



Predicting miRNA-Disease Association Based on Modularity Preserving Heterogeneous Network Embedding

Wei Peng^{1,2*}, Jieli Du¹, Wei Dai^{1,2} and Wei Lan³

¹ Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China,

² Computer Technology Application Key Laboratory of Yunnan Province, Kunming University of Science and Technology, Kunming, China, ³ Guangxi Key Laboratory of Multimedia Communications and Network Technology, Guangxi University, Nanning, China

OPEN ACCESS

Edited by:

Lei Deng,
Central South University, China

Reviewed by:

Bolin Chen,
Northwestern Polytechnical University,
China
Xiujuan Lei,
Shaanxi Normal University, China

*Correspondence:

Wei Peng
weipeng1980@gmail.com

Specialty section:

This article was submitted to
Molecular Medicine,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 07 September 2020

Accepted: 23 March 2021

Published: 10 June 2021

Citation:

Peng W, Du J, Dai W and Lan W
(2021) Predicting miRNA-Disease
Association Based on Modularity
Preserving Heterogeneous Network
Embedding.
Front. Cell Dev. Biol. 9:603758.
doi: 10.3389/fcell.2021.603758

MicroRNAs (miRNAs) are a category of small non-coding RNAs that profoundly impact various biological processes related to human disease. Inferring the potential miRNA-disease associations benefits the study of human diseases, such as disease prevention, disease diagnosis, and drug development. In this work, we propose a novel heterogeneous network embedding-based method called MDN-NMTF (Module-based Dynamic Neighborhood Non-negative Matrix Tri-Factorization) for predicting miRNA-disease associations. MDN-NMTF constructs a heterogeneous network of disease similarity network, miRNA similarity network and a known miRNA-disease association network. After that, it learns the latent vector representation for miRNAs and diseases in the heterogeneous network. Finally, the association probability is computed by the product of the latent miRNA and disease vectors. MDN-NMTF not only successfully integrates diverse biological information of miRNAs and diseases to predict miRNA-disease associations, but also considers the module properties of miRNAs and diseases in the course of learning vector representation, which can maximally preserve the heterogeneous network structural information and the network properties. At the same time, we also extend MDN-NMTF to a new version (called MDN-NMTF2) by using modular information to improve the miRNA-disease association prediction ability. Our methods and the other four existing methods are applied to predict miRNA-disease associations in four databases. The prediction results show that our methods can improve the miRNA-disease association prediction to a high level compared with the four existing methods.

Keywords: heterogeneous network embedding, matrix factorization, miRNA, disease, miRNA-disease association prediction

INTRODUCTION

MicroRNA (miRNA) is a category of small endogenous single-stranded non-coding RNA molecules with about 22 nucleotides in length. They play an essential role in regulating gene expression and complex gene regulatory networks by repressing target mRNAs expression at the post-transcriptional level (Bartel, 2004; Meister and Tuschl, 2004). Studies show that about 60% of human protein-coding genes are targeted by miRNAs, where the 5' region of miRNA binds to 3'

UTR of the target mRNAs (Friedman et al., 2009). With the rapid development of biotechnology, increasing research has demonstrated that miRNAs play crucial roles at multiple stages of many critical biological processes such as early cell growth, development, proliferation, differentiation, tumor invasion, and apoptosis (Ambros, 2003). Furthermore, studies have shown that abnormality and dysregulations of disease-related miRNAs may cause human diseases (Garzon et al., 2010). Therefore, inferring the potential miRNA-disease association is of great benefit to studying human diseases, such as disease prevention, disease diagnosis, and drug development. As we all know, discovering the miRNA-disease associations through traditional biological experiments is a time-consuming and labor-intensive process. Instead, computational models would serve as a low-cost, and high-efficiency way of predicting miRNA-disease associations.

Previous researches observe that similar miRNAs tend to associate with the same diseases and similar diseases are highly likely related to the same miRNAs. Hence, many computational methods construct disease similarity network and miRNA similarity network and infer miRNA-disease associations based on the associations between or within the disease or miRNAs (Peng et al., 2016, 2017; Zou et al., 2016; Huang et al., 2019). Xuan et al. (2013) construct a miRNA similarity network according to the degree of two miRNAs sharing similar disease and consider the k most similar neighbors of each miRNA to infer miRNA-disease associations. Chen X. et al. (2012) implement a random walk on the miRNA functional similarity network and explore the potential miRNA-disease associations from the global network information. Xuan et al. (2015) divide the miRNA nodes in the miRNA similarity network into two categories: the given disease-related and the given disease-unrelated nodes. They assign different transition weights to different types of nodes and implement random walk on the miRNA similarity network to predict miRNA-disease associations. Besides the single network, some researchers build a heterogeneous network that consists of miRNAs, diseases, and their inter and intro associations. Liu et al. (2017) construct the miRNA similarity network, disease similarity network and known miRNA-disease association network. After that, they run a random walk on the heterogeneous network to propagate information and exploit potential miRNA-disease associations. Considering the difference in the network structure of the miRNA similarity network and disease similarity network, Luo and Xiao (2017) use an unbalanced Bi-Random walk (called UBiRW) on the heterogeneous network of disease similarity network, miRNA functional similarity network and a known miRNA-disease association network to infer potential miRNA-disease associations. Zeng et al. (2016) enumerate all of the paths from miRNA/disease to disease/miRNA in the heterogeneous network, and the final score between a miRNA and a disease is a linear combination of their path scores. You et al. (2017) construct the heterogeneous network by integrating known human miRNA-disease associations, miRNA functional similarity, disease semantic similarity, and the Gaussian Interaction Profile (GIP) kernel similarity. After that, they do a depth-first search to find the paths between the miRNAs and diseases on the heterogeneous network. Then they filter the

long paths and calculate the association of miRNA and disease by combining all their paths.

Recently, a group of researchers proposes the network embedding-based method to predict miRNA-disease associations. The network embedding method designs an objective function and converts the network nodes into a low dimensional vector while maximally preserves the network structural information. Chen and Yan (2014) develop a regularized least square method to learn the latent vectors for miRNAs and diseases on the miRNA similarity network and disease similarity network. They combine the two vectors to give the final solution of predicting new miRNA-disease associations. Lan et al. (2016) construct multi-kernels to store the miRNA functional similarity network, miRNA sequence similarity network and disease semantic similarity network. Then they employ a Bayesian matrix factorization method to infer potential miRNA-disease associations by integrating these data sources. Yan et al. (2019) develop a dynamic neighborhood regularized logistic matrix factorization method called DNRLMF-MDA to learn representation vectors for miRNAs and diseases and predict potential miRNA-disease associations. Li et al. (2017) design an objective function to ensure the scores of known miRNA-disease association matrix are close to those in the predicted miRNA-disease association matrix. They utilize the matrix completion algorithm to update the matrix of known miRNA-disease associations and to predict the potential associations. Xiao et al. (2018) use a graph regularized non-negative matrix factorization framework (named GRNMF) to identify possible associations for all diseases simultaneously. Similarly, Chen's group proposes two matrix completion-based methods, namely IMCMDA (Chen et al., 2018), and NCMCMDA (Chen et al., 2020) for miRNA-disease association prediction. The differences are IMCMDA uses inductive matrix completion for miRNA-disease association prediction, while NCMCMDA integrates neighborhood constraint in the course of matrix completion.

The methods mentioned above have achieved great success in predicting miRNA-disease associations. However, there are still some shortcomings in these existing methods. Firstly, the single network-based methods only use the miRNA similarity or disease similarity network. They may ignore the relationship between diseases or miRNAs. Secondly, seldom heterogeneous network-based methods consider the miRNA/disease similarity network's modular structure. Although some network embedding-based methods, i.e., DNRLMF-MDA, NCMCMDA, learn node representation only considering the constraint from part of neighbors, most of them ignore the modular information of miRNAs and diseases. Lu et al. (2008) constructed disease network by giving two diseases an edge if they share at least one common associated miRNA. Diseases cluster together, which suggests that some diseases form modules sharing similar associations at the miRNA level. Moreover, the disease-associated miRNAs show various dysfunctions, such as mutation, upregulation, deleted, and downregulation. On the other hand, groups of homologous miRNA belong to the same miRNA families. They might have similar functions, and therefore, their dysfunction would lead to a similar phenotype. By analyzing members in disease modules or miRNA modules, researchers

found that most of the members in the miRNA module are related to the same disease, and the members in the disease modules are mostly related to the same miRNA too. Therefore, this finding can guide us to predict novel disease-related miRNAs.

In this work, we propose a novel heterogeneous network embedding-based method for predicting miRNA-disease associations. We calculate the disease semantic similarity, diseases functional similarity, miRNA functional similarity, and compute the GIP kernel similarity of miRNAs and diseases. Then, we integrate these similarities and construct a heterogeneous network of the disease similarity network, miRNA functional similarity network and a known miRNA-disease association network. After that, we propose a Module-based Dynamic Neighborhood Non-negative Matrix Tri-Factorization (MDN-NMTF) to learn the latent vector representation for miRNAs and diseases in the heterogeneous network. Finally, the association probability is computed by the product of the latent miRNA and disease vectors. MDN-NMTF not only successfully integrates diverse biological information of miRNAs and diseases to predict miRNA-disease associations, but also considers the module properties of miRNAs and diseases in the course of learning vector representation, which can maximally preserve the heterogeneous network structural information and the network properties. Meanwhile, we also extend MDN-NMTF to a new version (called MDN-NMTF2) by using the modular information to improve the prediction ability of MDN-NMTF. Our methods, as well as the other four existing methods [DNRLMF-MDA (Yan et al., 2019), IMCMA (Chen et al., 2018), UBIRW (Luo and Xiao, 2017), and GRNMF (Xiao et al., 2018)], are applied to predict miRNA-disease associations on four data sets. The prediction results show that compared with the four existing methods, our methods can improve the performance of miRNA-disease association prediction to a high level.

MATERIALS

Four datasets (see **Table 1**), namely HMDD2.0-You (Li et al., 2014), HMDD2.0-Lan (Lan et al., 2016), HMDD2.0-Yan, and HMDD3.0¹, were used to evaluate our methods and the other existing methods. HMDD2.0-You, HMDD2.0-Lan, and HMDD2.0-Yan were from HMDD database version 2.0. The HMDD2.0-You dataset includes 495 miRNAs, 380 diseases and

¹<http://www.cuilab.cn/hmdd>

TABLE 1 | The number of miRNAs, diseases and miRNA-disease associations in four datasets.

Dataset	n_m	n_d	n_{md}
HMDD2.0-You	495	380	5424
HMDD2.0-Lan	550	329	6084
HMDD2.0-Yan	576	356	6391
HMDD3.0	1207	894	18732

The table shows the differences among the four datasets, where n_m , n_d , and n_{md} indicate the number of miRNAs, diseases and miRNA-disease associations, respectively.

5,424 miRNA-disease associations. The HMDD2.0-Lan dataset consists of 550 miRNAs, 329 diseases and 6,084 miRNA-disease associations. The HMDD2.0-Yan dataset includes 576 miRNAs, 356 diseases, and 6,391 miRNA-disease associations. The HMDD3.0 dataset came from HMDD database version 3.0, which involves 1,207 miRNAs, 894 diseases, and 18,732 miRNA-disease associations. To calculate the functional similarity of diseases, we extracted the functional similarity scores of gene-gene pairs from the HumanNet database that contains 16,243 genes and 476,399 associations (Lee et al., 2011). The disease-gene associations of HMDD2.0-You, HMDD2.0-Yan, and HMDD3.0 were obtained from the DisGeNET database, where includes 13,000 diseases, over 16,000 genes, and 380,000 disease-gene associations (Piñero et al., 2015). The disease-gene associations of HMDD2.0-Lan were downloaded from the SIDD database, containing 2,603 genes, 2,817 diseases, and 117,190 disease-gene associations (Cheng et al., 2013).

METHODS

The MDN-NMTF model aims to learn the representation vectors for miRNAs and diseases and to achieve better prediction for disease-related miRNAs. It can maximally maintain their features in original spaces, i.e., known miRNA-disease associations, miRNA similarity network structure, and disease similarity network structure. The preparing process of MDN-NMTF is broadly divided into four steps: building the networks, learning feature representation; reconstructing the miRNA-disease association network, predicting miRNA-disease associations (see **Figure 1**).

Disease Semantic Similarity

The disease semantic similarity between diseases is calculated using Mesh descriptors of diseases (Nelson et al., 2002). The disease terms can be represented as a direct acyclic graph (DAG), where nodes represent disease terms and edges represent the associations between diseases. The similarity between diseases can be calculated according to their common ancestors in the DAGs.

Let DAG_d represent disease d , $DAG_d = (T_d, E_d)$. T_d is the set composed of all parent disease nodes of d and itself, and E_d is the set of all edges between disease nodes within T_d . The formula to calculate the semantic value $DV_d(t)$ of diseases t and d is as follows:

$$DV_d(t) = \begin{cases} 1, & \text{if } t = d \\ \max \left\{ \Delta \times DV_d(t') \mid t' \in \text{children of } t \right\}, & \text{if } t \neq d \end{cases} \quad (1)$$

Where t is the set of all common ancestors of diseases d . Δ is the semantic contribution factor, whose value is between 0 and 1. We set the value of Δ to 0.5 in this study, similar to the values in (Yan et al., 2019). $DS(d)$ is the semantic values of a disease d in DAG.

$$DS(d) = \sum_{t \in T_d} DV_d(t) \quad (2)$$

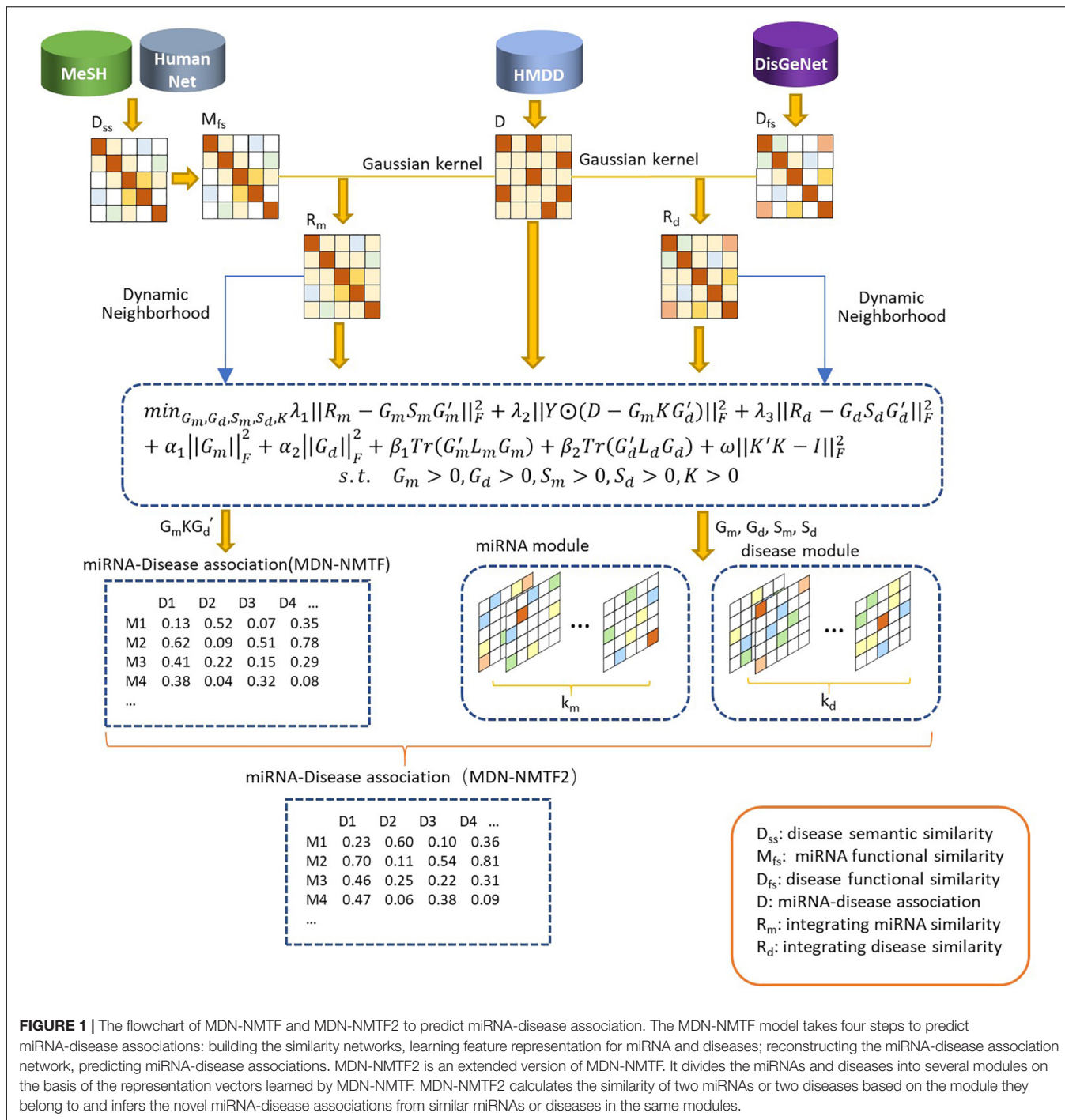


FIGURE 1 | The flowchart of MDN-NMTF and MDN-NMTF2 to predict miRNA-disease association. The MDN-NMTF model takes four steps to predict miRNA-disease associations: building the similarity networks, learning feature representation for miRNA and diseases; reconstructing the miRNA-disease association network, predicting miRNA-disease associations. MDN-NMTF2 is an extended version of MDN-NMTF. It divides the miRNAs and diseases into several modules on the basis of the representation vectors learned by MDN-NMTF. MDN-NMTF2 calculates the similarity of two miRNAs or two diseases based on the module they belong to and infers the novel miRNA-disease associations from similar miRNAs or diseases in the same modules.

The semantic similarity between disease d_i and disease d_j is as follows:

$$D_{ss}(d_i, d_j) = \frac{\sum_{t \in T_{d_i} \cap T_{d_j}} (DV_{d_i}(t) + DV_{d_j}(t))}{DS(d_i) + DS(d_j)} \quad (3)$$

(Cheng et al., 2014). Therefore, given a pair of diseases d_a and d_b , the functional similarity is defined as:

$$D_{fs}(d_a, d_b) = \frac{\sum_{1 \leq i \leq m} GFS_{G_b}(g_{ai}) + \sum_{1 \leq j \leq n} GFS_{G_a}(g_{bj})}{m+n} \quad (4)$$

Disease Functional Similarity

Calculating disease functional similarity is based on the assumption that similar diseases target similar disease genes

Where $G_a = \{g_{a1}, g_{a2}, \dots, g_{am}\}$ and $G_b = \{g_{b1}, g_{b2}, \dots, g_{bn}\}$ are two gene sets which associate with diseases d_a and d_b , respectively, and m and n are the numbers of genes in G_a and

G_b , respectively. $GFS_{G_b}(g_{ai})$ denotes the functional similarity between gene g_{ai} and the genes in G_b . It can be defined as below:

$$GFS_{G_b}(g_{ai}) = \max_{1 \leq j \leq n} (FS(g_{ai}, g_{bj})) \quad (5)$$

Where $FS(g_{ai}, g_{bj})$ denotes the functional similarity between gene g_{ai} and gene g_{bj} , which is obtained from HumanNet dataset (Lee et al., 2011) in this work. In the same way, the $GFS_{G_a}(g_{bj})$ can be computed.

MiRNA Functional Similarity

The miRNA functional similarity between two miRNAs m_1 and m_2 is calculated based on the semantic similarity of diseases to which they are related. It can be defined as follows:

$$M_{fs}(m_1, m_2) = \frac{\sum_{1 \leq i \leq n_1} MFS_{DT_2}(dt_{1i}) + \sum_{1 \leq j \leq n_2} MFS_{DT_1}(dt_{2j})}{n_1 + n_2} \quad (6)$$

Where n_1 and n_2 are the number of diseases that are associated with miRNAs m_1 and m_2 , respectively. DT_1 and DT_2 are the sets of diseases that are associated with miRNAs m_1 and m_2 , respectively. $MFS_{DT_2}(dt_{1i})$ is the semantic similarity of disease dt_{1i} and the diseases in DT_2 , which is defined as below:

$$MFS_{DT_2}(dt_{1i}) = \max_{1 \leq j \leq n_2} (D_{ss}(dt_{1i}, dt_{2j})) \quad (7)$$

$$MFS_{DT_1}(dt_{2j}) = \max_{1 \leq i \leq n_1} (D_{ss}(dt_{1i}, dt_{2j})) \quad (8)$$

Where $D_{ss}(dt_{1i}, dt_{2j})$ is the semantic similarity between diseases dt_{1i} and dt_{2j} .

GIP Kernel Similarity

It is observed that miRNAs with similar functions are more likely to be associated with similar diseases and vice versa. According to this observation, GIP kernel similarity is constructed to describe the miRNA similarity and disease similarity (Laarhoven et al., 2011). First, we defined a binary vector $IP(m_i)$ to represent the interaction profile of miRNA m_i by observing whether or not there is a known association between miRNA m_i and every disease. Then, the GIP similarity between miRNA m_i and m_j can be calculated as:

$$KM_{GIP}(m_i, m_j) = \exp(-\gamma_m \|IP(m_i) - IP(m_j)\|^2) \quad (9)$$

Where, γ_m controls the kernel bandwidth, which normalizes another bandwidth parameter γ_m' by the average number of related miRNAs per disease. γ_m is defined as follows:

$$\gamma_m = \gamma_m' / (\frac{1}{n_m} \sum_{i=1}^{n_m} \|IP(m_i)\|^2) \quad (10)$$

Here γ_m' is set to be 1 based on the previous study (Yan et al., 2019). n_m is the number of miRNAs.

Thus, the GIP kernel similarity between disease d_i and d_j is defined as follows:

$$KD_{GIP}(d_i, d_j) = \exp(-\gamma_d \|IP(d_i) - IP(d_j)\|^2) \quad (11)$$

$$\gamma_d = \gamma_d' / (\frac{1}{n_d} \sum_{i=1}^{n_d} \|IP(d_i)\|^2) \quad (12)$$

Where γ_d' is also set to 1 and n_d is the number of diseases.

Integrating Similarity for miRNAs and Diseases

Because not all miRNA-miRNA pairs have functional similarity, the GIP kernel similarity for miRNA is interpolated to the miRNA functional similarity to obtain the integrated similarity for miRNA. The final miRNA similarity matrix between miRNA m_i and miRNA m_j is defined as follows:

$$R_m(m_i, m_j) = \begin{cases} M_{fs}(m_i, m_j), & \text{if } M_{fs}(m_i, m_j) > 0 \\ KM_{GIP}(m_i, m_j), & \text{otherwise} \end{cases} \quad (13)$$

Similarly, the final disease similarity between disease d_i and disease d_j is defined as follows:

$$R_d(d_i, d_j) = \begin{cases} D_{fs}(d_i, d_j), & \text{if } D_{fs}(d_i, d_j) > 0 \\ KD_{GIP}(d_i, d_j), & \text{otherwise} \end{cases} \quad (14)$$

Regularized by Dynamic Neighborhood

Similar to previous DNRLMF-MDA method (Yan et al., 2019), we only preserve the relationships between a miRNA or a disease and their closest neighbors, when projecting the miRNA or disease to their latent spaces. Let $N(m_i)$ and $N(d_j)$ denote the set of nearest neighbors of miRNA m_i and disease d_j , respectively. The numbers of nearest neighbors of the miRNAs are not fixed but are dynamically determined according to Eq. (15). For miRNA m_i , $h(m_i)$ denotes its number of nearest neighbors, which can be as follows:

$$h(m_i) = \begin{cases} \max(H), & \text{if } \frac{1 - rs(m_i)_l}{l} \leq \epsilon^l, 1 \leq l \leq H \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

Where ϵ is the control parameter. It is set to 0.56 via cross-validation. The $rs(m_i)$ is a ranked vector based on the similarity between miRNA m_i and other miRNAs from high to low, and $rs(m_i)_l$ is the l th most similar value. H integer ranges from 1 to the total number of m_i 's neighbors and l (the exponent of ϵ) is a dynamic variable integer to satisfy the constraint. Similarly, for disease d_i , its number of nearest neighbors ($h(d_j)$) also can be formulated as follows:

$$h(d_j) = \begin{cases} \max(H), & \text{if } \frac{1 - rs(d_j)_l}{l} \leq \epsilon^l, 1 \leq l \leq H \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

Where $rs(d_j)$ is a ranked vector based on the similarities between disease d_j and other diseases from high to low, and $rs(d_j)_l$ is the l th most similar value.

Let matrix A be the dynamic nearest neighborhood matrix of miRNAs, its element $a_{i\mu}$ is calculated as below:

$$a_{i\mu} = \begin{cases} R_m(m_i, m_\mu), & \text{if } m_\mu \in N(m_i) \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

Similarity, let matrix B be the dynamic nearest neighborhood matrix of diseases, its element b_{jv} can be calculated as below:

$$b_{jv} = \begin{cases} R_d(d_j, d_v), & \text{if } d_v \in N(d_j) \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

In this work, we assume that if the two miRNAs or two diseases are the nearest neighbors in their original similarity networks, they should show similar representations in the corresponding latent spaces. Hence, the following two regularization terms are designed for miRNA and disease, respectively, which will be incorporated into the MDN-NMTF objective function. The regularization term for miRNAs can be defined as the following equation (Liu et al., 2016):

$$\sum_{i=1}^{n_m} \sum_{\mu=1}^{n_m} a_{i\mu} \|g_{m_i} - g_{m_\mu}\|_F^2 = \text{Tr}(G'_m L_m G_m) \quad (19)$$

Where $\text{Tr}()$ is the trace of a matrix, and $L_m = (D_m + D'_m) - (A + A')$, in which D_m and D'_m are the diagonal matrices, whose diagonal elements are $D_{mii} = \sum_{\mu=1}^{n_m} a_{i\mu}$ and $D'_{m\mu\mu} = \sum_{i=1}^{n_m} a_{i\mu}$, respectively. G_m represents the latent matrices of all miRNAs. Similarly, the regularization term for diseases can be defined as the following equation:

$$\sum_{j=1}^{n_d} \sum_{v=1}^{n_d} b_{jv} \|g_{d_j} - g_{d_v}\|_F^2 = \text{Tr}(G'_d L_d G_d) \quad (20)$$

Where $L_d = (D_d + D'_d) - (B + B')$, in which D_d and D'_d are the diagonal matrices, whose diagonal elements are $D_{djj} = \sum_{v=1}^{n_d} b_{jv}$ and $D'_{dvv} = \sum_{j=1}^{n_d} b_{jv}$, respectively. And G_d represents the latent matrices of all diseases.

The MDN-NMTF Model

Let $R_m \in \mathbb{R}^{n_m \times n_m}$ and $R_d \in \mathbb{R}^{n_d \times n_d}$ denote the adjacency matrix of the miRNA similarity network and disease similarity network, respectively. The latent matrices of all miRNAs and diseases are represented as $G_m \in \mathbb{R}^{n_m \times k_m}$ and $G_d \in \mathbb{R}^{n_d \times k_d}$, respectively. $K \in \mathbb{R}^{k_m \times k_d}$ denotes the association matrix between miRNA modules and disease modules. Let $D \in \mathbb{R}^{n_m \times n_d}$ be the matrix storing the known miRNA-disease associations. The MDN-NMTF learns the representation vectors for miRNAs (G_m) and disease (G_d) by optimizing the following objective function.

$$\begin{aligned} & \min_{G_m, G_d, S_m, S_d, K} \lambda_1 \|R_m - G_m S_m G'_m\|_F^2 \\ & + \lambda_2 \|Y \odot (D - G_m K G'_d)\|_F^2 + \lambda_3 \|R_d - G_d S_d G'_d\|_F^2 \\ & + \alpha_1 \|G_m\|_F^2 + \alpha_2 \|G_d\|_F^2 + \beta_1 \text{Tr}(G'_m L_m G_m) \\ & + \beta_2 \text{Tr}(G'_d L_d G_d) + \omega \|K'K - I\|_F^2 \\ \text{s.t. } & G_m > 0, G_d > 0, S_m > 0, S_d > 0, K > 0 \end{aligned} \quad (21)$$

In Eq. (21), the term of $\|R_m - G_m S_m G'_m\|_F^2$ captures the intrinsic module structure within the original miRNA similarity

matrix. Because the values in G_m record the modules the miRNAs belong to and S_m records the relationship of these modules. \odot is the Hadamard product. The term of $\|Y \odot (D - G_m K G'_d)\|_F^2$ indicates the miRNAs and diseases share similar relationship both in their original space and the latent space at the module level. We only want to use the known miRNA-disease information to learn their representation matrixes. Hence, let $Y \in \mathbb{R}^{n_m \times n_d}$ be a label weighted matrix [see Eq. (22)], where the elements of Y are set to 1 if the miRNA is known to associate with the disease. The elements of Y are set to 0.2 if the miRNA is known to not associate with the disease. Otherwise, the elements of Y are set to 0. Here, we set different weight for knowing to have or have no miRNA-disease associations. Because it is hard to prove that the miRNAs do not associate to certain diseases and some associations are temporarily not annotated due to the limitation of techniques.

$$Y(m_i, d_j) = \begin{cases} 1, & \text{if } D(m_i, d_j) \text{ is known and } D(m_i, d_j) = 1 \\ 0.2, & \text{if } D(m_i, d_j) \text{ is known and } D(m_i, d_j) = 0 \\ 0, & \text{if } D(m_i, d_j) \text{ is unknown} \end{cases} \quad (22)$$

The terms of $\text{Tr}(G'_m L_m G_m)$ [see Eq. (19)] and $\text{Tr}(G'_d L_d G_d)$ [see Eq. (20)] are used to preserve the network structure of the original miRNA similarity network and disease similarity network, respectively. We introduce L_m and L_d to represent the dynamic neighborhood of miRNAs and diseases, respectively, (see section “The MDN-NMTF model”). Two terms of $\|G_m\|_F^2$ and $\|G_d\|_F^2$ are adopted to penalize the magnitudes of the G_m and G_d for avoiding overfitting. $\|K'K - I\|_F^2$ is relaxed the constraint to $K'K = I$. λ_1, λ_2 , and λ_3 are balance parameters of matrix tri-factorization. α_1 and α_2 are regularization term parameters. β_1 and β_2 are the dynamic neighborhood regularization parameters. ω is the k -constraint parameter. In this work, the values of $k_m, k_d, \lambda_1, \lambda_2, \lambda_3, \alpha_1, \alpha_2, \beta_1, \beta_2$, and ω are set to 200, 200, 0.001, 5, 0.1, 0.2, 0.8, 90, 1.5, and 160, respectively (Supplementary Table 1).

Computation of S_m, S_d, G_m, G_d , and K

To obtain the optimal solution of S_m, S_d, G_m, G_d , and K in the objective function of MDN-NMTF model Eq. (21), we take the partial derivative of the objective function with respect to S_m, S_d, G_m, G_d , and K , respectively. Following the Karush-Kuhn-Tucker (KKT) condition for the non-negativity of S_m, S_d, G_m, G_d , and K and setting the partial derivative equal to zero, we can update S_m, S_d, G_m, G_d , and K as follows.

$$S_m \leftarrow S_m \odot \frac{G'_m R_m G_m}{G'_m G_m S_m G'_m G_m} \quad (23)$$

$$S_d \leftarrow S_d \odot \frac{G'_d R_d G_d}{G'_d G_d S_d G'_d G_d} \quad (24)$$

$$G_m \leftarrow G_m \odot \frac{2\lambda_1 R_m G_m S_m + \lambda_2 Y \odot D G'_d K' + \beta_1 A G_m}{2\lambda_1 G_m S_m G'_m G_m S_m + \lambda_2 Y \odot (G_m K G'_d) G'_d K' + \alpha_1 G_m + \beta_1 D_m G_m} \quad (25)$$

$$G_d \leftarrow G_d \odot \frac{2\lambda_3 R_d G_d S_d + \lambda_2 Y' \odot D' G'_m K + \beta_2 B G_d}{2\lambda_3 G_d S_d G'_d G_d S_d + \lambda_2 Y' \odot (G'_d K' G'_m) G'_m K + \alpha_2 G_d + \beta_2 D'_d G_d} \quad (26)$$

$$K \leftarrow K \odot \sqrt{\frac{\lambda_2 G'_m (Y \odot D) G_d + 2\omega K}{\lambda_2 G'_m (Y \odot (G_m K G'_d)) G_d + 2\omega K K' K}} \quad (27)$$

In this algorithm, \odot denotes the Hadamard product, and \div is entry-wise division for matrices. As shown in section “The MDN-NMTF model,” A and B are the dynamic nearest neighborhood matrix of miRNAs and diseases, respectively.

Predicting miRNA-Disease Associations

After getting the low-rank matrixes G_m , K , and G_d , we rebuild matrix DI by the produce of the matrixes G_m , K , and G_d ($DI = G_m K G'_d$) to predict miRNA-disease associations. The elements in DI denote the probability between miRNAs and diseases. Following is the pseudocode of MDN-NMTF algorithm.

Algorithm MDN-NMTF

Input: miRNA similarity R_m ; disease similarity R_d ; miRNA-disease association D ; the parameters $\lambda_1, \lambda_2, \lambda_3, \alpha_1, \alpha_2, \beta_1, \beta_2$, and ω .

Output: G_m, G_d, S_m, S_d, K , and DI

1: Initialize matrixes G_m, G_d, S_m, S_d , and K with random non-negative matrices, while S_m, S_d are symmetric matrices.

2: Calculate the dynamic neighbor matrices A and B by Eqs (17, 18), and then calculate the diagonal matrices D_m and D_d , and the Laplacian matrix L_m and L_d .

3: While objective function value in Eq. (21) not converge do

- (1) Fix G_m, G_d, S_d , and K and update S_m with Eq. (23).
- (2) Fix G_m, G_d, S_m , and K and update S_d with Eq. (24)
- (3) Fix G_d, S_m, S_d , and K and update G_m with Eq. (25).
- (4) Fix G_m, S_m, S_d , and K and update G_d with Eq. (26).
- (5) Fix G_m, G_d, S_m , and S_d and update K with Eq. (27).

end while

4: Rebuild miRNA-disease association matrix $DI = G_m K G'_d$.

In the third step of the while loop, each update iteration replaces the zero value in the matrices with 10^{-9} to guarantee the constraint condition in Eq. (21). The convergence condition is that the difference between two objective functions in the iteration is less than 10^{-6} or the number of iterations reaches the maximum number of iterations of 1,000.

Predicting miRNA-Disease Associations With Modular Information

At the same time, we also extend MDN-NMTF to a new version (called MDN-NMTF2, see Algorithm MDN-NMTF2) by using the modular information to improve the miRNA-disease association prediction ability of MDN-NMTF. Since the factorized matrices G_m and G_d obtained from MDN-NMTF record the modules the miRNAs or diseases belong to. MDN-NMTF2 utilizes the G_m and G_d values to partition the miRNAs and diseases into different models. Given $G_m \in \mathbb{R}^{n_m \times k_m}$ and $G_d \in$

$\mathbb{R}^{n_d \times k_d}$, there are k_m miRNA modules and k_d disease modules. The elements with relatively large values of each column of G_m (G_d) is assigned to the members of the corresponding module. We calculate the threshold for each miRNA (i.e., each row $g_m(i, \cdot)$ of G_m) with:

$$Th(i) = \mu(i) + t\sigma(i) \quad (28)$$

where $\mu(i) = \frac{1}{k_m} \sum_{k=1}^{k_m} g_m(i, k)$, $\sigma(i) =$

$$\sqrt{\frac{1}{k_m-1} \sum_{k=1}^{k_m} (g_m(i, k) - \mu(i))^2}$$

t is a given threshold. Based on this rule, we determined miRNA m_i as the k th module member if the entries of $g_m(i, k)$ are larger than $Th(i)$. In the same way, the threshold for each disease [each row $g_d(i, \cdot)$ of G_d] can be calculated. According to the settings of Ma et al. (2020), we also set $t = 1.5$ to identify miRNA and disease modules with proper resolution.

Then we calculate the similarity of two miRNAs based on the module they belong to. If two miRNAs m_i and m_j belong to the same miRNA module, their similarity in the k th miRNA module (ms_k) can be constructed as:

$$ms_k(m_i, m_j) = Sim(m_i, m_j), \text{ where } m_i \in ms_k, m_j \in ms_k \quad (29)$$

The $Sim(u, v)$ can be calculate as Eq.(30):

$$Sim(u, v) = \frac{\sum_{k=1}^{k_m} u_k v_k}{\sqrt{\sum_{k=1}^{k_m} u_k^2} \sqrt{\sum_{k=1}^{k_m} v_k^2}} \quad (30)$$

Here, k_m represents the dimension of the vectors u and v . u_k and v_k represent the k th element of the vectors u and v . Similarly, if two diseases d_i and d_j belong to the same disease module, we can construct their similarity in the k th disease module (ds_k) as

$$ds_k(d_i, d_j) = Sim(d_i, d_j), \text{ where } d_i \in ds_k, d_j \in ds_k \quad (31)$$

Based on the assumption that the miRNAs in the same modules are highly likely related to the same diseases, vice versa, we use $Scorem_k(m_i, d_j)$ to represent the correlation score between the disease d_j and the miRNA m_i in the k th miRNA module. It can be calculated according to Eq. (32):

$$Scorem_k(m_i, d_j) = \frac{\sum_{q=1}^{n_m} ms_k(m_i, m_q) D(q, j)}{\sum_{q=1}^{n_m} Sim(m_i, m_q)} \quad (32)$$

Where $D \in \mathbb{R}^{n_m \times n_d}$ is the matrix storing the known miRNA-disease associations. Thus, let D_m be the miRNA-disease associations that are predicted based on miRNA modules, which can be defined as:

$$D_m = \frac{1}{k_m} \sum_{i=1}^{k_m} Scorem_i \quad (33)$$

Similarly, we can get the correlation score ($Scored_k$) in the k th disease module and predict miRNA-disease associations (D_m) based on disease modules as follows:

$$Scored_k(m_i, d_j) = \frac{\sum_{q=1}^{n_d} ds_k(d_j, d_q) D(i, q)}{\sum_{q=1}^{n_d} Sim(d_j, d_q)} \quad (34)$$

$$D_d = \frac{1}{k_d} \sum_{j=1}^{k_d} \text{Score}_j \quad (35)$$

Here, k_m and k_d , as shown in the previous section, denote the number of miRNA modules and disease modules. The final predicted miRNA-disease associations of MDN-NMTF2 can be calculated by:

$$D2 = D1^* + \frac{D_m^*}{2} + \frac{D_d^*}{2} \quad (36)$$

Here, * denotes the Min-Max Normalization of the matrix. Following is the pseudocode of MDN-NMTF2.

Algorithm MDN-NMTF2

Input: miRNA similarity matrix R_m ; disease similarity matrix R_d and miRNA-disease association matrix D ; the parameters $\lambda_1, \lambda_2, \lambda_3, \alpha_1, \alpha_2, \beta_1, \beta_2$, and ω .

Output: $D2$

- 1: Get G_m, G_d, S_m, S_d, K , and $D1$ by the Algorithm MDN-NMTF
 - 2: Determine miRNA m_i as the k th module member if the entries of $g_m(i, k)$ are larger than $Th(i)$ [see Eq. (28)]
 3. Determine disease d_j as the k th module member if the entries of $g_d(i, k)$ are larger than $Th(i)$ [see Eq. (28)]
 4. Calculate similarity of two miRNAs If they belong to the same miRNA module by Eq. (29)
 5. Calculate D_m as the miRNA-disease associations that are predicted based on miRNA modules [see Eq. (33)]
 6. Calculate similarity of two diseases If they belong to the same disease module by Eq. (31)
 7. Calculate D_d as the miRNA-disease associations that are predicted based on disease modules [see Eq. (35)]
 - 8: Calculate the final predicted miRNA-disease associations $D2 = D1^* + D_m^*/2 + D_d^*/2$.
-

RESULTS

Performance Evaluation

To evaluate the performances of MDN-NMTF and MDN-NMTF2, we compared them with four state-of-the-art methods (DNRLMF-MDA, IMCMA, UBiRW, and GRNMF). The UBiRW uses an unbalanced Bi-Random walk on the heterogeneous network to propagate information and to infer potential miRNA-disease associations. DNRLMF-MDA, IMCMA, and GRNMF are three latest network embedding-based methods. DNRLMF-MDA adopts a dynamic neighborhood regularized logistic matrix factorization method to predict potential miRNA-disease associations. IMCMA uses Inductive matrix completion for miRNA-disease association prediction. GRNMF infer possible associations for all disease by a graph regularized non-negative matrix factorization framework. Considering there is no available interaction observed for new

diseases or miRNAs, GRNMF develops a preprocessing step to construct the miRNA-disease associations according to the neighbors' information. We implemented cross-validation under two different settings to evaluate the performance of the proposed methods. The two different settings are 5-fold randomly zeroing and single-column zeroing. For 5-fold randomly zeroing cross-validation, all the known miRNA-disease associations are randomly and equally divided into five non-overlapping parts. In each round, one of the five parts is for testing and the corresponding values in matrix D are cleared as 0, and the other four parts are as positive samples for training. Note that the miRNA and disease similarity network should be recalculated in each round. Single-column zeroing is to clear all miRNA-disease associations of a particular column of diseases and take them as testing data, others as training sets, and finally sum all AUCs to get the mean value. We repeat the cross-validation 20 times on four different datasets and show the average values in the following sections. For HMDD2.0-You, HMDD2.0-Lan, HMDD2.0-Yan and HMDD3.0 datasets, we select the way illustrated in section "Materials" to calculate the miRNA similarity network and disease similarity network.

To make the comparison fair, we tuned the parameters for every method to perform them the best in all of our experiments through randomly zeroing 5-fold cross-validation on HMDD2.0-You, HMDD2.0-Lan, and HMDD2.0-Yan. The detailed information, please see the online **Supplementary Files (Supplementary Table 2)**.

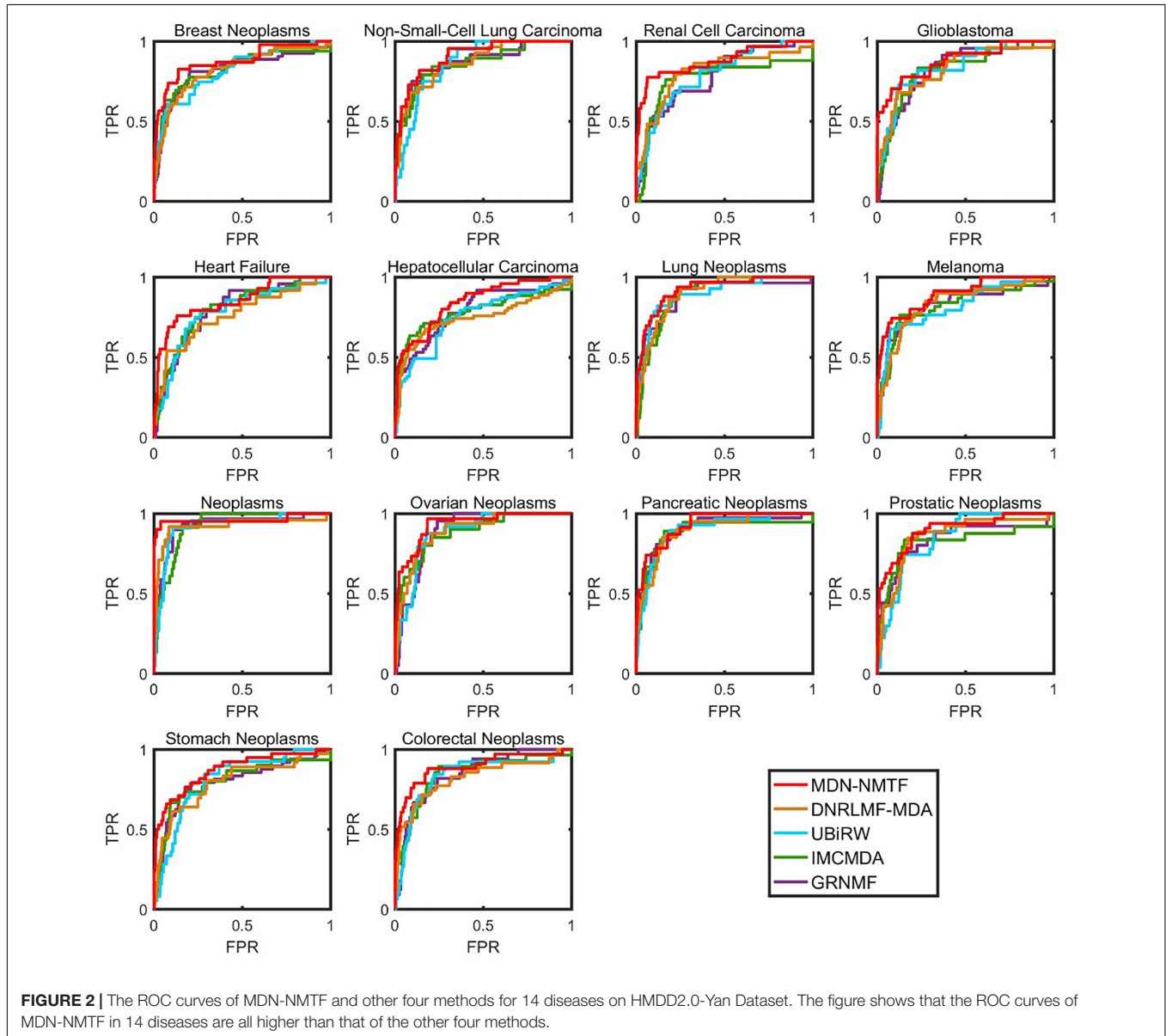
Randomly Zeroing Cross-Validation

As we can see from **Table 2**, MDN-NMTF and MDN-NMTF2 possess the highest two performance among the four methods on all the four datasets in terms of AUC values. On HMDD2.0-You dataset, compared with other methods (DNRLMF-MDA: 0.9301 ± 0.0036 , IMCMA: 0.8285 ± 0.0068 , UBiRW: 0.9196 ± 0.0036 , and GRNMF: 0.9031 ± 0.0049), the AUC values of MDN-NMTF and MDN-NMTF2 achieve 0.9335 ± 0.0037 and 0.9354 ± 0.0035 , respectively. On HMDD2.0-Yan dataset, the prediction performance of MDN-NMTF and MDN-NMTF2 are the best two because their AUC values are 0.9409 ± 0.0030 and 0.9424 ± 0.0033 , compared with other methods (DNRLMF-MDA: 0.9384 ± 0.0031 , IMCMA: 0.8045 ± 0.0062 , UBiRW: 0.9191 ± 0.0030 , GRNMF: 0.9153 ± 0.0045). On HMDD2.0-Lan dataset, the AUC values of MDN-NMTF and MDN-NMTF2 are 0.9391 ± 0.0033 and 0.9415 ± 0.0033 , which are both superior to the other results of DNRLMF-MDA (0.9369 ± 0.0030), IMCMA (0.7216 ± 0.0072), UBiRW (0.9198 ± 0.0032), and GRNMF (0.9157 ± 0.0044). On HMDD3.0 dataset, the AUC values of the MDN-NMTF and MDN-NMTF2 are 0.9435 ± 0.0021 and 0.9467 ± 0.0020 , which is better than that of DNRLMF-MDA method (0.9390 ± 0.0015), that of IMCMA method (0.6572 ± 0.0052), that of UBiRW method (0.9280 ± 0.0016) and that of GRNMF method (0.9247 ± 0.0023). We observe that DNRLMF-MDA leads to the highest performance among the four existing methods. It may be the DNRLMF-MDA method adopts a dynamic neighborhood regularized logistic matrix factorization method to

TABLE 2 | The AUC values for the models on different databases by randomly zeroing cross validation.

Methods	HMDD2.0-You	HMDD2.0-Yan	HMDD2.0-Lan	HMDD3.0
MDN-NMTF	0.9335 ± 0.0037	0.9409 ± 0.0030	0.9391 ± 0.0033	0.9435 ± 0.0021
MDN-NMTF2	0.9354 ± 0.0035	0.9424 ± 0.0033	0.9415 ± 0.0033	0.9467 ± 0.0020
DNRLMF-MDA	0.9301 ± 0.0036	0.9384 ± 0.0031	0.9369 ± 0.0030	0.9390 ± 0.0015
IMCMDA	0.8285 ± 0.0068	0.8045 ± 0.0062	0.7216 ± 0.0072	0.6572 ± 0.0052
UBiRW	0.9196 ± 0.0036	0.9191 ± 0.0030	0.9198 ± 0.0032	0.9280 ± 0.0016
GRNMF	0.9031 ± 0.0049	0.9153 ± 0.0045	0.9157 ± 0.0044	0.9247 ± 0.0023

It can be seen from the picture that the AUC values of our methods MDN-NMTF and MDN-NMTF2 on four datasets are both higher than the other four methods, and our extended method MDN-NMTF2 is a little higher than the original method MDN-NMTF. The highest AUC values of all methods in different datasets are in bold.



predict potential miRNA-disease associations. Both our MDN-NMTF and MDN-NMTF2 methods and the DNRLMF-MDA method utilize the dynamic neighborhood regularized restriction to construct the miRNA and disease feature vectors. We observe

that our MDN-NMTF and MDN-NMTF2 methods outperform DNRLMF-MDA. It can be partially attributed to the high quality of miRNA and disease features extracted by our MDN-NMTF and MDN-NMTF2 method from the heterogeneous network

TABLE 3 | AUC values of MDN-NMTF and other four compared methods for the 14 diseases on HMDD2.0-Yan Dataset.

Disease name	MDN-NMTF	DNRLMF-MDA	IMCMDA	UBiRW	GRNMF
Breast neoplasms	0.8732	0.8274	0.8340	0.8194	0.8183
Non-small-cell lung carcinoma	0.8989	0.8748	0.8658	0.8614	0.8582
Renal cell carcinoma	0.8699	0.8089	0.7936	0.7592	0.7846
Glioblastoma	0.8740	0.8280	0.8414	0.8248	0.8336
Heart failure	0.8528	0.7797	0.7864	0.7962	0.8120
Hepatocellular carcinoma	0.8408	0.7585	0.7588	0.7916	0.7846
Lung neoplasms	0.9207	0.9077	0.8963	0.8922	0.8885
Melanoma	0.8898	0.8375	0.8211	0.8216	0.8251
Neoplasms	0.9555	0.9253	0.9227	0.9223	0.9264
Ovarian neoplasms	0.9262	0.8941	0.8863	0.8835	0.8885
Pancreatic neoplasms	0.9274	0.9035	0.8943	0.8886	0.9057
Prostatic neoplasms	0.8919	0.8623	0.8395	0.8184	0.8261
Stomach neoplasms	0.8641	0.8054	0.8164	0.8055	0.8071
Colorectal neoplasms	0.8899	0.8292	0.8350	0.8463	0.8425

The table shows that the AUC values of MDN-NMTF in 14 diseases are all higher than that of the other four methods. The highest AUC values of all methods in different datasets are in bold.

under the consideration of the networks' module properties. MDN-NMTF2 employs the miRNA and disease features extracted by MDN-NMTF method to partition miRNA and disease modules. It infers potential miRNA-disease associations by considering the miRNAs' neighbors and diseases' neighbors in the same modules, which makes the MDN-NMTF2 method achieves a clear improvement than the MDN-NMTF method when predicting the missing miRNA-disease associations.

Besides, we test the performance of each method on 14 common diseases related to at least 110 miRNAs. **Figure 2** illustrates the Receiver operating characteristics curves of each method on the 14 disease. **Table 3** GRNMF lists the corresponding area under the curves (AUC). Both results show that MDN-NMTF outperforms the other four methods for all the 14 diseases.

Single-Column Zeroing Cross Validation

It still is a challenging task to infer miRNA associations for a new disease. To assess whether the MDN-NMTF and MDN-NMTF2 methods can successfully predict related miRNA for new diseases, we perform single-column zeroing cross-validation. **Table 4** lists the AUC values of different methods on four datasets. The AUC values of MDN-NMTF and MDN-NMTF2 still control the highest two in the four datasets. Compared to DNRLMF-MDA that has relatively better performance among the four existing methods, the MDN-NMTF method achieves 1.04% improvement on HMDD2.0-You dataset, 0.37% improvement on HMDD2.0-Yan dataset, 0.72% improvement on HMDD2.0-Lan dataset, and 1.18% improvement on HMDD3.0 dataset. The results prove that our methods considering the intrinsic module structure of miRNA and disease networks can extract the high quality of miRNA and disease features to predict related miRNAs for new diseases successfully. We observe that MDN-NMTF2 has a little lower performance than MDN-NMTF across the four datasets. It may be MDN-NMTF2 fails to infer the associations for new disease from the miRNA in the same modules.

Case Study

To further illustrate the performance of MDN-NMTF, we evaluate its miRNA prediction ability for some cancer types, such as Stomach Neoplasms (gastric Neoplasms) and Lymphoma. The dbDEMC database and miRCaner database are used as the benchmark datasets.

Among cancer-related deaths worldwide, Stomach Neoplasms ranks the third. Increasing evidence indicates that many miRNAs interact with Stomach Neoplasms by regulating the related genes of Stomach Neoplasms. **Table 5** demonstrates the top 50 predicted novel Stomach Neoplasms-related miRNAs predicted by MDN-NMTF on HMDD2.0-Yan dataset and the corresponding evidence. **Table 5** shows 35 of the 50 miRNAs are validated by dbDEMC database and miRCaner database. The remaining 15 miRNAs are all found to be related to human diseases in the literature. miR-181b modulates multidrug resistance by targeting BCL2 in human cancer cell lines (Zhu et al., 2010). MicroRNA-125b affects the proliferation of gastric cancer cells (Yang et al., 2013). miR-15b and miR-16 modulate multidrug resistance by targeting BCL2 in human gastric cancer cells (Xia et al., 2008). miR-101-2, miR-125b-2, and miR-451a act

TABLE 4 | The AUC values of each method on four different datasets by single-column zeroing cross validation.

Methods	HMDD2.0-You	HMDD2.0-Yan	HMDD2.0-Lan	HMDD3.0
MDN-NMTF	0.8570 ± 0.1223	0.8482 ± 0.1265	0.8445 ± 0.1339	0.8917 ± 0.1108
MDN-NMTF2	0.8561 ± 0.1240	0.8473 ± 0.1292	0.8447 ± 0.1342	0.8896 ± 0.1142
DNRLMF-MDA	0.8482 ± 0.1355	0.8451 ± 0.1431	0.8385 ± 0.1487	0.8813 ± 0.1181
IMCMDA	0.8329 ± 0.1297	0.8214 ± 0.1290	0.8158 ± 0.1357	0.8781 ± 0.1308
UBiRW	0.8512 ± 0.1343	0.8403 ± 0.1356	0.8326 ± 0.1499	0.8794 ± 0.1341
GRNMF	0.7833 ± 0.1505	0.7504 ± 0.1618	0.7895 ± 0.1465	0.8245 ± 0.1502

The table lists the AUC values of different methods on four different datasets. The AUC values of MDN-NMTF and MDN-NMTF2 are still the highest two. The results prove that our method considering the intrinsic module structure of miRNA and disease networks can improve the performance on miRNA-disease association predictions. The bold AUC values represent the highest AUC value of all method in the four different datasets.

TABLE 5 | Top 50 Related miRNAs of Stomach Neoplasms predicted by MDN-NMTF on HMDD2.0-Yan Dataset.

Top1-25miRNA	Evidence	Top26-50miRNA	Evidence
hsa-mir-21	dbDEMC, miRCancer	hsa-mir-199a-1	PMID:22956063
hsa-mir-214	dbDEMC, miRCancer	hsa-mir-22	dbDEMC, miRCancer
hsa-mir-200b	dbDEMC, miRCancer	hsa-mir-375	dbDEMC, miRCancer
hsa-mir-200c	miRCancer	hsa-mir-486	PMID:21415212
hsa-mir-182	dbDEMC, miRCancer	hsa-mir-106a	dbDEMC, miRCancer
hsa-mir-221	dbDEMC, miRCancer	hsa-mir-16-1	miRCancer
hsa-mir-181b-1	PMID:20162574	hsa-mir-222	dbDEMC, miRCancer
hsa-mir-148a	dbDEMC, miRCancer	hsa-mir-101-1	PMID:22450781
hsa-mir-34c	miRCancer	hsa-mir-10b	dbDEMC, miRCancer
hsa-mir-146b	miRCancer	hsa-mir-195	dbDEMC, miRCancer
hsa-mir-34a	dbDEMC, miRCancer	hsa-mir-141	dbDEMC, miRCancer
hsa-mir-125b-1	PMID:23128435	hsa-mir-101-2	PMID:26458815
hsa-mir-200a	dbDEMC, miRCancer	hsa-mir-146a	miRCancer
hsa-mir-31	dbDEMC, miRCancer	hsa-mir-199a-2	PMID:22956063
hsa-mir-145	dbDEMC, miRCancer	hsa-mir-106b	dbDEMC, miRCancer
hsa-mir-126	dbDEMC, miRCancer	hsa-mir-143	dbDEMC, miRCancer
hsa-mir-34b	miRCancer	hsa-mir-124-1	PMID:21365509
hsa-mir-16-2	PMID:18449891	hsa-mir-124-2	PMID:21365509
hsa-mir-125b-2	PMID:26458815	hsa-mir-103a-2	PMID:20726036
hsa-mir-107	dbDEMC, miRCancer	hsa-mir-130a	dbDEMC, miRCancer
hsa-mir-223	dbDEMC, miRCancer	hsa-mir-27b	dbDEMC, miRCancer
hsa-mir-183	dbDEMC, miRCancer	hsa-mir-155	dbDEMC, miRCancer
hsa-mir-27a	dbDEMC, miRCancer	hsa-mir-335	miRCancer
hsa-mir-25	dbDEMC, miRCancer	hsa-mir-151a	PMID:22956063
hsa-mir-181b-2	PMID:22539488	hsa-mir-7-1	PMID:22139078

From the table, 35 of the 50 miRNAs are validated by dbDEMC database and miRCancer database. The remaining 15 miRNAs are all found to be disease-related in the literature.

as potential tumor suppressors in primary GCs as well as in GC-derived AGS cells (Riquelme et al., 2016). MicroRNA-181b targets cAMP-responsive element-binding protein 1 in gastric adenocarcinomas (Chen L. et al., 2012). Plasma miRNA-199a-3p and miRNA-151-5p are significantly elevated ($p < 0.05$) and are significantly reduced after surgery ($p < 0.05$) in gastric cancer patients (Li et al., 2012). Genomic loss of miR-486 regulates tumor progression and the OLFM4 antiapoptotic factor

TABLE 6 | Top 50 Related miRNAs of Lymphoma predicted by MDN-NMTF on HMDD2.0-Yan Dataset.

Top1-25 miRNA	Evidence	Top26-50 miRNA	Evidence
hsa-mir-17	dbDEMC, miRCancer	hsa-mir-363	dbDEMC
hsa-mir-20a	dbDEMC, miRCancer	hsa-mir-150	dbDEMC, miRCancer
hsa-mir-155	dbDEMC, miRCancer	hsa-mir-126	dbDEMC
hsa-mir-18a	dbDEMC, miRCancer	hsa-mir-200b	dbDEMC
hsa-mir-19a	dbDEMC, miRCancer	hsa-mir-184	dbDEMC
hsa-mir-19b-1	miRCancer	hsa-mir-200a	dbDEMC
hsa-mir-92a-1	PMID:21383985	hsa-mir-499a	PMID:19690137
hsa-mir-15a	dbDEMC, miRCancer	hsa-mir-34a	dbDEMC
hsa-mir-146a	dbDEMC	hsa-mir-210	dbDEMC
hsa-mir-19b-2	miRCancer	hsa-mir-200c	dbDEMC
hsa-mir-16-1	miRCancer	hsa-mir-205	dbDEMC
hsa-mir-16-2	miRCancer	hsa-mir-145	dbDEMC
hsa-mir-21	dbDEMC, miRCancer	hsa-mir-24-1	PMID:19177201
hsa-mir-92a-2	PMID:21383985	hsa-mir-125b-1	dbDEMC
hsa-mir-181a-1	dbDEMC	hsa-mir-20b	dbDEMC
hsa-mir-181a-2	PMID:21910161	hsa-mir-125a	dbDEMC
hsa-mir-26a-2	dbDEMC	hsa-mir-124-1	PMID:22395483
hsa-mir-26a-1	PMID:19197161	hsa-mir-141	dbDEMC
hsa-mir-122	dbDEMC	hsa-mir-125b-2	PMID:23527180
hsa-mir-101-1	PMID:21960592	hsa-mir-18b	dbDEMC
hsa-mir-101-2	PMID:21960592	hsa-mir-138-2	dbDEMC
hsa-mir-342	dbDEMC	hsa-mir-29c	dbDEMC
hsa-mir-486	dbDEMC	hsa-mir-138-1	PMID:21960592
hsa-mir-203	dbDEMC	hsa-mir-708	dbDEMC
hsa-mir-223	dbDEMC, miRCancer	hsa-mir-143	dbDEMC

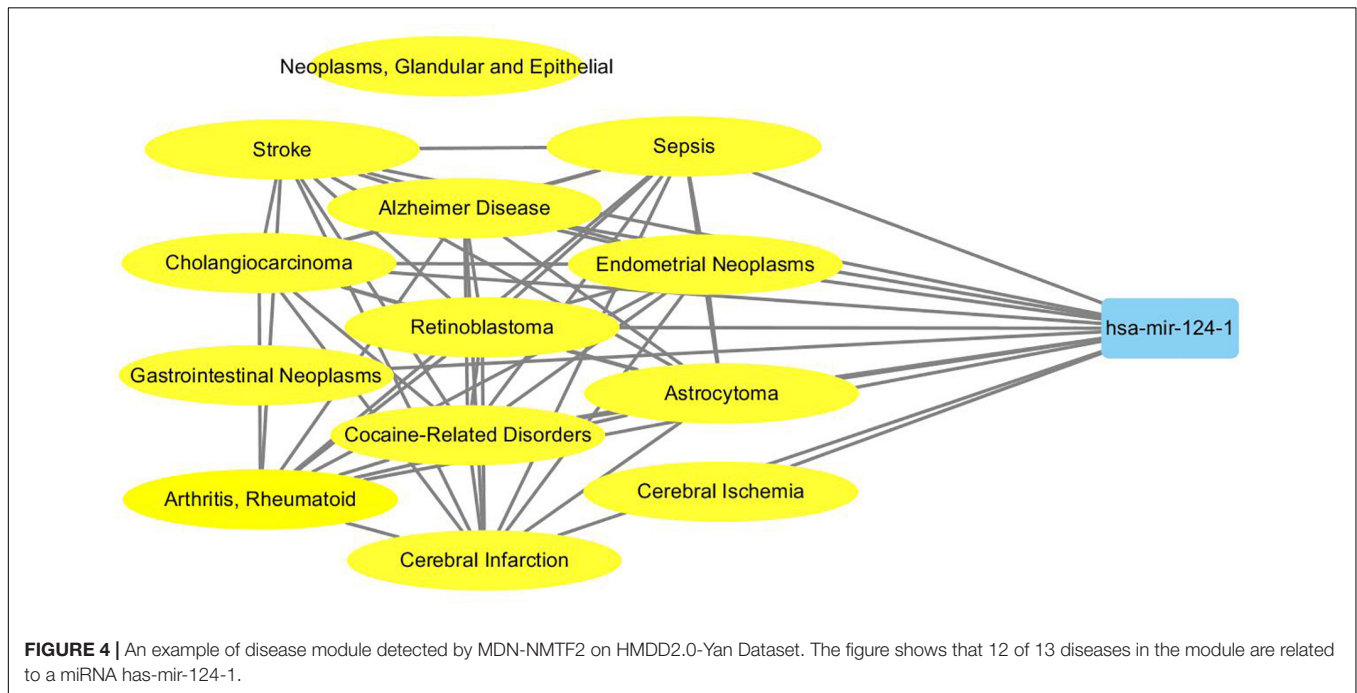
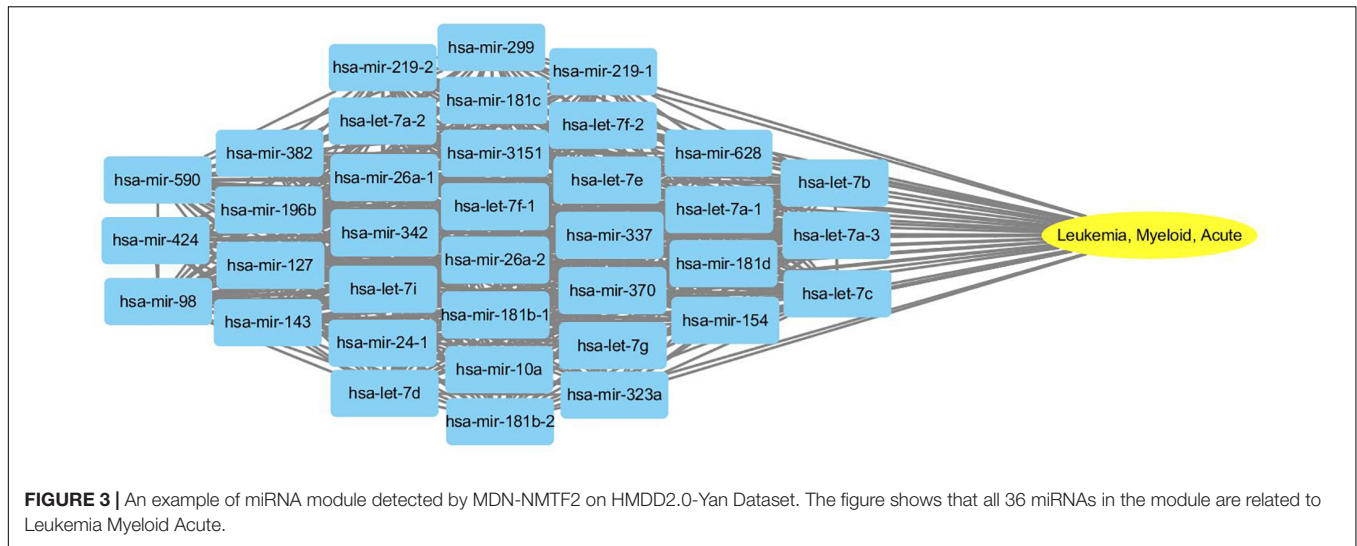
From the table, 39 of the 50 miRNAs are validated by dbDEMC database and miRCancer database. The remaining 11 miRNAs are all found to be disease-related in the literature.

TABLE 7 | miRNA modules and disease modules detected by MDN-NMTF2 on HMDD2.0-Yan Dataset.

Modules	NM	AvgSize	AvgSim	AvgPc
miRNA	127	40	0.4409	82.20%
disease	142	22	0.0939	61.28%

This table list the detailed information of the miRNA modules and disease modules detected by MDN-NMTF2 on the HMDD2.0-Yan dataset. Here, NM is the number of modules with more than one member. AvgSize is the average size of the modules. AvgSim is the average function similarity of the members in the modules. AvgPc is the average percentage of miRNAs (diseases) in the same modules share associations with common diseases (miRNA). It can be seen from the table that the average function similarity of the members in the miRNA modules was 0.4409, the values in the disease modules was 0.0939. Meanwhile, average 82% of miRNAs in the same module are related to the same disease and 61% of the disease in the same module is related to the same miRNA.

in gastric cancer (Oh et al., 2011). Lack of microRNA-101 causes E-cadherin functional deregulation through EZH2 upregulation in intestinal gastric cancer (Carvalho et al., 2012). Significant



associations are found between hypermethylation of the hsa-miR-124a and tumor size, differentiation, lymphatic metastasis, and invasion depth (Pei et al., 2011). miR-103, miR-21, miR-145, miR-106b, miR-146a, and miR-148a separate node-positive from node-negative gastric cancers (Tchernitsa et al., 2010). miR-7 is a novel mechanism by which the inflammatory response promotes gastric tumorigenesis (Kong et al., 2012).

Lymphoma is a type of cancer that begins in immune system cells. It is one of the top 10 deadly diseases. **Table 6** shows the result of top 50 Lymphoma-related miRNAs detected by MDN-NMTF on the HMDD2.0-Yan dataset. It shows that 39 of the 50 miRNAs are validated by dbDEMOC database and miRCancer database. The remaining 11 miRNAs are all found to be disease-related in the literature. The plasma miR-92a value

could be a novel biomarker not only for diagnosis but also for monitoring lymphoma patients after chemotherapy (Ohyashiki et al., 2011). Compared with healthy canine peripheral blood mononuclear cells and normal lymph nodes, miR-181a shows a decreased expression level (Uhl et al., 2011). miR-26a is repressed by MYC (Sander et al., 2009). The down-regulation of miR-16, miR-26a, miR-101, miR-29c, and miR138 in the *t*(14;18)-negative FL (follicular lymphoma) subset is associated with profound mRNA expression changes of potential target genes involving cell cycle control, apoptosis and B-cell differentiation. miR-16 targets CHEK1 showing increased expression on the protein level in *t*(14;18)-negative FL, while reducing TCL1A expression, in line with a partial loss of the germinal center B-cell phenotype in this FL subset (Leich et al., 2011). mir-499a

is deregulated hypermutations (Navarro et al., 2009). miR-24 is overexpressed (Gibcus et al., 2009). A distinct set of five microRNAs (miR-150, miR-550, miR-124a, miR-518b, and miR-539) is shown to be differentially expressed in gastritis as opposed to MALT lymphoma (Thorns et al., 2012). miR-125b-5p not only regulates tumor growth *in vivo* but also increases cellular resistance to proteasome inhibitors via modulation of MAD4 (Manfè et al., 2013).

Module Analysis

To probe why the modules help the MDN-NMTF2 to obtain better result, we analyze the miRNA or disease modules detected by MDN-NMTF2 on the dataset HMDD2.0-Yan (Supplementary Texts 3, 4). Table 7 lists the details of these modules. There are 127 miRNA modules with more than one member after removing the modules. The average size of these modules is 40. There are 142 disease modules with more than one member and their average size is 22. The average function similarity of the members in the miRNA modules was 0.4409, which was 113.20% higher than the average value of 0.2068 of the whole miRNA function similarity network (Supplementary Text 5). Similarly, the average function similarity of the disease modules was 0.0939, which was 160.11% higher than the average value of 0.0361 of the whole disease function similarity network (Supplementary Text 6). It suggests that the miRNA modules and disease modules detected by MDN-NMTF2 consist of members with similar functions. We also find that average 82% of miRNAs in the same module are related to the same disease (Supplementary Text 7). Figure 3 shows an example of miRNA module that consists of 36 miRNAs. All of these miRNAs are associated with Leukemia Myeloid Acute. On the other hand, 61% of the disease in the same module relate to the same miRNA (Supplementary Text 8). Figure 4 illustrates an example of disease module with 13 members. 12 of 13 diseases in the module relate to a common miRNA has-mir-124-1 that expresses in human embryonic stem cells. Hence, the MDN-NMTF2 infers miRNA-disease associations from miRNAs or diseases in the same modules, which helps it achieve better prediction results.

CONCLUSION

Inferring miRNA-disease associations is a crucial step to manifest principles of disease prevention, disease diagnosis and drug development. In this study, we have presented a novel method named MDN-NMTF to predict miRNA-disease associations. It constructs a heterogeneous network of disease similarity network, miRNA functional similarity network and a known miRNA-disease association network. After that, it learns the vector representation for miRNAs and diseases in the heterogeneous network by a matrix tri-factorization method under the constraint of the module structure and dynamic neighborhood. Finally, MDN-NMTF predicts novel miRNA-disease association probability by the product of the miRNA and disease latent vectors. At the same time, we also extend MDN-NMTF to a new version (called MDN-NMTF2) by using the modular information. Compared with the previous network

propagation-based method, like UBiRW, MDN-NMTF, and MDN-NMTF2 project miRNAs and diseases to a latent space. It can successfully integrate diverse biological information of miRNAs and diseases to predict miRNA-disease associations. Compared with the network embedding-based methods, like DNRLMF-MDA, IMCMDA and GRNMF, and MDN-NMTF and MDN-NMTF2 consider the module properties of miRNAs and diseases in the course of learning vector representation, which can maximally preserves the heterogeneous network structural information and the network properties. In particular, MDN-NMTF2 not only considers the modularity in the feature learning process but also uses the miRNA module and disease module information when reconstructing the miRNA-disease association matrix. We test our methods and the other four existing methods on four different datasets by implementing randomly zero cross-validation and single-column zero cross-validation. The results show that our methods outperform the state-of-the-art methods not only on predicting the missing miRNA-disease associations but also on recommending related miRNA for new diseases.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/weiba/MDN-NMTF>.

AUTHOR CONTRIBUTIONS

WP and JD obtained and analyzed miRNA-related data, disease-related data, and miRNA-disease associations. WP, JD, WD, and WL designed the new method MDN-NMTF and analyzed the results. WP and JD drafted the manuscript together. WP, JD, WD, and WL participated in revising the draft. All authors have read and approved the manuscript.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under grant Nos. 61972185 and 62072124, Natural Science Foundation of Yunnan Province of China (2019FA024), Yunnan Key Research and Development Program (2018IA054), and Yunnan Ten Thousand Talents Plan Young, Natural Science Foundation of Guangxi (2018GXNSFBA281193).

ACKNOWLEDGMENTS

This manuscript is recommended by the 5th Computational Bioinformatics Conference.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.603758/full#supplementary-material>

REFERENCES

- Ambros, V. (2003). MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* 114, 673–676. doi: 10.1016/s0092-8674(03)00428-8
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.
- Carvalho, J., Grieken, N. C. V., Pereira, P. M., Sousa, S., Tijssen, M., Buffart, T. E., et al. (2012). Lack of microRNA-101 causes E-cadherin functional deregulation through EZH2 up-regulation in intestinal gastric cancer. *J. Pathol.* 228, 31–44.
- Chen, L., Yang, Q., Kong, W.-Q., Liu, T., Liu, M., Li, X., et al. (2012). MicroRNA-181b targets cAMP responsive element binding protein 1 in gastric adenocarcinomas. *IUBMB Life* 64, 628–635. doi: 10.1002/iub.1030
- Chen, X., and Yan, G.-Y. (2014). Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* 4:5501.
- Chen, X., Liu, M.-X., and Yan, G.-Y. (2012). RWRMDA: predicting novel human microRNA-disease associations. *Mol. Biosyst.* 8, 2792–2798. doi: 10.1039/c2mb25180a
- Chen, X., Sun, L.-G., and Zhao, Y. (2020). NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief. Bioinform.* 22, 485–496. doi: 10.1093/bib/bbz159
- Chen, X., Wang, L., Qu, J., Guan, N.-N., and Li, J.-Q. (2018). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 34, 4256–4265.
- Cheng, L., Li, J., Ju, P., Peng, J., and Wang, Y. (2014). SemFunSim: a new method for measuring disease similarity by integrating semantic and gene functional association. *PLoS One* 9:e99415. doi: 10.1371/journal.pone.0099415
- Cheng, L., Wang, G., Li, J., Zhang, T., Xu, P., and Wang, Y. (2013). SIDD: a semantically integrated database towards a global view of human disease. *PLoS One* 8:e75504. doi: 10.1371/journal.pone.0075504
- Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105. doi: 10.1101/gr.082701.108
- Garzon, R., Marcucci, G., and Croce, C. M. (2010). Targeting microRNAs in cancer: rationale, strategies and challenges. *Nat. Rev. Drug Discov.* 9, 775–789. doi: 10.1038/nrd3179
- Gibcus, J. H., Tan, L. P., Harms, G., Schakel, R. N., Jong, D. D., Blokzijl, T., et al. (2009). Hodgkin lymphoma cell lines are characterized by a specific miRNA expression profile. *Neoplasia* 11, 167–176. doi: 10.1593/neo.08980
- Huang, Z., Liu, L., Gao, Y., Shi, J., and Zhou, Y. (2019). Benchmark of computational methods for predicting microRNA-disease associations. *Genome Biol.* 20:202.
- Kong, D., Piao, Y.-S., Yamashita, S., Oshima, H., Oguma, K., Fushida, S., et al. (2012). Inflammation-induced repression of tumor suppressor miR-7 in gastric tumor cells. *Oncogene* 31, 3949–3960. doi: 10.1038/onc.2011.558
- Laarhoven, T. V., 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
- Lan, W., Wang, J., Li, M., Liu, J., Wu, F.-X., and Pan, Y. (2016). Predicting MicroRNA-disease associations based on improved MicroRNA and disease similarities. *ACM Trans. Comput. Biol.* 15, 1774–1782. doi: 10.1109/tcbb.2016.2586190
- Lee, I., Blom, U. M., Wang, P. I., Shim, J. E., and Marcotte, E. M. (2011). Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* 21, 1109–1121. doi: 10.1101/gr.118992.110
- Leich, E., Zamo, A., Horn, H., Haralambieva, E., Puppe, B., Gascoyne, R. D., et al. (2011). MicroRNA profiles of t(14;18)-negative follicular lymphoma support a late germinal center B-cell phenotype. *Blood* 118, 5550–5558.
- Li, C., Li, J. F., Cai, Q., Qiu, Q. Q., Yan, M., Liu, B. Y., et al. (2012). miRNA-199a-3p in plasma as a potential diagnostic biomarker for gastric cancer. *Ann. Surg. Oncol.* 20, S397–S405.
- Li, J.-Q., Rong, Z.-H., Chen, X., Yan, G.-Y., and You, Z.-H. (2017). MCMDA: matrix completion for MiRNA-disease association prediction. *Oncotarget* 8, 21187–21199. doi: 10.18632/oncotarget.15061
- Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., et al. (2014). HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* 42, D1070–D1074.
- Liu, Y., Wu, M., Miao, C., Zhao, P., and Li, X.-L. (2016). Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput. Biol.* 12:e1004760. doi: 10.1371/journal.pcbi.1004760
- Liu, Y., Zeng, X., He, Z., and Zou, Q. (2017). Inferring MicroRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *ACM Trans. Comput. Biol. Bioinform.* 14, 905–915. doi: 10.1109/tcbb.2016.2550432
- Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., et al. (2008). An analysis of human MicroRNA and disease associations. *PLoS One* 3:e3420. doi: 10.1371/journal.pone.0003420
- Luo, J., and Xiao, Q. (2017). A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network. *J. Biomed. Inform.* 66, 194–203. doi: 10.1016/j.jbi.2017.01.008
- Ma, Y., Liu, G., Ma, Y., and Chen, Q. (2020). Integrative analysis for identifying co-modules of microbe-disease data by matrix tri-factorization with phylogenetic information. *Front. Genet.* 11:83.
- Manfè, V., Biskup, E., Willumsgaard, A., Skov, A. G., Palmieri, D., Gasparini, P., et al. (2013). cMyc/miR-125b-5p signalling determines sensitivity to bortezomib in preclinical model of cutaneous T-cell lymphomas. *PLoS One* 8:e59390. doi: 10.1371/journal.pone.0059390
- Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* 431, 343–349. doi: 10.1038/nature02873
- Navarro, A., Beà, S., Fernández, V., Prieto, M., Salaverría, I., Jares, P., et al. (2009). MicroRNA expression, chromosomal alterations, and immunoglobulin variable heavy chain hypermutations in Mantle cell lymphomas. *Cancer Res.* 69, 7071–7078. doi: 10.1158/0008-5472.can-09-1095
- Nelson, S. J., Aronson, A. R., Doszkocs, T. E., Chang, H. F., Mork, J., and McCray, A. T. (2002). Automated assignment of medical subject headings. *Proc. Amia Symp.* 6:1127.
- Oh, H.-K., Tan, A. L.-K., Das, K., Ooi, C.-H., Deng, N.-T., Tan, I. B., et al. (2011). Genomic loss of miR-486 regulates tumor progression and the OLFM4 antiapoptotic factor in gastric cancer. *Clin. Cancer Res.* 17, 2657–2667. doi: 10.1158/1078-0432.ccr-10-3152
- Ohyashiki, K., Umezumi, T., Yoshizawa, S.-I., Ito, Y., Ohyashiki, M., Kawashima, H., et al. (2011). Clinical impact of down-regulated plasma miR-92a levels in non-Hodgkin's lymphoma. *PLoS One* 6:e16408. doi: 10.1371/journal.pone.0016408
- Pei, L., Xia, J.-Z., Huang, H.-Y., Zhang, R.-R., Yao, L.-B., Zheng, L., et al. (2011). [Role of miR-124a methylation in patients with gastric cancer]. *Chin. J. Gastrointestinal Surgery* 14, 136–139.
- Peng, W., Lan, W., Yu, Z., Wang, J., and Pan, Y. (2016). A Framework for integrating multiple biological networks to predict microRNA-disease associations. *IEEE Trans. NanoBiosci.* 16, 100–107. doi: 10.1109/tnb.2016.2633276
- Peng, W., Lan, W., Zhong, J., Wang, J., and Pan, Y. (2017). A novel method of predicting microRNA-disease associations based on microRNA, disease, gene and environment factor networks. *Methods* 124, 69–77. doi: 10.1016/j.jmeth.2017.05.024
- Piñero, J., Queralt-Rosinach, N., Bravo, À, Deu-Pons, J., Bauer-Mehren, A., Baron, M., et al. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database J. Biol. Databases Curation* 2015:bav028. doi: 10.1093/database/bav028
- Riquelme, I., Tapia, O., Leal, P., Sandoval, A., Varga, M. G., Letelier, P., et al. (2016). miR-101-2, miR-125b-2 and miR-451a act as potential tumor suppressors in gastric cancer through regulation of the PI3K/AKT/mTOR pathway. *Cell Oncol.* 39, 23–33. doi: 10.1007/s13402-015-0247-3
- Sander, S., Bullinger, L., and Wirth, T. (2009). Repressing the repressor: a new mode of MYC action in lymphomagenesis. *Cell Cycle* 8, 556–559. doi: 10.4161/cc.8.4.7599
- Tchernitsa, O., Kasajima, A., Schäfer, R., Kuban, R.-J., Ungethüm, U., Györfy, B., et al. (2010). Systematic evaluation of the miRNA-ome and its downstream effects on mRNA expression identifies gastric cancer progression. *J. Pathol.* 222, 310–319. doi: 10.1002/path.2759
- Thorns, C., Kuba, J., Bernard, V., Senft, A., Szymczak, S., Feller, A. C., et al. (2012). Deregulation of a distinct set of microRNAs is associated with transformation of gastritis into MALT lymphoma. *Virchows Arch.* 460, 371–377. doi: 10.1007/s00428-012-1215-1

- Uhl, E., Krimer, P., Schliekelman, P., Tompkins, S. M., and Suter, S. (2011). Identification of altered MicroRNA expression in canine lymphoid cell lines and cases of B- and T-Cell lymphomas. *Genes Chromosomes Cancer* 50, 950–967. doi: 10.1002/gcc.20917
- Xia, L., Zhang, D., Du, R., Pan, Y., Zhao, L., Sun, S., et al. (2008). miR-15b and miR-16 modulate multidrug resistance by targeting BCL2 in human gastric cancer cells. *Int. J. Cancer* 123, 372–379. doi: 10.1002/ijc.23501
- Xiao, Q., Luo, J., Liang, C., Cai, J., and Ding, P. (2018). A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 34, 239–248. doi: 10.1093/bioinformatics/btx545
- Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., et al. (2013). Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One* 8:e70204. doi: 10.1371/journal.pone.0070204
- Xuan, P., Han, K., Guo, Y., Li, J., Li, X., Zhong, Y., et al. (2015). Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics* 31, 1805–1815. doi: 10.1093/bioinformatics/btv039
- Yan, C., Wang, J., Ni, P., Lan, W., Wu, F.-X., and Pan, Y. (2019). DNRLMF-MDA: predicting microRNA-disease associations based on similarities of microRNAs and diseases. *ACM Trans. Comput. Biol. Bioinform.* 16, 233–243. doi: 10.1109/tcbb.2017.2776101
- Yang, Z.-X., Lu, C.-Y., Yang, Y.-L., Dou, K.-F., and Tao, K.-S. (2013). MicroRNA-125b expression in gastric adenocarcinoma and its effect on the proliferation of gastric cancer cells. *Mol. Med. Rep.* 7, 229–232. doi: 10.3892/mmr.2012.1156
- 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
- Zeng, X., Zhang, X., Liao, Y., and Pan, L. (2016). Prediction and validation of association between microRNAs and diseases by multipath methods. *Biochim. Biophys. Acta Gen. Subj.* 1860, 2735–2739. doi: 10.1016/j.bbagen.2016.03.016
- Zhu, W., Shan, X., Wang, T., Shu, Y., and Liu, P. (2010). miR-181b modulates multidrug resistance by targeting BCL2 in human cancer cell lines. *Int. J. Cancer* 127, 2520–2529. doi: 10.1002/ijc.25260
- Zou, Q., Li, J., Song, L., Zeng, X., and Wang, G. (2016). Similarity computation strategies in the microRNA-disease network: a survey. *Brief. Funct. Genomics* 15, 55–64.

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 © 2021 Peng, Du, Dai and Lan. 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.