



Human Microbe-Disease Association Prediction With Graph Regularized Non-Negative Matrix Factorization

Bin-Sheng He^{1*}, Li-Hong Peng^{2*} and Zejun Li^{3,4*}

¹ The First Affiliated Hospital, Changsha Medical University, Changsha, China, ² School of Information Engineering, Changsha Medical University, Changsha, China, ³ College of Information Science and Engineering, Hunan University, Changsha, China, ⁴ School of Computer and Information Science, Hunan Institute of Technology, Hengyang, China

OPEN ACCESS

Edited by:

Hongsheng Liu,
Liaoning University, China

Reviewed by:

Edwin Wang,
University of Calgary, Canada
Jiawei Luo,
Hunan University, China

*Correspondence:

Bin-Sheng He
hbcsmsu@163.com
Li-Hong Peng
plhnhu@163.com
Zejun Li
lzjfox@163.com

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 16 August 2018

Accepted: 08 October 2018

Published: 01 November 2018

Citation:

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

A microbe is a microscopic organism which may exist in its single-celled form or in a colony of cells. In recent years, accumulating researchers have been engaged in the field of uncovering microbe-disease associations since microbes are found to be closely related to the prevention, diagnosis, and treatment of many complex human diseases. As an effective supplement to the traditional experiment, more and more computational models based on various algorithms have been proposed for microbe-disease association prediction to improve efficiency and cost savings. In this work, we developed a novel predictive model of Graph Regularized Non-negative Matrix Factorization for Human Microbe-Disease Association prediction (GRNMFHMDA). Initially, microbe similarity and disease similarity were constructed on the basis of the symptom-based disease similarity and Gaussian interaction profile kernel similarity for microbes and diseases. Subsequently, it is worth noting that we utilized a preprocessing step in which unknown microbe-disease pairs were assigned associated likelihood scores to avoid the possible negative impact on the prediction performance. Finally, we implemented a graph regularized non-negative matrix factorization framework to identify potential associations for all diseases simultaneously. To assess the performance of our model, cross validations including global leave-one-out cross validation (LOOCV) and local LOOCV were implemented. The AUCs of 0.8715 (global LOOCV) and 0.7898 (local LOOCV) proved the reliable performance of our computational model. In addition, we carried out two types of case studies on three different human diseases to further analyze the prediction performance of GRNMFHMDA, in which most of the top 10 predicted disease-related microbes were verified by database HMDAD or experimental literatures.

Keywords: microbe, disease, association prediction, graph regularization, matrix factorization

INTRODUCTION

Antonie Van Leeuwenhoek, the father of microbiology, was the first to discover, observe, describe, study, and conduct scientific experiments with microbes, using simple single-lensed microscopes of his own design in 1673 (Leeuwenhoek, 1683-1775). From then on, with the development of biological theory and technology, a great mass of microbes has been discovered. It has been suggested that the amount of organisms living below the Earth's surface is comparable with the amount of life on or above the surface (Gold, 1992). As we know, microbes are very closely related

to humans in many fields, such as food production (Smid and Lacroix, 2013), water treatment (Tabatabaei et al., 2010), energy (Tanaka, 1999), and human health (Thiele et al., 2013). Especially, many studies have demonstrated that one of the most important effects of microbes on humans is the associations between microbes and complex human diseases. For example, Boleij et al. (2015) proved that the *Bacteroides fragilis* toxin gene is associated with colorectal neoplasia, especially in late-stage colorectal cancer (CRC). Moreover, Galiana et al. (2014) found that *Actinomyces* can be as an indicator in the evolution of chronic obstructive pulmonary disease (COPD) patients because their study confirmed a strong association between the presence or absence of *Actinomyces* and the severity of the clinical condition. Another example is that periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* stimulate tumorigenesis of oral squamous cell carcinoma (OSCC) via direct interaction with oral epithelial cells through Toll-like receptors which is beneficial to the development of corresponding prevention and treatment schemes (Binder Gallimidi et al., 2015). Thus, due to the fact that detecting potential microbiological markers could help to provide a better understanding of the pathogenesis of diseases and the role played by the microbiota in its severity, it is of great significance to explore the potential associations between microbes and diseases. However, since traditional experimental methods always suffer from the time constraints and capital limitations, proposing novel computational models is able to be an effective complement for uncovering potential microbe-disease associations. Recently, many feasible and effective prediction models have been developed by researchers.

In the last few years, some prediction models were proposed based on network analysis. Ma et al. (2017) developed an analysis method based on the microbe-based human disease network (Human Microbe Disease Network, HMDN) to infer the associations between microbes and disease genes, symptoms, chemical fragments, and drugs. In the method, they first utilized a large-scale text mining-based method to build the microbe-disease association network, on which the cosine similarity was calculated for each disease pair to construct the HMDN. Taking microbe-disease gene association prediction as an example, the potential related disease genes of a microbe in the HMDN can be finally obtained by finding the highly overlapped genes among the microbe-related diseases in the gene-based human disease network (Human Gene Disease Network, HGDN). Besides, in a similar way, this analysis method can also be used between HMDN and symptom-based human disease network (Human Symptoms Disease Network, HSDN), chemical fragment-based human disease network (Human Chemical Fragments Disease Network, HCDN), and drug-based human disease network (Human Drug Disease Network, HDDN) to infer the associations between microbes and disease symptoms, chemical fragments, and drugs, respectively. However, the prediction performance of this analysis method is limited by the small microbe-based disease network. Thereafter, Chen et al. (2017a) was the first to propose a computational model of KATZ measure for Human Microbe-Disease Association prediction (KATZHMDA) on a large scale. Firstly, they integrated the known microbe-disease associations network and Gaussian

interaction profile kernel similarity networks of microbes and diseases into a heterogeneous graph. Through summarizing all walks with different weighted lengths (i.e., the walk with shorter length was assigned larger coefficient) for each microbe-disease pair, they finally calculated the association probability between each microbe and disease. Moreover, KATZHMDA is applicable for new diseases/microbes without known associations if there are additional available similarity information between the new disease/microbe and other diseases/microbes in the known microbe-disease association network. One limitation of KATZHMDA is that the optimal value of the number of walks is still hard to select. Later, Huang Z. A. et al. (2017) proposed a model of Path-Based Human Microbe-Disease Association Prediction (PBHMDA) by integrating known microbe-disease association network and Gaussian interaction profile kernel similarity network for microbes and diseases into a heterogeneous interlinked network in which a threshold was set to remove the edges that represent weak correlations. In the heterogeneous interlinked network, the weights of all paths between a microbe-disease pair were finally aggregated to represent the association probability between the microbe and the disease, while the weight of each path was calculated by multiplying the weights of all edges in the path without overlap and then penalizing the path with a decay coefficient. The limitation existing in PBHMDA is that it will cause bias to microbes or diseases with more known associations. Moreover, PBHMDA cannot work well for new microbes and new diseases.

In addition, some proposed models were not based on network analysis. Since the negative microbe-disease samples (i.e., microbe-disease pairs that are confirmed to have no associations) are unavailable, Wang et al. (2017) presented a semi-supervised learning-based computational model of Laplacian Regularized Least Squares for Human Microbe-Disease Association prediction (LRLSHMDA) by optimizing the Laplacian regularized least squares classifiers in microbe space and disease space. Finally, they used a simple weighted average operation on the above two optimal classifiers to obtain the final probability matrix that indicates the potential association probabilities between microbes and diseases. However, LRLSHMDA is still faced with the problem of being unable to be implemented to new diseases without known associated microbes. Similarly, with no need for negative samples, Huang Y. A. et al. (2017) developed the method of a Neighbor- and Graph-based combined Recommendation model for Human Microbe-Disease Association prediction (NGRHMDA) by combining two recommendation models that are neighbor-based collaborative filtering model and topological information-based model. In the neighbor-based collaborative filtering model, considering that different microbe-disease pairs may share the same microbes or diseases, they computed two association possibility matrices respectively from the microbe perspective and disease perspective and then averaged them to obtain a prediction matrix. While in the topological information-based model, they introduced a two-step diffusion approach on the microbe-disease bipartite graph to obtain another prediction matrix. Ultimately, the above two prediction matrices

were simply averaged to get the final association possibilities for all microbe-disease pairs. What is worth noting is that NGRHMDA shares the same aforementioned disadvantage with LRLSHMDA.

In summary, all of the above models have their own limitations in predicting microbe-disease associations. Due to the lack of measurements for microbe/disease similarity, some models are only based on the Gaussian interaction profile kernel similarity of microbes and diseases that leads to unavoidable bias to those well-investigated diseases and microbes. Besides, some models cannot predict for new microbes/diseases and optimal parameters in some models are not easy to select. In this work, considering some of the above limitations, we developed a novel computational model of Graph Regularized Non-negative Matrix Factorization for Human Microbe-Disease Association prediction (GRNMFHMDA). First of all, the information of Gaussian interaction profile kernel similarity of microbes and diseases, symptom-based disease similarity and known microbe-disease associations in HMDAD (Ma et al., 2017) were combined as the input to start the whole prediction process. Here, after data preparation, the prediction process consists of two main steps, the preprocessing step and the step of GRNMF. In the preprocessing step, the weighted K nearest neighbor profiles for microbes and diseases were calculated to reconstruct the original adjacency matrix obtained based on the known microbe-disease associations so that we could avoid the possible negative impact on the final prediction performance from unknown microbe-disease pairs. While in the step of GRNMF, Tikhonov (L_2) and graph Laplacian regularization were introduced into the standard NMF framework to obtain a smoother solution from matrix factorization and take full advantage of the geometric structure of our data, respectively. In addition, global leave-one-out cross validation (LOOCV), local LOOCV and two types of case studies were carried out to evaluate the prediction performance of our model. As a result, GRNMFHMDA obtained AUCs of 0.8715 (global LOOCV) and 0.7898 (local LOOCV). More than that, 9 (Asthma), 9 (Obesity), and 8 (Type 1 diabetes) out of the top 10 predicted disease-related microbes were confirmed by HMDAD or experimental literatures. Thus, it is obvious that our model would perform well in microbe-disease association prediction according to the aforementioned results.

MATERIALS AND METHODS

Method Overview

Here, to predict potential associations between microbes and diseases, the model of GRNMFHMDA (See **Figure 1**) can be decomposed into three steps: (1) data preparation, in which adjacency matrix, microbe similarity, and disease similarity were established; (2) the preprocessing step, in which unknown microbe-disease pairs were assigned with associated likelihood scores based on the calculation of weighted K nearest neighbor profiles for microbes and diseases; (3) GRNMF, in which Tikhonov (L_2) and Graph Laplacian regularization were introduced into the standard NMF framework to obtain the final score matrix.

Human Microbe-Disease Associations

From the Human Microbe-Disease Association Database (HMDAD, <http://www.cuilab.cn/hmdad>) (Ma et al., 2017), we can download 483 known microbe-disease associations between 292 microbes and 39 human diseases. However, since some microbe-disease associations we downloaded are the same, there were only 450 known associations after removing the duplicate parts according to different evidences. In order to represent the associations information in a more convenient and efficient way, we defined an adjacency matrix $Y \in R^{m \times n}$, where m and n denoted the number of microbes and diseases, respectively. Moreover, the element $Y(m_i, d_j)$ was set to 1 if microbe m_i and disease d_j had known association, otherwise 0.

Gaussian Interaction Profile Kernel Similarity for Microbes

There is a hypothesis that similar microbes (i.e., microbes exhibiting a similar pattern of interaction and non-interaction with the diseases of a microbe-disease association network) are inclined to be associated with the same disease, on which many previous studies had relied to construct the Gaussian interaction profile kernel similarity for microbes (Chen et al., 2017a; Huang Z. A. et al., 2017). In this article, based on the same assumption, we first represent the interaction profile for each microbe with a binary vector involving the association information between the microbe and each disease in the known microbe-disease association network. On the basis of the definition of adjacency matrix Y , the i th row vector ($Y(m_i) = (Y_{i1}, Y_{i2}, \dots, Y_{in})$) can be used to denote the interaction profile of microbe m_i . Thus, according to the method of van Laarhoven et al. (2011), the Gaussian interaction profile kernel similarity between microbe m_i and m_j can be defined as follows:

$$S^m(m_i, m_j) = \exp(-\gamma_m \|Y(m_i) - Y(m_j)\|^2) \quad (1)$$

where

$$\gamma_m = \gamma'_m / \left(\frac{1}{m} \sum_{i=1}^m \|Y(m_i)\|^2 \right) \quad (2)$$

Here, γ_m is the adjustment coefficient that can be obtained by normalizing another bandwidth parameter γ'_m .

Gaussian Interaction Profile Kernel Similarity for Diseases

The construction of the Gaussian interaction profile kernel similarity for diseases is based on the assumption that similar diseases (i.e., diseases exhibiting a similar pattern of interaction and non-interaction with the microbes of a microbe-disease association network) are more likely to be associated with similar microbes. Here, the interaction profile for each disease is also represented by a binary vector containing the association information between the disease and each microbe in the known microbe-disease association network. Based on the same method of van Laarhoven et al. (2011), the j th column vector ($Y(d_j) =$

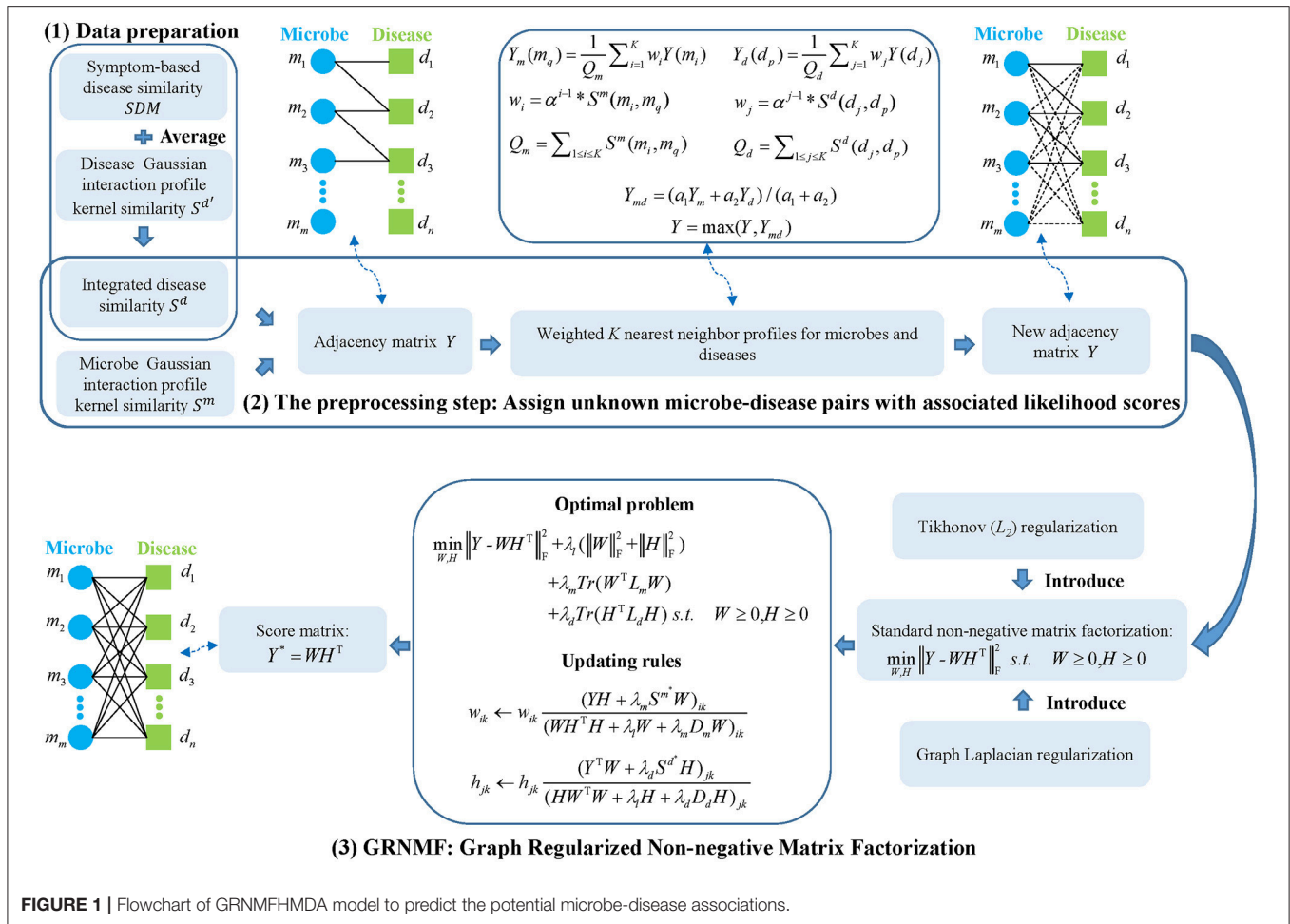


FIGURE 1 | Flowchart of GRNMFHMDA model to predict the potential microbe-disease associations.

$(Y_{1j}, Y_{2j}, \dots, Y_{mj})$ denotes the interaction profile of disease d_j and the Gaussian interaction profile kernel similarity between disease d_i and d_j can be defined as follows:

$$S^d(d_i, d_j) = \exp(-\gamma_d \|Y(d_i) - Y(d_j)\|^2) \tag{3}$$

where

$$\gamma_d = \gamma'_d / \left(\frac{1}{n} \sum_{j=1}^n \|Y(d_j)\|^2 \right) \tag{4}$$

Similarly, γ_d is the adjustment coefficient that can be calculated by normalizing another bandwidth parameter γ'_d .

Integrated Symptom-Based Disease Similarity

As we have mentioned above, Gaussian interaction profile kernel similarity is used in our model to measure the similarity of microbes and diseases. However, since the Gaussian interaction profile kernel similarity is an association information-based measurement, it is essential to combine more types of microbe

or disease similarities based on other available biological information. Indeed, according to different biological data, many researchers have developed their own method to measure the similarity of microbes or diseases. For instance, Zhou et al. (2014) proposed a model of symptom-based human disease network (HSDN) to measure the disease similarity based on co-occurrence of disease/symptom terms recorded in different literatures. In this work, we implemented HSDN to calculate symptom-based disease similarity (SDM) and then constructed a new disease similarity matrix (S^d) by integrating SDM with $S^{d'}$ in an average way according to the study of Chen et al. (2017a):

$$S^d = \frac{S^{d'} + SDM}{2} \tag{5}$$

Weighted K Nearest Neighbor Profiles for Microbes and Diseases

Due to the fact that values in interaction profiles of microbes or diseases without known associations are all zeros, the prediction performance may be affected to some extent. Considering that, to deal with the above mentioned problem, we came up with a preprocessing step to establish new interaction profiles both for

microbes and diseases. For each microbe m_q , we first find out its K nearest known microbes, each of which must has at least one known association. Next, the similarity information between m_q and its K nearest known microbes together with the information of their corresponding K interaction profiles are combined to calculate the new interaction profile as follows:

$$Y_m(m_q) = \frac{1}{Q_m} \sum_{i=1}^K w_i Y(m_i) \tag{6}$$

where

$$w_i = \alpha^{i-1} * S^m(m_i, m_q) \tag{7}$$

$$Q_m = \sum_{1 \leq i \leq K} S^m(m_i, m_q) \tag{8}$$

Here, m_1 to m_K denote the K nearest known microbes of m_q which were sorted in descending order based on the similarity values between them. The function of the weight coefficient w_i is that the corresponding similarity value is assigned higher weight if m_i is more similar to m_q . Besides, α is a decay term whose value is in the range of $[0,1]$ and Q_m is the normalization term.

In a similar way, the new interaction profile for each disease d_p can be defined as follows:

$$Y_d(d_p) = \frac{1}{Q_d} \sum_{j=1}^K w_j Y(d_j) \tag{9}$$

$$w_j = \alpha^{j-1} * S^d(d_j, d_p) \tag{10}$$

$$Q_d = \sum_{1 \leq j \leq K} S^d(d_j, d_p) \tag{11}$$

After calculating the new interaction profiles from microbe perspective and disease perspective, we combine Y_m and Y_d as follows:

$$Y_{md} = (a_1 Y_m + a_2 Y_d) / (a_1 + a_2) \tag{12}$$

where a_1 and a_2 are two weight coefficient and both of them are set to 1 for simplicity.

Finally, to replace the element $Y(m_i, d_j) = 0$ with an associated likelihood score, we use the following equation to update the original adjacency matrix Y .

$$Y = \max(Y, Y_{md}) \tag{13}$$

GRNMF

As a common method, the purpose of the standard NMF is to find two non-negative matrices whose product is an optimal approximation to the original matrix (Sotiras et al., 2015; Xu et al., 2015). Therefore, the adjacency matrix $Y \in R^{m \times n}$ can be decomposed into two parts after implementing NMF, namely, $W \in R^{m \times k}$ and $H \in R^{n \times k}$ ($Y \approx WH^T$). Accordingly, we can further get the following standard optimization problem:

$$\min_{W,H} \|Y - WH^T\|_F^2 + L(W,H) \tag{14}$$

where $L(W,H)$ is a regularization term to prevent overfitting.

Here, motivated by the study of Xiao et al. (2017) and the standard NMF framework, we introduced other two terms, the Tikhonov (L_2) (Guan et al., 2011) and graph Laplacian regularization (Cai et al., 2011), to predict microbe-disease associations. The utilizing of Tikhonov regularization aims to obtain a smooth solution (W and H), while the purpose of introducing graph regularization is to ensure a part-based representation through taking full advantage of the data geometric structure. Thus, we can construct the optimization problem of GRNMF as follows:

$$\begin{aligned} \min_{W,H} & \|Y - WH^T\|_F^2 + \lambda_l (\|W\|_F^2 + \|H\|_F^2) \\ & + \lambda_m \sum_{i,p=1}^n \|w_i - w_p\|^2 S^{m*}_{ip} + \lambda_d \sum_{j,q=1}^m \|h_j - h_q\|^2 S^{d*}_{jq} \text{ s.t. } W \geq 0, H \geq 0 \end{aligned} \tag{15}$$

Here, λ_l , λ_m and λ_d are the corresponding regularization coefficients. Besides, w_i and h_j are defined as i th rows of W and j th rows of H , respectively. In order to avoid negative affects to the prediction performance of our model, we introduced sparse weight matrices of S^{d*} and S^{m*} that are constructed on the basis of the geometrical information of disease and microbe data spaces (S^d and S^m), respectively. Then, Equation (14) can be transformed into:

$$\begin{aligned} \min_{W,H} & \|Y - WH^T\|_F^2 + \lambda_l (\|W\|_F^2 + \|H\|_F^2) \\ & + \lambda_m \text{Tr}(W^T L_m W) + \lambda_d \text{Tr}(H^T L_d H) \text{ s.t. } W \geq 0, H \geq 0 \end{aligned} \tag{16}$$

where $\text{Tr}(\bullet)$ represents the trace of a matrix. Here, L_m and L_d are the corresponding graph Laplacian matrices for S^{m*} and S^{d*} that can be calculated as follows:

$$L_m = D_m - S^{m*} \tag{17}$$

$$L_d = D_d - S^{d*} \tag{18}$$

where D_m and D_d are the diagonal matrices whose entries are row (or column) sums of S^{m*} and S^{d*} , respectively.

Based on the information of the nearest neighbor graph on a scatter of data points, researchers came up with a conclusion that local geometric structure is able to be effectively modeled (Cai et al., 2011; Li et al., 2017). Since microbes or diseases appearing in the same cluster are more likely to behave similarly, according to the above conclusion, we construct the graph matrices S^{m*} and S^{d*} in terms of microbe space and disease space respectively on the basis of the p nearest neighbors and corresponding clustering information. Here, we use the ClusterONE method (Nepusz et al., 2012) to construct the graph S^{m*} from microbe space, in which the

weight matrix X^m is generated based on the microbe similarity matrix S^m as follows:

$$X_{ij}^m = \begin{cases} 1 & i \in N(m_j) \ \& \ j \in N(m_i), m_j \in C \\ 0 & i \notin N(m_j) \ \& \ j \notin N(m_i), m_j \notin C \\ 0.5 & \text{otherwise} \end{cases} \quad (19)$$

where $N(m_i)$ and $N(m_j)$ are the sets of p nearest neighbors of m_i and m_j , respectively. C denotes to any one of the clusters obtained by ClusterONE method and we define the graph matrix S^{m*} for microbes as follows:

$$\forall i, j \ S_{ij}^{m*} = X_{ij}^m S_{ij}^m \quad (20)$$

In a similar way as the computation of S^{m*} , we calculate the graph matrix S^{d*} according to the disease similarity matrix S^d .

Here, we defined $\Phi=[\varphi_{ik}]$ and $\Psi=[\psi_{jk}]$ as the Lagrange multipliers for the constrains $w_{ik} \geq 0$ and $h_{jk} \geq 0$, respectively. In this work, we first convert the optimization problem in Equation (15) to an unconstraint problem, then minimize this problem by utilizing the corresponding Lagrange function L_f as follows:

$$L_f = Tr(YY^T) - 2Tr(YHW^T) + Tr(WH^T HW^T) + \lambda_l Tr(WW^T) + \lambda_i Tr(HH^T) + \lambda_m Tr(W^T L_m W) + \lambda_d Tr(H^T L_d H) + Tr(\Phi W^T) + Tr(\Psi H^T) \quad (21)$$

To solve the above problem, we first calculate the partial derivatives with respect to W and H as follows:

$$\frac{\partial L_f}{\partial W} = -2YH + 2WH^T H + 2\lambda_l W + 2\lambda_m L_m W + \Phi \quad (22)$$

$$\frac{\partial L_f}{\partial H} = -2Y^T W + 2HW^T W + 2\lambda_i H + 2\lambda_d L_d H + \Psi \quad (23)$$

After using the Karush-Kuhn-Tucker (KKT) conditions of $\varphi_{ik} w_{ik} = 0$ and $\psi_{jk} h_{jk} = 0$ (Facchinei et al., 2014), we can obtain the equations for w_{ik} and h_{jk} as follows:

$$- (YH)_{ik} w_{ik} + (WH^T H)_{ik} w_{ik} + (\lambda_l W)_{ik} w_{ik} + [\lambda_m (D_m - S^{m*}) W]_{ik} w_{ik} = 0 \quad (24)$$

$$- (Y^T W)_{jk} h_{jk} + (HW^T W)_{jk} h_{jk} + (\lambda_i H)_{jk} h_{jk} + [\lambda_d (D_d - S^{d*}) H]_{jk} h_{jk} = 0 \quad (25)$$

Finally, on the basis of the above two equations, we can get the updating rules for w_{ik} and h_{jk} as follows:

$$w_{ik} \leftarrow w_{ik} \frac{(YH + \lambda_m S^{m*} W)_{ik}}{(WH^T H + \lambda_l W + \lambda_m D_m W)_{ik}} \quad (26)$$

$$h_{jk} \leftarrow h_{jk} \frac{(Y^T W + \lambda_d S^{d*} H)_{jk}}{(HW^T W + \lambda_i H + \lambda_d D_d H)_{jk}} \quad (27)$$

Based on the above two updating formulas, we can obtain the final two non-negative matrices W and H until convergence. Subsequently, we calculate the score matrix Y^* for microbe-disease pairs by utilizing $Y^* = WH^T$, in which the higher score of a microbe-disease pair indicates that the microbe is more likely to be associated with the corresponding disease. In addition, for better understanding, we provided the pseudocode of the whole GRNMF algorithm (See **Figure 2**).

RESULTS

Performance Evaluation

Cross validation, a widely used assessment method, was introduced to evaluate the prediction performance of GRNMFHMDA. In this study, we utilized two types of cross validations, namely, global LOOCV and local LOOCV. For the global LOOCV, each of the known microbe-disease associations was in turn considered to be the test sample while the remaining known associations were treated as the training samples. Besides, all of the unknown microbe-disease pairs were regarded as the candidate samples which would be used in the ranking process. After implementing GRNMFHMDA, we ranked each test sample with all candidate samples according to their predicted scores. As for local LOOCV, the difference is that the test sample was only ranked with the candidate samples involving the investigated disease.

In each cross validation process, we would consider that the test sample was successfully predicted if the ranking of the test sample was higher than the given threshold. Further, based on the ranks of all test samples, we drew a receiver operating characteristic (ROC) curve through calculating the ratio between true positive rate (TPR, sensitivity) and false positive rate (FPR, 1-specificity) under different thresholds both for global LOOCV and local LOOCV. Sensitivity meant the ratio between the number of test samples ranking higher than the given threshold and the number of positive samples (known microbe-disease associations), while 1-specificity denoted the percentage of the number of negative microbe-disease pairs whose ranks were lower than the given threshold. Moreover, area under the ROC curve (AUC) was calculated to make quantitative evaluation for our model's prediction performance. The model would be considered to be able to perfectly predict all associations if the value of AUC equaled to 1, while the model was only supposed to be able to make random prediction if the value of AUC equaled to 0.5. As a result, GRNMFHMDA obtained AUCs of 0.8715 and 0.7898 in global LOOCV and local LOOCV, respectively. Furthermore, the prediction performance of our model outperformed the KATZHMDA both in global LOOCV (0.8644) and local LOOCV (0.6998), which proved the superior accuracy and reliability of our model in predicting microbe-disease associations (See **Figure 3**).

Case Study

Here, we put forward two types of case studies on three different common human diseases with the purpose of further assessing the prediction performance of GRNMFHMDA. On the basis of the known microbe-disease associations in HMDAD,

Algorithm: GRNMF

Input: Matrix $Y \in R^{m \times n}$, $S^m \in R^{m \times m}$, and $S^d \in R^{n \times n}$, decay term α , subspace dimensionality k , neighborhood sizes K and p , regularization coefficients λ_1 , λ_m and λ_d

Output: Predicted association matrix Y^*

Initialize: randomly initialize the two non-negative matrices $W \in R^{m \times k}$ and $H \in R^{n \times k}$

1. calculate weighted K nearest neighbor profiles for microbes:

for each microbe m_q ($q \in [1, m]$) **do**

for $i \leftarrow 1$ to K **do**

$$w_i = \alpha^{i-1} * S^m(m_i, m_q)$$

end for

$$Q_m = \sum_{1 \leq i \leq K} S^m(m_i, m_q)$$

$$Y_m(m_q) = \frac{1}{Q_m} \sum_{i=1}^K w_i Y(m_i)$$

end for

2. calculate weighted K nearest neighbor profiles for diseases:

for each disease d_p ($p \in [1, n]$) **do**

for $j \leftarrow 1$ to K **do**

$$w_j = \alpha^{j-1} * S^d(d_j, d_p)$$

end for

$$Q_d = \sum_{1 \leq j \leq K} S^d(d_j, d_p)$$

$$Y_d(d_p) = \frac{1}{Q_d} \sum_{j=1}^K w_j Y(d_j)$$

end for

3. combine the two weighted K nearest neighbor profiles of microbes and diseases:

$$Y_{md} = (a_1 Y_m + a_2 Y_d) / (a_1 + a_2)$$

4. update the original adjacency matrix Y :

$$Y = \max(Y, Y_{md})$$

5. construct matrices S^{m^*} and S^{d^*} from S^m and S^d , respectively.

6. **while** not converged **do**

$$w_{ik} \leftarrow w_{ik} \frac{(YH + \lambda_m S^{m^*} W)_{ik}}{(WH^T H + \lambda_1 W + \lambda_m D_m W)_{ik}}$$

$$h_{jk} \leftarrow h_{jk} \frac{(Y^T W + \lambda_d S^{d^*} H)_{jk}}{(HW^T W + \lambda_1 H + \lambda_d D_d H)_{jk}}$$

end while

7. **return** $Y^* = WH^T$

FIGURE 2 | The pseudocode of the whole GRNMF algorithm.

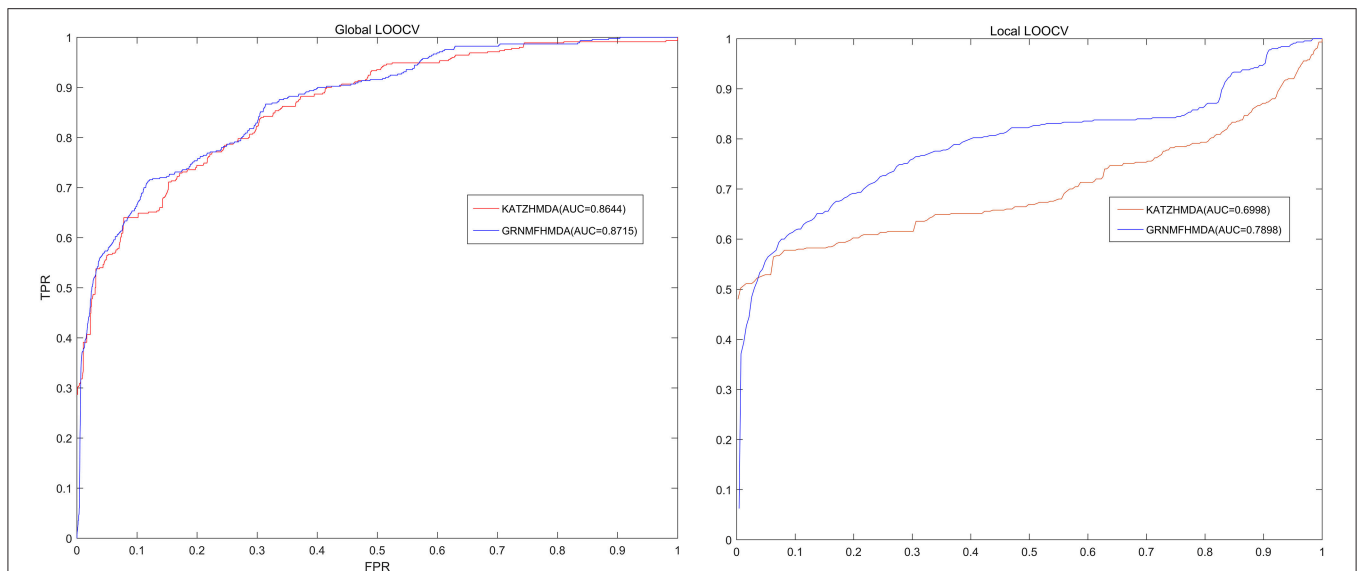


FIGURE 3 | The comparison of prediction performance between GRNMFHMDA and the classical model of KATZHMDA both in global LOOCV and local LOOCV. As a result, GRNMFHMDA achieved AUCs of 0.8715 and 0.7898 in the global and local LOOCV, which exceeded the first computational model of KATZHMDA in the field of microbe-disease association prediction.

we implemented GRNMFHMDA to predict disease-related microbes and then validated the top 10 predicted microbes by HMDAD or recent literatures.

Asthma, a common long-term inflammatory disease of the airways of the lungs, often starts during childhood and its average number of deaths and death rates (per 100,000 people) respectively reached to 38 and 0.1 in 2016 in the World Health Organization (WHO) European region among 10–14 years old children (Kyu et al., 2018). Here, under the GRNMFHMDA framework, asthma was treated as an investigated disease to explore its potential associated microbes. As a result, 9 out of the top 10 microbes in the prediction list were confirmed to be associated with asthma by experimental literatures (See **Table 1**). For example, *Lactobacillus casei rhamnosus* Lcr35, a species of *Lactobacillus* (1st in the prediction list), was found to be able to attenuate airway inflammation and hyperreactivity in a mouse model of asthma through oral treatment before sensitization (Yu et al., 2010). Besides, Ding et al. (2018) discovered that exosomes derived by *Pseudomonas* (2nd in the prediction list) *aeruginosa* could induce protection against allergic sensitization in asthma mice. Another example is that there is a distinct alteration of the sputum microbiota with a greater prominence of *Firmicutes* (4th in the prediction list) in severe asthma (Zhang et al., 2016).

Obesity, a medical condition in which accumulated excess body fat reaches a certain level that may have a negative effect on health, is a leading preventable cause of death worldwide (Reinier and Chugh, 2015). In recent years, plenty of studies have shown certain associations between obesity and microbes that helps a lot to the prevention and treatment of obesity. For instance, many researchers have demonstrated that *methanogens* play a specific role in weight gain and the development of obesity in human subjects (Armougom et al., 2009; Krajmalnik-Brown et al., 2012). Not only that, many studies have now been conducted into

TABLE 1 | Prediction list of the top 10 potential asthma-related microbes based on the known associations in HMDAD database and the corresponding validation evidences (experimental literatures in PubMed) for these associations.

Rank	Microbe	Evidence
1	Lactobacillus	PMID: 20592920
2	Pseudomonas	PMID: 29795208
3	Burkholderia	unconfirmed
4	Firmicutes	PMID: 27078029
5	Actinobacteria	PMID: 23265859
6	Clostridium coccoides	PMID: 21477358
7	Streptococcus	PMID: 17928596
8	Clostridia	PMID: 22047069
9	Lachnospiraceae	PMID: 26512904
10	Fusobacterium	PMID: 24024497

the potential of probiotics to ameliorate obesity and diabetes (Delzenne et al., 2011; Peterson et al., 2015). Therefore, taking obesity as another investigated disease in the first type of case study, we found that 9 out of the top 10 predicted obesity-related microbes were confirmed by experimental literatures (See **Table 2**). For the phylum *Proteobacteria* (1st in the prediction list) which belongs to gram-negative bacteria, the existing study already discovered that it was abundant in the obese group compared with lean group (Park et al., 2015). Besides, as a species of *Clostridia* (2nd in the prediction list), the presence of *Clostridium ramosum* in simplified human intestinal (SIHUMI) enhanced diet-induced obesity according to the experiment data of Woting et al. (2014). Moreover, *Bacillus*, a genus of *Clostridia* (3rd in the prediction list), was found to have outgrown dramatically in the obesity group by Gao et al. (2018).

TABLE 2 | Prediction list of the top 10 potential obesity-related microbes based on the known associations in HMDAD database and the corresponding validation evidences (experimental literatures in PubMed) for these associations.

Rank	Microbe	Evidence
1	Proteobacteria	PMID: 25407880
2	Clostridia	PMID: 25271283
3	Bacilli	PMID: 29280312
4	Faecalibacterium prausnitzii	PMID: 19849869
5	Clostridium	PMID: 23645850
6	Betaproteobacteria	PMID: 29312822
7	Clostridium coccooides	PMID: 23147032
8	Lactobacillus	PMID: 23056479
9	Fusobacterium nucleatum	unconfirmed
10	Prevotella	PMID: 21695273

More than that, in order to facilitate future researchers to study the disease-related microbes that they are interested in, based on the known associations in HMDAD, we provided the whole prediction list including all pairs between 292 microbes and 39 diseases as well as their predicted association scores (See **Supplementary Table 1**).

In addition, to prove the predictive applicability of our model on new diseases without known associated microbes, we carried out another case study on a disease via removing all its known associations in HMDAD. In this way, the prediction process of seeking the investigated disease-related microbes can only depend on the information of other known microbe-disease associations (training samples) and the relevant similarity measures. What needs to be emphasized is that only candidate samples (all microbe-disease pairs including the investigated disease) were ranked and then verified in HMDAD. Hence, there was no overlap between training samples and prediction list. In other words, the verification of predicted associations was independent of HMDAD. Type 1 diabetes, a form of diabetes mellitus, is believed to involve a combination of genetic and environmental factors such as dietary agents (Serena et al., 2015), viral infections (Rewers and Ludvigsson, 2016) and gut microbiota (Bibbò et al., 2017). Especially in gut microbiota, the previous study confirmed that the genus *Bacteroides* is the largest representative of type 1 diabetes-associated dysbiosis that can be modulated by diet (Mejía-León and Barca, 2015). Thus, considering the significance of studying type 1 diabetes-related microbes, we took type 1 diabetes as the investigated disease to predict its potential associated microbes under the framework of the second type of case study. After implementing GRNMFHMDA, we obtained the ranks of type 1 diabetes' candidate microbes in terms of their association scores (See **Table 3**). As a result, 8 out of the top 10 predictions were confirmed by HMDAD or recent literatures. For example, Giongo et al. (2011) demonstrated that the *Clostridia* (1st in the prediction list) sequences increased in control samples (samples of general population) as the abundance of *Clostridia* decreased overtime in the case samples (samples of patients with type 1 diabetes). Moreover, at the phylum level and at p -values <

TABLE 3 | Prediction list of the top 10 potential type 1 diabetes-related microbes via removing all the known type 1 diabetes-microbe associations in HMDAD database.

Rank	Microbe	Evidence
1	Clostridia	confirmed by HMDAD
2	Proteobacteria	confirmed by HMDAD
3	Clostridium coccooides	unconfirmed
4	Lactobacillus	confirmed by HMDAD
5	Bacteroidetes	confirmed by HMDAD
6	Firmicutes	confirmed by HMDAD
7	Faecalibacterium prausnitzii	PMID: 23934614
8	Clostridium	PMID: 23433344
9	Betaproteobacteria	unconfirmed
10	Bacilli	PMID: 24930037

The validation evidences denote to whether the predicted associations were confirmed by the HMDAD database or experimental literatures in PubMed.

0.001, *Proteobacteria* (2nd in the prediction list) was found to be higher in case samples than that in control samples (Brown et al., 2011). Another example is that *Lactobacillus strains* (a species of *Lactobacillus* ranking 4th in the prediction list) was found to be able to induce specific changes in the immune system of non-obese diabetic (NOD) mice that can increase or decrease diabetes (Brown et al., 2011).

According to the results presented, GRNMFHMDA consistently achieved an excellent predictive performance in the two types of case studies. With the continuous experimental research on microbe-disease associations, we expect that more and more microbes in the prediction lists generated by our model would be verified in the future.

DISCUSSION

In this article, we proposed a novel prediction model of GRNMFHMDA based on the known microbe-disease associations in HMDAD, Gaussian interaction profile kernel similarity of microbes and diseases and symptom-based disease similarity. To eliminate the possible problem caused by unknown microbe-disease pairs that may affect our final prediction performance, we first implemented a preprocessing step to establish new interaction profiles both for microbes and diseases. Then, after introducing Tikhonov (L_2) and graph Laplacian regularization under the standard NMF framework, we finally obtained reliable and satisfactory prediction performance both in LOOCV and case studies. Therefore, we can conclude that our prediction model is able to play critical role in revealing the associations between microbes and diseases, thus improving the prevention, diagnosis and treatment of many complex human diseases in the future.

Here, the reason why GRNMFHMDA performed well in microbe-disease association prediction lies in the following facts. Firstly, in the study of Wang et al. (2015), to model cancer hallmark traits and networks, nodes and links in the network were weighted, and certain scoring functions were developed to

represent gene regulatory logics/strengths on networks. Inspired by that, based on the data extracted from the acknowledged databases, we implemented proper and effective measurements to quantify microbe-disease association network, microbe similarity network and disease similarity network, which guaranteed the reliable prediction performance of our model. Secondly, before implementing GRNMF, we constructed new interaction profiles both for microbes and diseases to further assign those unknown microbe-disease pairs with associated likelihood score, which also improved our model's performance in some degree. Thirdly, different from the standard NMF, we introduced Tikhonov (L_2) and graph Laplacian regularization that ensured the final two non-negative matrices smoothness and guaranteed a part-based representation via fully exploiting the data geometric structure, respectively.

Nevertheless, here are also some limitations restricting the accuracy of our model that need to be overcome in future studies. Initially, the types of similarities for microbes and diseases are not enough yet and we believe that our model would be significantly improved with more biological data and similarity measurements being taken into consideration. Successful advance in association prediction research in various fields of computational biology would also accelerate the development of effective models for microbe-disease association prediction (Chen and Yan, 2013; Chen et al., 2016, 2017b, 2018a,b,c; Chen and Huang, 2017; You et al., 2017). Secondly, as shown in the research of Hao et al. (2018), three representative genome-scale cellular networks, genome-scale metabolic network (GMN), transcriptional regulatory network (TRN), and signal transduction network (STN), were found to be able to become a necessary tool in the systematic analysis of microbes through network integration. Therefore, whether there

are similar molecular networks between two microbes is well worth studying in constructing our prediction model. Thirdly, the selection of the optimal parameters is still worth studying. Finally, GRNMFHMDA would inevitably cause bias to diseases that have more known associated microbes and vice versa. Hence, we would come up with optimization strategies to deal with those limitations in our next work.

AUTHOR CONTRIBUTIONS

B-SH conceived the project, developed the prediction method, analyzed the result, and revised the paper. L-HP designed the experiments, implemented the experiments, analyzed the result, and wrote the paper. ZL analyzed the result and revised the paper. All authors read and approved the final manuscript.

FUNDING

B-SH was supported by Key Program of Hunan Provincial Education Department (Grant No. 15A026), General Program of Hunan Provincial Philosophy and Social Science Planning Fund office (Grant No. 15YBA035). L-HP was supported by Natural Science Foundation of Hunan province under Grant No. 2018JJ3570. ZL was supported by National Nature Science Foundation of China (Grant No. 61672223).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02560/full#supplementary-material>

REFERENCES

- Armougom, F., Henry, M., Vialettes, B., Raccach, D., and Raoult, D. (2009). Monitoring bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in obese patients and *Methanogens* in anorexic patients. *PLoS ONE* 4:e7125. doi: 10.1371/journal.pone.0007125
- Bibbò, S., Dore, M. P., Pes, G. M., Delitala, G., and Delitala, A. P. (2017). Is there a role for gut microbiota in type 1 diabetes pathogenesis? *Ann. Med.* 49, 11–22. doi: 10.1080/07853890.2016.1222449
- Binder Gallimidi, A., Fischman, S., Revach, B., Bulvik, R., Maliutina, A., Rubinstein, A. M., et al. (2015). Periodontal pathogens *Porphyromonas gingivalis* and *Fusobacterium nucleatum* promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget* 6, 22613–22623. doi: 10.18632/oncotarget.4209
- Bolej, A., Hechenbleikner, E. M., Goodwin, A. C., Badani, R., Stein, E. M., Lazarev, M. G., et al. (2015). The *Bacteroides fragilis* toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clin. Infect. Dis.* 60, 208–215. doi: 10.1093/cid/ciu787
- Brown, C. T., Davis-Richardson, A. G., Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., et al. (2011). Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. *PLoS ONE* 6:e25792. doi: 10.1371/journal.pone.0025792
- Cai, D., He, X., Han, J., and Huang, T. S. (2011). graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 1548–1560. doi: 10.1109/TPAMI.2010.231
- Chen, X., and Huang, L. (2017). LRSSLMDA: Laplacian Regularized Sparse Subspace Learning for MiRNA-Disease Association prediction. *PLoS Comput. Biol.* 13:e1005912. doi: 10.1371/journal.pcbi.1005912
- Chen, X., Huang, Y. A., You, Z. H., Yan, G. Y., and Wang, X. S. (2017a). A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* 33, 733–739. doi: 10.1093/bioinformatics/btw715
- Chen, X., Ren, B., Chen, M., Wang, Q., Zhang, L., and Yan, G. (2016). NLSS: Predicting Synergistic Drug Combinations Based on Semi-supervised Learning. *PLoS Comput. Biol.* 12:e1004975. doi: 10.1371/journal.pcbi.1004975
- Chen, X., Wang, L., Qu, J., Guan, N. N., and Li, J. Q. (2018a). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* doi: 10.1093/bioinformatics/bty503. [Epub ahead of print].
- 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, 3178–3186. doi: 10.1093/bioinformatics/bty333
- Chen, X., Xie, D., Zhao, Q., and You, Z. H. (2017b). MicroRNAs and complex diseases: from experimental results to computational models. *Brief. Bioinformatics*. doi: 10.1093/bib/bbx130. [Epub ahead of print].
- Chen, X., and Yan, G. Y. (2013). Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* 29, 2617–2624. doi: 10.1093/bioinformatics/btt426
- Chen, X., Yin, J., Qu, J., and Huang, L. (2018c). MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLoS Comput. Biol.* 14:e1006418. doi: 10.1371/journal.pcbi.1006418

- Delzenne, N. M., Neyrinck, A. M., Bäckhed, F., and Cani, P. D. (2011). Targeting gut microbiota in obesity: effects of prebiotics and probiotics. *Nat. Rev. Endocrinol.* 7, 639–646. doi: 10.1038/nrendo.2011.126
- Ding, F. X., Liu, B., Zou, W. J., Li, Q. B., Tian, D. Y., and Fu, Z. (2018). *Pseudomonas aeruginosa*-derived exosomes ameliorates allergic reactions via inducing the Treg response in asthma. *Pediatr. Res.* 84, 125–133. doi: 10.1038/s41390-018-0020-1
- Facchinei, F., Kanzow, C., and Sagratella, S. (2014). Solving quasi-variational inequalities via their KKT conditions. *Math. Program.* 144, 369–412. doi: 10.1007/s10107-013-0637-0
- Galiana, A., Aguirre, E., Rodriguez, J. C., Mira, A., Sañtibañez, M., Candela, I., et al. (2014). Sputum microbiota in moderate versus severe patients with COPD. *Eur. Respir. J.* 43, 1787–1790. doi: 10.1183/09031936.00191513
- Gao, R., Zhu, C., Li, H., Yin, M., Pan, C., Huang, L., et al. (2018). Dysbiosis signatures of gut microbiota along the sequence from healthy, young patients to those with overweight and obesity. *Obesity* 26, 351–361. doi: 10.1002/oby.22088
- Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., Novelo, L. L., Casella, G., et al. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J.* 5, 82–91. doi: 10.1038/ismej.2010.92
- Gold, T. (1992). The deep, hot biosphere. *Proc. Natl. Acad. Sci. U.S.A.* 89, 6045–6049. doi: 10.1073/pnas.89.13.6045
- Guan, N., Tao, D., Luo, Z., and Yuan, B. (2011). Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent. *IEEE Trans. Image Process.* 20, 2030–2048. doi: 10.1109/TIP.2011.2105496
- Hao, T., Wu, D., Zhao, L., Wang, Q., Wang, E., and Sun, J. (2018). The genome-scale integrated networks in microorganisms. *Front. Microbiol.* 9:296. doi: 10.3389/fmicb.2018.00296
- Huang, Y. A., You, Z. H., Chen, X., Huang, Z. A., Zhang, S., and Yan, G. Y. (2017). Prediction of microbe-disease association from the integration of neighbor and graph with collaborative recommendation model. *J. Transl. Med.* 15, 209. doi: 10.1186/s12967-017-1304-7
- Huang, Z. A., Chen, X., Zhu, Z., Liu, H., Yan, G. Y., You, Z. H., et al. (2017). PBHMDA: Path-Based Human Microbe-Disease Association Prediction. *Front. Microbiol.* 8:233. doi: 10.3389/fmicb.2017.00233
- Krajmalnik-Brown, R., Ilhan, Z. E., Kang, D. W., and Dibaise, J. K. (2012). Effects of gut microbes on nutrient absorption and energy regulation. *Nutr. Clin. Pract.* 27, 201–214. doi: 10.1177/0884533611436116
- Kyu, H. H., Stein, C. E., Boschi Pinto, C., Rakovac, I., Weber, M. W., Dannemann Purnat, T., et al. (2018). Causes of death among children aged 5–14 years in the WHO European Region: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Child Adolesc. Health* 2, 321–337. doi: 10.1016/S2352-4642(18)30095-6
- Leeuwenhoek, A. V. (1683-1775). Part of a Letter from Mr Antony van Leeuwenhoek, F. R. S. concerning Green Weeds Growing in Water, and Some Animalcula Found about Them. *Philos. Trans.* 23, 1304–1311.
- Li, X., Cui, G., and Dong, Y. (2017). Graph regularized non-negative low-rank matrix factorization for image clustering. *IEEE Trans. Cybern.* 47, 3840–3853. doi: 10.1109/TCYB.2016.2585355
- Ma, W., Zhang, L., Zeng, P., Huang, C., Li, J., Geng, B., et al. (2017). An analysis of human microbe-disease associations. *Brief. Bioinformatics* 18, 85–97. doi: 10.1093/bib/bbw005
- Mejjía-León, M. E., and Barca, A. M. (2015). Diet, microbiota and immune system in type 1 diabetes development and evolution. *Nutrients* 7, 9171–9184. doi: 10.3390/nu7115461
- Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* 9, 471–472. doi: 10.1038/nmeth.1938
- Park, H. J., Lee, S. E., Kim, H. B., Isaacson, R. E., Seo, K. W., and Song, K. H. (2015). Association of obesity with serum leptin, adiponectin, and serotonin and gut microflora in beagle dogs. *J. Vet. Intern. Med.* 29, 43–50. doi: 10.1111/jvim.12455
- Peterson, C. T., Sharma, V., Elmén, L., and Peterson, S. N. (2015). Immune homeostasis, dysbiosis and therapeutic modulation of the gut microbiota. *Clin. Exp. Immunol.* 179, 363–377. doi: 10.1111/cei.12474
- Reinier, K., and Chugh, S. S. (2015). Obesity and sudden death: visceral response? *Heart* 101, 165–166. doi: 10.1136/heartjnl-2014-306921
- Rewers, M., and Ludvigsson, J. (2016). Environmental risk factors for type 1 diabetes. *Lancet* 387, 2340–2348. doi: 10.1016/S0140-6736(16)30507-4
- Serena, G., Camhi, S., Sturgeon, C., Yan, S., and Fasano, A. (2015). The role of gluten in celiac disease and type 1 diabetes. *Nutrients* 7, 7143–7162. doi: 10.3390/nu7095329
- Smid, E. J., and Lacroix, C. (2013). Microbe-microbe interactions in mixed culture food fermentations. *Curr. Opin. Biotechnol.* 24, 148–154. doi: 10.1016/j.copbio.2012.11.007
- Sotiras, A., Resnick, S. M., and Davatzikos, C. (2015). Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. *Neuroimage* 108, 1–16. doi: 10.1016/j.neuroimage.2014.11.045
- Tabatabaei, M., Rahim, R. A., Abdullah, N., Wright, A. D. G., Shirai, Y., Sakai, K., et al. (2010). Importance of the methanogenic archaea populations in anaerobic wastewater treatments. *Process Biochem.* 45, 1214–1225. doi: 10.1016/j.procbio.2010.05.017
- Tanaka, K. (1999). From the fryer to the fuel tank: the complete guide to using vegetable oil as an alternative fuel. *Green Teacher* 15, 46.
- Thiele, I., Heinken, A., and Fleming, R. M. (2013). A systems biology approach to studying the role of microbes in human health. *Curr. Opin. Biotechnol.* 24, 4–12. doi: 10.1016/j.copbio.2012.10.001
- van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500
- Wang, E., Zaman, N., Mcgee, S., Milanese, J. S., Masoudi-Nejad, A., and O'Connor-Mccourt, M. (2015). Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin. Cancer Biol.* 30, 4–12. doi: 10.1016/j.semcancer.2014.04.002
- Wang, F., Huang, Z. A., Chen, X., Zhu, Z., Wen, Z., Zhao, J., et al. (2017). LRLSHMDA: Laplacian regularized least squares for human microbe-disease association prediction. *Sci. Rep.* 7:7601. doi: 10.1038/s41598-017-08127-2
- Woting, A., Pfeiffer, N., Loh, G., Klaus, S., and Blaut, M. (2014). Clostridium ramosum promotes high-fat diet-induced obesity in gnotobiotic mouse models. *MBio* 5, e01530–e01514. doi: 10.1128/mBio.01530-14
- Xiao, Q., Luo, J., Liang, C., Cai, J., and Ding, P. (2017). A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 34, 239–248. doi: 10.1093/bioinformatics/btx545
- Xu, J., Xiang, L., Wang, G., Ganesan, S., Feldman, M., Shih, N. N., et al. (2015). Sparse Non-negative Matrix Factorization (SNMF) based color unmixing for breast histopathological image analysis. *Comput. Med. Imaging Graph* 46(Pt 1), 20–29. doi: 10.1016/j.compmedimag.2015.04.002
- You, Z. H., Huang, Z. A., Zhu, Z., Yan, G. Y., Li, Z. W., Wen, Z., et al. (2017). PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* 13:e1005455. doi: 10.1371/journal.pcbi.1005455
- Yu, J., Jang, 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, 199–205. doi: 10.4168/aaair.2010.2.3.199
- Zhang, Q., Cox, M., Liang, Z., Brinkmann, F., Cardenas, P. A., Duff, R., et al. (2016). Airway microbiota in severe asthma and relationship to asthma severity and phenotypes. *PLoS ONE* 11:e0152724. doi: 10.1371/journal.pone.0152724
- Zhou, X., Menche, J., Barabási, A. L., and Sharma, A. (2014). Human symptoms-disease network. *Nat. Commun.* 5, 4212. doi: 10.1038/ncomms5212

Conflict of Interest Statement: 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 © 2018 He, Peng and Li. 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.