe-disease associations in descending order of their scores. We used different thresholds to determine the correct predictions and wrong predictions and calculated corresponding FPR (false positive rate) and TPR (true positive rate) according to Equation (19). Finally, the results were presented in the ROC (receiver operating characteristics) curve plot of TPR against FPR.

$$TPR = \frac{TP}{FN + TP}, \quad FPR = \frac{FP}{TN + FP} \quad (19)$$

where FN is the number of false negative predictions (i.e. the cases whose prediction scores below the threshold), and TP is the number of true positive predictions (i.e. the cases whose prediction scores are not smaller than the threshold). FP is the number of the predicted associations that are not in the HMDAD

dataset with scores not smaller than the threshold, and TN is the number of predicted associations that are not in the HMDAD dataset with scores smaller than the threshold. The area under a ROC curve is called AUC, and AUC is generally utilized to compare the power of predictive models. AUC of 0.5 indicates an entirely random prediction while AUC = 1 means a completely correct prediction.

In order to further test the prediction power of KATZBNRA, we also adopted 5-fold cross validation and 2-fold cross validation besides LOOCV. 5-fold (or 2-fold) cross validation randomly divides the microbe-disease associations equally into five (or two) parts and one of the five (or two) parts is reserved as the verification data while the remaining is used as training data. Considering the potential random sampling bias, we repeated each LOOCV, 2-fold and 5-fold cross validation test 100 times, and all ROC curves and AUCs are the average results of the 100 repeated tests. Meanwhile, we compared KATZBNRA with several state-of-the-art predictive methods using these validations.

For our method KATZBNRA, the walk length k plays a critical role. To test the effect of k , we changed the value of k , and carried out a series of LOOCV experiments. As shown in **Figure 3**, when k is set to 2, 3 and 4, the AUCs of each walk lengths are 0.9098, 0.8968, and 0.8827, respectively. Obviously, when parameter $k = 2$, KATZBNRA achieved the best prediction performance and walks of longer lengths may make the association prediction worse. Therefore, in the following experiments, we set $k = 2$. KATZBNRA has two more parameters, γ' and β . The test in a previous work (Chen et al., 2016) showed AUC tended to decrease when γ' was increased from 1.0 to 1.5, 2.0 and 2.5, and β was increased from 0.01 to 0.05 and 0.1. We also evaluated the AUC of KATZBNRA with different values of parameter γ' and c in Equation (2) and Equation (3), and the test results are shown in **Tables 1** and **2**, showing similar results as Chen et al., 2016. Therefore, we set $\gamma'=1.0$ and $\beta=0.01$.

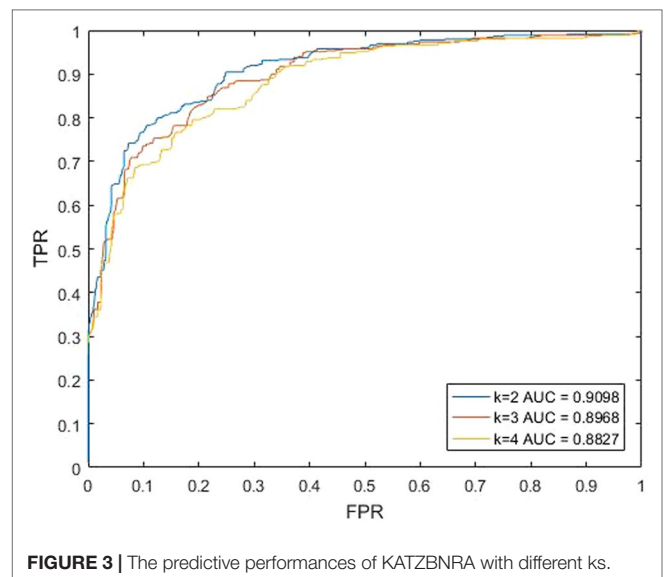


TABLE 1 | The AUC of KATZBNRA with γ' set different values.

γ'	AUC
1	0.9098
1.5	0.9083
2	0.9033

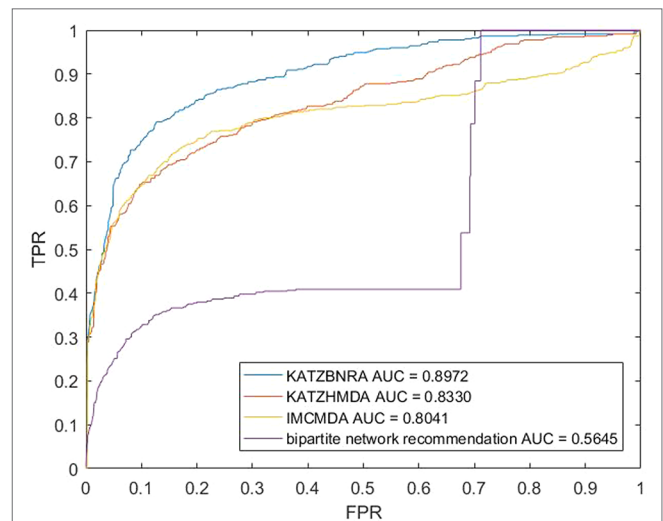
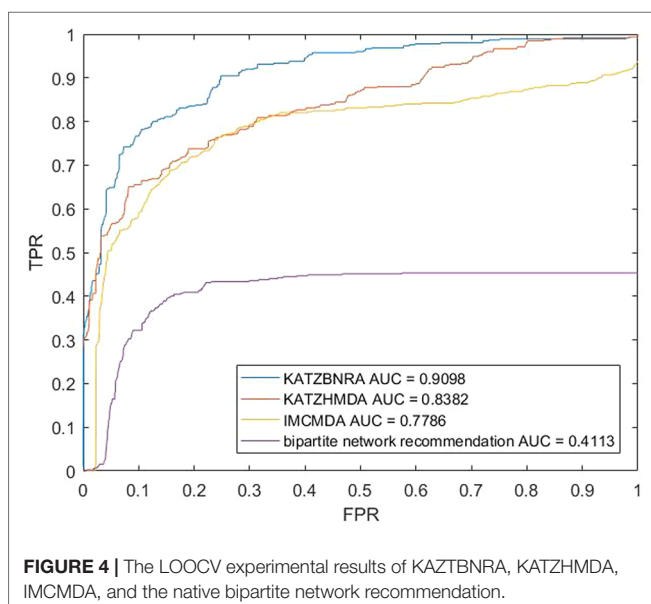
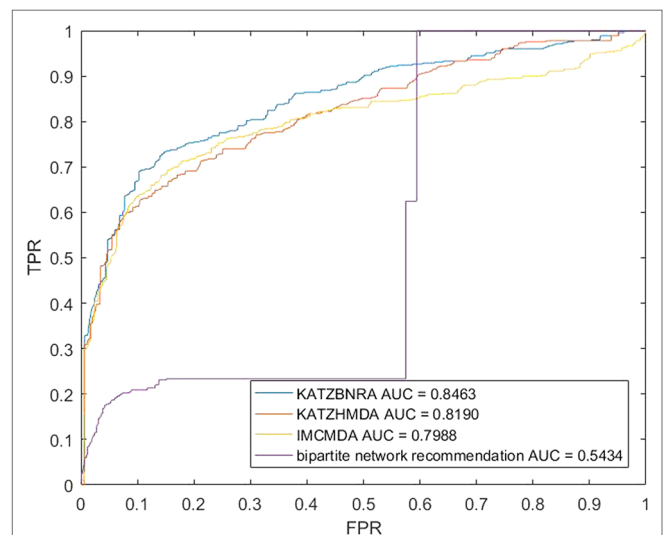
TABLE 2 | The AUC of KATZBNRA with c set different values.

c	AUC
-15	0.9098
-10	0.9038
-5	0.8935

We compared KATZBNRA with another three prediction methods, the native bipartite network recommendation (BNR) (Zhou et al., 2007), KATZHMDA (Chen et al., 2017), and IMCMA (Chen et al., 2018a) using LOOCV, 5-fold cross validation and 2-fold cross validation. The global LOOCV showed that the AUCs of KATZBNRA, KATZHMDA, IMCMA and BNR were 0.9098, 0.8382, 0.7786, and 0.4113, respectively, as shown in **Figure 4–6** show the 5-fold cross validation experimental results and the 2-fold cross validation experimental results, respectively. In 5-fold cross validation KATZBNRA, KATZHMDA, IMCMA, and BNR obtained AUCs of 0.8972, 0.8330, 0.8041, and 0.5645, respectively, and in 2-fold cross validation, their AUCs were 0.8463, 0.8190, 0.7988 and 0.5434, respectively. In all the above experiments, the curves of KATZBNRA are above those of the other methods, which means that among the four methods, KATZBNRA achieved the best prediction performance.

Case Studies

We studied asthma and inflammatory bowel disease (IBD) of microbe-related diseases of human beings based on recent

**FIGURE 5** | The 5-fold cross validation experimental results of KATZBNRA, KATZHMDA, IMCMA, and the native bipartite network recommendation.**FIGURE 6** | The 2-fold cross validation experimental results of KATZBNRA, KATZHMDA, IMCMA, and the native bipartite network recommendation.

published clinical and biological reports to further evaluate the ability of our method. The predicted disease-microbe associations which are contained in the HMDAD dataset are sorted according to their prediction scores in descending order. For asthma and IBD, we observed the microbes in the top 10 associations of the lists. This guarantees absolute independence between the verification candidate and the known association for model training.

As a common chronic lung inflammatory disease, asthma causes difficulty in breathing (Martinez, 2007). It is believed that asthma is caused by the environment and a combination of genes. For severe asthma, one of the leading causes is a microbe (Huang et al., 2011). All of top predicted 10 candidate microbes of KATZBNRA (**Table 3**) have been verified by recent studies.

TABLE 3 | The Asthma-related microbe prediction of KATZBNRA. All of top 10 microbes were confirmed by recent studies.

Rank	Microbe	Evidence
1	Firmicutes	PMID: 23265859(Marri et al., 2013)
2	Actinobacteria	PMID: 23265859(Marri et al., 2013)
3	Clostridium coccoides	PMID:21477358(Vael et al., 2011)
4	Streptococcus	PMID: 17950502(Preston et al., 2007)
5	Lactobacillus	PMID: 20592920(Yu et al., 2010)
6	Lachnospiraceae	PMID:17433177(Rados et al., 2007)
7	Pseudomonas	PMID:13268970(Fein, 1955)
8	Burkholderia	PMID:24451910(Beigelman et al., 2014)
9	Fusobacterium	Dang et al., 2013(Dang et al., 2013)
10	Propionibacterium	PMID:27433177(Jung et al., 2016)

TABLE 4 | Top 10 potential IBD-related microbes predicted by KATZBNRA

Rank	Microbe	Evidence
1	Clostridium coccoides	PMID:19235886(Sokol et al., 2009)
2	Firmicutes	PMID:25307765(Walters et al., 2014)
3	Bacteroidetes	PMID:25307765(Walters et al., 2014)
4	Staphylococcus	PMID:28174737(Pedamallu et al., 2016)
5	Prevotella	PMID:25307765(Walters et al., 2014)
6	Streptococcus	PMID:23679203(Kojima et al., 2014)
7	Propionibacterium	unconfirmed
8	Propionibacterium acnes	unconfirmed
9	Bacteroidaceae	PMID:17897884(Takaishi et al., 2008)
10	Haemophilus	PMID:24013298(Said et al., 2014)

As a typical chronic GI (gastrointestinal) tract inflammatory bowel disease, IBD includes ulcerative colitis and Crohn's disease (Lomax et al., 2006). We listed the top 10 IBD-related candidate microbes predicted by KATZBNRA in **Table 4**, among which eight microbes have been previously validated.

DISCUSSION

Based on the bipartite network recommendation and the KATZ model, the paper introduced a novel disease-microbe association prediction method called KATZBNRA. KATZBNRA uses the

REFERENCES

- Aroniadis, O. C., and Brandt, L. J. (2013). Fecal microbiota transplantation: past, present and future. *Curr. Opin. Gastroenterol.* 29 (1), 79–84. doi: 10.1097/MOG.0b013e32835a4b3e
- Bao, W., Jiang, Z., and Huang, D. S. (2017). Novel human microbe-disease association prediction using network consistency projection. *BMC Bioinf.* 18 (Suppl 16), 543. doi: 10.1186/s12859-017-1968-2
- Beigelman, A., Weinstock, G. M., and Bacharier, L. B. (2014). The relationships between environmental bacterial exposure, airway bacterial colonization, and asthma. *Curr. Opin. Allergy Clin. Immunol.* 14 (2), 137–142. doi: 10.1097/ACI.0000000000000036
- Chen, H., and Zhang, Z. (2013). Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med. Genomics* 6, 12. doi: 10.1186/1755-8794-6-12

Gaussian interaction profile kernel to calculate the similarity of diseases and microbes in the bipartite network containing the known microbe-disease associations from the HMDAD database, and the bipartite network recommendation score on the KATZ model enables KATZBNRA to predict potential disease-microbe associations with high accuracy. The experimental results of LOOCV, 5-fold cross validation, 2-fold cross validation and the IBD and asthma case studies have demonstrated the excellent and reliable prediction ability of KATZBNRA. With regard to similar prediction problems such as predicting lncRNA-disease, drug-target, gene-disease, miRNA-disease, and other biological associations, this model can be applied with small modifications.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.cuilab.cn/hmdad>.

AUTHOR CONTRIBUTIONS

SL and MX conceived the study and the conceptual design of the work. SL and XL collected the data and implemented the KATZBNRA algorithm. SL tested the algorithms and drafted the manuscript. MX and XL polished the manuscript. All authors have read and approved the manuscript.

FUNDING

This work is supported by the National Natural Science Foundation of China (Grant No. 61772197).

ACKNOWLEDGMENTS

We sincerely thank all reviewers for their valuable comments that we have used to improve the quality of our manuscript. The paper is an extended version of the abstract presented orally at CBC 2019 and has been recommended by CBC 2019.

- Chen, X. (2015). KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Sci. Rep.* 5, 16840. doi: 10.1038/srep16840
- Chen, X., Huang, Y. A., Wang, X. S., You, Z. H., and Chan, K. C. (2016). FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget* 7 (29), 45948–45958. doi: 10.18632/oncotarget.10008
- Chen, X., Huang, Y. A., You, Z. H., Yan, G. Y., and Wang, X. S. (2017). A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* 33 (5), 733–739. doi: 10.1093/bioinformatics/btw715
- Chen, X., Wang, L., Qu, J., Guan, N. N., and Li, J. Q. (2018a). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 34 (24), 4256–4265. doi: 10.1093/bioinformatics/bty503
- Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., and Liu, H. (2018b). BNPMDA: Bipartite Network Projection for MiRNA-Disease Association prediction. *Bioinformatics* 34 (18), 3178–3186. doi: 10.1093/bioinformatics/bty333

- Chen, X., and Yan, G. Y. (2013). Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* 29 (20), 2617–2624. doi: 10.1093/bioinformatics/btt426
- Dang, H. T., Park, H. K., Shin, J. W., Park, S.-G., and Kim, W. (2013). Analysis of oropharyngeal microbiota between the patients with bronchial asthma and the non-asthmatic persons. *J. Bacteriology Virol.* 43 (4), 270–278. doi: 10.4167/jbv.2013.43.4.270
- Fein, B. T. (1955). Bronchial asthma caused by *Pseudomonas aeruginosa* diagnosed by bronchoscopic examination. *Ann. Allergy* 13 (6), 639–641.
- Fry, L., and Baker, B. S. (2007). Triggering psoriasis: the role of infections and medications. *Clin. Dermatol.* 25 (6), 606–615. doi: 10.1016/j.clindermatol.2007.08.015
- Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312 (5778), 1355–1359. doi: 10.1126/science.1124234
- Grice, E. A., and Segre, J. A. (2011). The skin microbiome. *Nat. Rev. Microbiol* 9 (4), 244–253. doi: 10.1038/nrmicro2537
- He, B. S., Peng, L. H., and Li, Z. (2018). Human microbe-disease association prediction with graph regularized non-negative matrix factorization. *Front. Microbiol* 9, 2560. doi: 10.3389/fmicb.2018.02560
- Huang, Y. J., Nelson, C. E., Brodie, E. L., Desantis, T. Z., Baek, M. S., Liu, J., et al. (2011). Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J. Allergy Clin. Immunol.* 127372–381 (2), e371–e373. doi: 10.1016/j.jaci.2010.10.048
- Human Microbiome Project, C. (2012a). A framework for human microbiome research. *Nature* 486 (7402), 215–221. doi: 10.1038/nature11209
- Human Microbiome Project, C. (2012b). Structure, function and diversity of the healthy human microbiome. *Nature* 486 (7402), 207–214. doi: 10.1038/nature11234
- Jung, J. W., Choi, J. C., Shin, J. W., Kim, J. Y., Park, I. W., Choi, B. W., et al. (2016). Lung microbiome analysis in steroid-naïve asthma patients by using whole sputum. *Tuberc Respir. Dis. (Seoul)* 79 (3), 165–178. doi: 10.4046/trd.2016.79.3.165
- Kojima, A., Nomura, R., Naka, S., Okawa, R., Ooshima, T., and Nakano, K. (2014). Aggravation of inflammatory bowel diseases by oral streptococci. *Dis.* 20 (4), 359–366. doi: 10.1111/odi.12125
- Lomax, A. E., Linden, D. R., Mawe, G. M., and Sharkey, K. A. (2006). Effects of gastrointestinal inflammation on enteroendocrine cells and enteric neural reflex circuits. *Auton Neurosci.* 126–127. doi: 10.1016/j.autneu.2006.02.015
- Ma, W., Zhang, L., Zeng, P., Huang, C., Li, J., Geng, B., et al. (2017). An analysis of human microbe-disease associations. *Brief Bioinform.* 18 (1), 85–97. doi: 10.1093/bib/bbw005
- Marri, P. R., Stern, D. A., Wright, A. L., Billheimer, D., and Martinez, F. D. (2013). Asthma-associated differences in microbial composition of induced sputum. *J. Allergy Clin. Immunol.* 131346–352 (2), e341–e343. doi: 10.1016/j.jaci.2012.11.013
- Martinez, F. D. (2007). Genes, environments, development and asthma: a reappraisal. *Eur. Respir. J.* 29 (1), 179–184. doi: 10.1183/09031936.00087906
- Pedamallu, C. S., Bhatt, A. S., Bullman, S., Fowler, S., Freeman, S. S., Durand, J., et al. (2016). Metagenomic characterization of microbial communities in situ within the deeper layers of the ileum in crohn's disease. *Cell Mol. Gastroenterol. Hepatol* 2563–566 (5), e565. doi: 10.1016/j.jcmgh.2016.05.011
- Preston, J. A., Essilfie, A. T., Horvat, J. C., Wade, M. A., Beagley, K. W., Gibson, P. G., et al. (2007). Inhibition of allergic airways disease by immunomodulatory therapy with whole killed *Streptococcus pneumoniae*. *Vaccine* 25 (48), 8154–8162. doi: 10.1016/j.vaccine.2007.09.034
- Qu, Y., Zhang, H., Liang, C., and Dong, X. (2018). KATZMDA: prediction of miRNA-disease associations based on KATZ model. *IEEE Access* 6, 3943–3950. doi: 10.1109/access.2017.2754409
- Rados, J., Dobric, I., Pasic, A., Lipozencic, J., Ledic-Drvar, D., and Stajminger, G. (2007). Normalization in the appearance of severely damaged psoriatic nails using soft x-rays. A case report. *Acta Dermatovenerol Croat* 15 (1), 27–32.
- Said, H. S., Suda, W., Nakagome, S., Chinen, H., Oshima, K., Kim, S., et al. (2014). Dysbiosis of salivary microbiota in inflammatory bowel disease and its association with oral immunological biomarkers. *DNA Res.* 21 (1), 15–25. doi: 10.1093/dnares/dst037
- Sokol, H., Seksik, P., Furet, J. P., Firmesse, O., Nion-Larmurier, I., Beaugerie, L., et al. (2009). Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflammation Bowel Dis.* 15 (8), 1183–1189. doi: 10.1002/ibd.20903
- Sommer, F., and Backhed, F. (2013). The gut microbiota—masters of host development and physiology. *Nat. Rev. Microbiol* 11 (4), 227–238. doi: 10.1038/nrmicro2974
- Sun, Y., Zhu, Z., You, Z. H., Zeng, Z., Huang, Z. A., and Huang, Y. A. (2018). FMSM: a novel computational model for predicting potential miRNA biomarkers for various human diseases. *BMC Syst. Biol.* 12 (Suppl 9), 121. doi: 10.1186/s12918-018-0664-9
- Takaishi, H., Matsuki, T., Nakazawa, A., Takada, T., Kado, S., Asahara, T., et al. (2008). Imbalance in intestinal microflora constitution could be involved in the pathogenesis of inflammatory bowel disease. *Int. J. Med. Microbiol* 298 (5–6), 463–472. doi: 10.1016/j.ijmm.2007.07.016
- Vael, C., Vanheirstraeten, L., Desager, K. N., and Goossens, H. (2011). Denaturing gradient gel electrophoresis of neonatal intestinal microbiota in relation to the development of asthma. *BMC Microbiol* 11, 68. doi: 10.1186/1471-2180-11-68
- van Laarhoven, T., Nabuurs, S. B., and Marchiori E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 27(21), 3036–3043.
- Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., and Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* 6 (1), e1000641. doi: 10.1371/journal.pcbi.1000641
- Walters, W. A., Xu, Z., and Knight, R. (2014). Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* 588 (22), 4223–4233. doi: 10.1016/j.febslet.2014.09.039
- Wang, J., and Jia, H. (2016). Metagenome-wide association studies: fine-mining the microbiome. *Nat. Rev. Microbiol* 14 (8), 508–522. doi: 10.1038/nrmicro.2016.83
- Wu, C., Gao, R., Zhang, D., Han, S., and Zhang, Y. (2018). PRWHMDA: human microbe-disease association prediction by random walk on the heterogeneous network with PSO. *Int. J. Biol. Sci.* 14 (8), 849–857. doi: 10.7150/ijbs.24539
- Yang, X., Gao, L., Guo, X., Shi, X., Wu, H., Song, F., et al. (2014). A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS One* 9 (1), e87797. doi: 10.1371/journal.pone.0087797
- Yu, J., Jiang, S. O., Kim, B. J., Song, Y. H., Kwon, J. W., Kang, M. J., et al. (2010). The Effects of *Lactobacillus rhamnosus* on the Prevention of Asthma in a Murine Model. *Allergy Asthma Immunol. Res.* 2 (3), 199–205. doi: 10.4168/air.2010.2.3.199
- Zeng, X., Liu, L., Lu, L., and Zou, Q. (2018). Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 34 (14), 2425–2432. doi: 10.1093/bioinformatics/bty112
- Zeng, X., Wang, W., Deng, G., Bing, J., and Zou, Q. (2019). Prediction of potential disease-associated MicroRNAs by using neural networks. *Mol. Ther. Nucleic Acids* 16, 566–575. doi: 10.1016/j.omtn.2019.04.010
- Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G., and Li, X. (2017). Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinf.* 18 (1), 18. doi: 10.1186/s12859-016-1415-9
- Zhang, W., Lu, X., Yang, W., Huang, F., Wang, B., Wang, A., et al. (2018a). "HNGRNMF: Heterogeneous Network-based Graph Regularized Nonnegative Matrix Factorization for predicting events of microbe-disease associations", in: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2018)*, (Madrid, Spain).
- Zhang, W., Yang, W., Lu, X., Huang, F., and Luo, F. (2018b). The bi-direction similarity integration method for predicting microbe-disease associations. *IEEE Access* 6, 38052–38061. doi: 10.1109/ACCESS.2018.2851751
- Zhang, W., Yue, X., Huang, F., Liu, R., Chen, Y., and Ruan, C. (2018c). Predicting drug-disease associations and their therapeutic function based on the

- drug-disease association bipartite network. *Methods* 145, 51–59. doi: 10.1016/j.ymeth.2018.06.001
- Zhang, W., Yue, X., Lin, W., Wu, W., Liu, R., Huang, F., et al. (2018d). Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinf.* 19 (1), 233. doi: 10.1186/s12859-018-2220-4
- Zhou, T., Ren, J., Medo, M., and Zhang, Y. C. (2007). Bipartite network projection and personal recommendation. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* 76 (4 Pt 2), 046115. doi: 10.1103/PhysRevE.76.046115
- Zou, S., Zhang, J., and Zhang, Z. (2017). A novel approach for predicting microbe-disease associations by bi-random walk on the heterogeneous network. *PLoS One* 12 (9), e0184394. doi: 10.1371/journal.pone.0184394

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Xie and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.