Bioinformatics, 36(8), 2020, 2538–2546
doi: 10.1093/bioinformatics/btz965
Advance Access Publication Date: 6 January 2020
Original Paper

Systems biology

Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction

Jin Li, Sai Zhang, Tao Liu, Chenxi Ning, Zhuoxuan Zhang and Wei Zhou*

School of Software, Yunnan University, Kunming 650091, China

*To whom correspondence should be addressed.

Associate Editor: Anthony Mathelier

Received on September 13, 2019; revised on December 17, 2019; editorial decision on December 20, 2019; accepted on December 31, 2019

Abstract

Motivation: Predicting the association between microRNAs (miRNAs) and diseases plays an import role in identifying human disease-related miRNAs. As identification of miRNA-disease associations via biological experiments is time-consuming and expensive, computational methods are currently used as effective complements to determine the potential associations between disease and miRNA.

Results: We present a novel method of neural inductive matrix completion with graph convolutional network (NIMCGCN) for predicting miRNA-disease association. NIMCGCN first uses graph convolutional networks to learn miRNA and disease latent feature representations from the miRNA and disease similarity networks. Then, learned features were input into a novel neural inductive matrix completion (NIMC) model to generate an association matrix completion. The parameters of NIMCGCN were learned based on the known miRNA-disease association data in a supervised end-to-end way. We compared the proposed method with other state-of-the-art methods. The area under the receiver operating characteristic curve results showed that our method is significantly superior to existing methods. Furthermore, 50, 47 and 48 of the top 50 predicted miRNAs for three high-risk human diseases, namely, colon cancer, lymphoma and kidney cancer, were verified using experimental literature. Finally, 100% prediction accuracy was achieved when breast cancer was used as a case study to evaluate the ability of NIMCGCN for predicting a new disease without any known related miRNAs.

Availability and implementation: https://github.com/ljatynu/NIMCGCN/

Contact: zwei@ynu.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

MicroRNAs (miRNAs) are a type of small endogenous non-coding RNAs (22 nucleotides in length), which play important roles in multiple biological processes. Dysfunction of miRNAs and their target mRNAs result in various diseases. Therefore, identification of disease-related miRNAs can contribute to the pathological study of diseases and disease biomarker detection (Chen *et al.*, 2017a, b, c, 2019a; Goh *et al.*, 2007). As identification of the associations between miRNAs and diseases using biological experiments is time-consuming and expensive, computational methods are being used to determine the potential associations between miRNAs and diseases (hereafter abbreviate MDA).

Existing computational methods are mainly summarized in the following two categories (Zeng *et al.*, 2016):

1. *Similarity measure-based prediction*: A basic assumption for this category approach is that functionally similar miRNAs are more likely to be associated with phenotypically similar diseases and vice versa. Chen *et al.* (2012) developed a random walk with restart for the miRNA-disease association (RWRMDA) to infer potential miRNA-disease interactions by implementing random walk on the miRNA function similarity network. Xuan *et al.* (2015) developed the MIDP based on random walk on the network. The network nodes were divided into labeled nodes and unlabeled nodes. The prior information of nodes can be completely utilized to improve MDA. You *et al.* (2017) proposed a path-based method called PBMDA that adopted a depth-first search algorithm for MDA. Zhang *et al.* (2019) developed a path prediction method based on Katz (Katz-SW, Katz-WL) by integrating miRNA family information, miRNA cluster information, experimentally valid miRNA target association and disease miRNA information. Chen et al. (2016, 2018b) explored the

global network similarity to capture the similarity relationship between diseases and miRNAs, respectively. Then, based on the consistency of diffusion profiles they got the miRNA-disease association scores. Chen et al. (2018a) proposed a novel information diffusion method based on network consistency for identifying disease-related miRNAs. Li et al. (2018) presented a label propagation model with linear neighborhood similarity to predict unobserved miRNA-disease associations. Recently, Chen et al. (2019) proposed a bipartite heterogeneous network link prediction method based on co-neighbor to predict miRNA-disease association.

2. *Machine learning-based predictions*: Over the past few years, many relationships between miRNAs and diseases have been identified using biological experiments, which provide opportunities for predicting the miRNA-disease association in a supervised machine learning manner. Chen *et al.* (2014) proposed a regularized least squares method for MDA, which uses a semi-supervised technique to identify the association between diseases and miRNAs. Luo et al. (2017) developed a collective prediction based on transduction learning to systematically prioritize miRNAs related to a disease. Li *et al.* (2017) proposed a Matrix Completion method for MDA (MCMDA) based on the known miRNA-disease associations. MCMDA utilized the matrix completion algorithm to update the adjacency matrix of known miRNA-disease associations for predicting the potential associations. Chen et al. (2017c) presented a computational model named Laplacian Regularized Sparse Subspace Learning for MDA. Chen et al. (2018b) developed a Matrix Decomposition and Heterogeneous Graph Inference (MDHGI) method for MDA. Chen et al. (2018c) proposed a novel computational model of bipartite network projection for MDA. Chen *et al.* (2018a) integrated different sources of information, such as miRNA functional similarity, disease semantic similarity and Gaussian interaction profile kernel similarity, into an inductive matrix completion model to infer miRNA-disease associations (IMCMDA). Cheng et al. (2019) developed a novel method to discover disease-related candidate miRNAs based on Adaptive Multi-View Multi-Label learning. Ding et al. (2019) replaced 0 in association matrix by the continuous value between 0 and 1 and proposed an improved inductive matrix completion method for association prediction. Furthermore, Chen et al. (2018d, 2019b, c) recently formulated a miRNA-disease association prediction as classification problems and proposed a series of effective prediction approaches. Experimental results show these approaches greatly improved the quality of association predictions. Recently, neural networks are also used to solve the problem of miRNA-disease association prediction. Xuan et al. (2018, 2019) presented two convolutional network-based methods for predicting candidate disease.

Despite the effectiveness of above-mentioned methods for MDA, there are still some limitations for current research results. On the one hand, the prediction qualities of similarity measure-based approaches were strongly limited by available linked information. Thus, these methods performed not very well on association predictions for new diseases or diseases with rare linked information. In addition, some useful information, such as diseases and miRNAs feature information, cannot be fully utilized to improve prediction accuracy for these methods. On the other hand, the feature representations of miRNAs and diseases and the prediction model are two critical issues for machine learning-based prediction approaches determining the quality of the results of MDA. However, both feature representation and prediction model must be reconsidered for further improving the prediction performance. First, existing approaches, for instance (Chen et al., 2018a; Ding et al., 2019; Luo et al., 2017), predicted miRNA-disease association ratings by inner products of the miRNA and the disease features projected onto a latent space. However, such a bilinear modeling of ratings simply combines the multiplication of latent features linearly and may not be enough to capture the complex and subtle interactions between the features of miRNAs and diseases. Second, classification-based approaches, such as Chen et al. (2018d, 2019b, c), were all faced with the challenge of effective negative samples selection strategies. Finally, existing neural networks-based approaches, such as Xuan et al. (2018, 2019), adopted convolutional networks to extract miRNAs and diseases feature representations. These approaches ignored the rich structural information contained in the miRNA and disease similarity networks and finally affected the quality of feature representations of the resulting miRNAs and diseases.

To overcome the mentioned limitations of current approaches for MDA, we proposed a novel method of Nonlinear Inductive Matrix Completion with Graph Convolutional Network (NIMCGCN) approach to address the problem of MDA in this study. The basic idea of our proposed method is as follows. First, we integrate information from different sources, such as the experimentally verified miRNA-disease associations, miRNA functional similarity, disease semantic similarity and Gaussian interaction profile kernel similarity, to construct a miRNA-disease heterogeneous network, which includes a miRNA functional similarity network, a disease semantic similarity network and a miRNA-disease association network. Second, miRNA and disease feature representations (embeddings) are learned from the constructed miRNA and disease similarity networks, respectively, using a graph convolutional network (GCN)-based approach. Third, to overcome the limitation of the bilinear rating model of IMC, we proposed a novel neural inductive matrix completion model, which replaced the feature projection matrix in IMC with a nonlinear neural network architecture that can learn an arbitrary function from data. This novel prediction framework significantly improved MDA.

2 Materials and methods

In this study, the interactions between miRNAs and diseases were captured using a heterogeneous network (HN), which consists of disease–disease semantic similarity, miRNA–miRNA functional similarity and the experimentally validated disease-miRNA associations. In this section, the methods of building a miRNA-disease HN based on different sources of information have been introduced.

2.1 The known human miRNA-disease associations

The known human miRNA-disease associations were obtained from the experimentally verified miRNA-disease database HMDD v2.0 (Li *et al.*, 2014). HMDD v2.0 contains 5430 experimentally verified human miRNA-disease associations. An adjacent matrix $\mathbf{T} \in \{0,\ 1\}^{m \times n}$ with 0-1 entries was constructed to represent the known miRNA-disease associations, where $\mathbf{T}(i,j) = 1$ if a miRNA $i \in \mathrm{M}$ is associated with a disease $j \in \mathrm{D}$. $\mathbf{T}(i,j) = 0$ if the association between a miRNA i and a disease j is unknown or unobserved.

2.2 MiRNA-miRNA similarity

miRNA similarity is measured using the miRNA's functional similarity score and the Gaussian interaction profile kernel similarity score. In particular, the miRNA functional similarity between a miRNA $i \in \mathrm{M}$ and a miRNA $j \in \mathrm{M}$ is defined as follows:

$$\mathrm{MS}(i,j) = \begin{cases} FS(i,j) & \text{the entry in MISIM database} \\ mGS(i,j) & \text{otherwise} \end{cases} \quad (1)$$

where $\mathrm{FS}(i,j)$ is the functional similarity scores download from the MISIM database: http://www.cuilab.cn/files/images/cuilab/misim.zip. $\mathrm{mGS}(i,j)$ is the Gaussian interaction profile kernel similarity score (Bandyopadhyay, 2010; Goh *et al.*, 2007; Lu *et al.*, 2008;

Wang *et al.*, 2010), which is used to supplement the missing entries in MISIM. Specifically, $\mathrm{mGS}(i,j)$ is calculated by follows

$$\mathrm{mGS}(i,j) = \exp\left(-\theta_{\mathrm{m}}\left|\left|\boldsymbol{T}[\boldsymbol{i},:] - \boldsymbol{T}[\boldsymbol{j},:]\right|\right|^2\right) \tag{2}$$

where $\mathbf{T}[\boldsymbol{i},:]$ represents the i-row in the adjacent matrix $\mathbf{T}$ and θ_{m} is the kernel bandwidth parameter which is calculated by the following formula

$$\theta_{\mathrm{m}} = \frac{1}{m}\sum_{i=1}^{m}\left|\left|\mathbf{T}[\boldsymbol{i},:]\right|\right|^2 \tag{3}$$

where m is the number of miRNAs, i.e. the row number of $\mathbf{T}$.

With $\mathrm{MS}(i,j)$, the miRNA functional similarity matrix is denoted by $\mathbf{A}_{\mathrm{m}} \in \mathbb{R}^{m\times m}$ and constructed by $[\boldsymbol{A}_m]_{ij} = \mathrm{MS}(i,j)$.

2.3 Disease–disease similarity

The MeSH database (http://www.ncbi.nlm.nih.gov/) is available for studying the relationship between different diseases. We obtained a hierarchical directed acyclic graph (DAG) directly from MeSH, where each node represents a disease and each directed edge in the DAG is from a general disease term to a specific disease term.

The semantic similarity scores between different diseases were calculated based on disease DAG. First, let $i \in \mathrm{D}$ be a disease. $\mathrm{dag}(i)$ indicates the node set, including node i and its ancestor nodes in the disease DAG. Then, the first semantic contribution of a disease $t \in \mathrm{D}$ to the disease i is denoted by $\mathrm{SC}_1(i,t)$ and can be formulated using the following equations (Chen *et al.*, 2018a),

$$\begin{cases} \mathrm{SC}_1(i,t) = 1 & \text{if } t = i \\ \mathrm{SC}_1(i,t) = \max\left\{\gamma\ \mathrm{SC}_1\left(i,t'\right) | t' \in \text{children of } t\right\} & \text{if } t \neq i \end{cases} \tag{4}$$

where γ is a semantic contribution decay factor, which shows that as the distances between disease t and its ancestor diseases increases, their contribution to the semantic value of disease d progressively decreases. γ was set as 0.5 according to previous literature (Wang *et al.*,2010).

Based on the definition of semantic contribution in Eq. (4), the first semantic similarity scores between different diseases, denoted by dS_1 was established. Let i,j be two different diseases. $dS_1(i,j)$ is defined as follows.

$$\mathrm{dS}_1(i,j) = \frac{\sum_{t\in\mathrm{dag}(i)\cap\mathrm{dag}(j)}\left(\mathrm{SC}_1(i,t)+\mathrm{SC}_1(j,t)\right)}{\sum_{t\in\mathrm{dag}(i)}\mathrm{SC}_1(i,t)+\sum_{t\in\mathrm{dag}(j)}\mathrm{SC}_1(j,t)} \tag{5}$$

Intuitively, $dS_1(i,j)$ is higher if the larger part of DAG is shared by i and j.

However, dS_1 ignores the significance of different disease contributions. Supposing that $i,t,q \in \mathrm{D}$, if disease t only appears in the $\mathrm{dag}(i)$, and q appears in both $\mathrm{dag}(i)$ and the dag of other diseases, t might have higher semantic contribution to i than q. Thus, the second semantic contribution score $\mathrm{SC}_2(i,t)$ was presented as follows:

$$SC_2(i,t) = -\log\left(\frac{\text{the number of dags including } t}{\text{the number of disease}}\right) \tag{6}$$

Based on $SC_2(i,t)$, the second semantic similarity score dS_2, between two diseases was presented as follows (Chen *et al.*, 2018a)

$$\mathrm{dS}_2(i,j) = \frac{\sum_{t\in\mathrm{dag}(i)\cap\mathrm{dag}(j)}\left(\mathrm{SC}_2(i,t)+\mathrm{SC}_2(j,t)\right)}{\sum_{t\in\mathrm{dag}(i)}\mathrm{SC}_2(i,t)+\sum_{t\in\mathrm{dag}(j)}\mathrm{SC}_2(j,t)} \tag{7}$$

As disease similarity measures calculated using dS_1 and dS_2 are both from the MeSH database, it provides only a part of the entries in diseases semantic similarity matrix. Hence, the Gaussian interaction profile kernel similarity was adopted to complement the remaining disease similarity entries.

Specifically, let $\mathbf{T} \in \{0,\ 1\}^{m\times n}$ be the adjacent matrix constructed using the known HMDD v2.0 miRNA-disease association data. $\mathbf{T}[:,\boldsymbol{j}]$ is the j-column binary vector representing disease j. Then, Gaussian interaction profile kernel similarity between disease i and disease j is defined as

$$\mathrm{dGS}(i,j) = \exp\left(-\theta_{\mathrm{d}}\left|\left|\boldsymbol{T}[:,\boldsymbol{i}] - \boldsymbol{T}[:,\boldsymbol{j}]\right|\right|^2\right) \tag{8}$$

where θ_{d} is the kernel bandwidth parameter calculated using the following formula

$$\theta_{\mathrm{d}} = \frac{1}{n}\sum_{i=1}^{n}\left|\left|\mathbf{T}[:,\boldsymbol{j}]\right|\right|^2 \tag{9}$$

where n is the number of diseases, i.e. the column number of $\mathbf{T}$.

With dS_1, dS_2 and dGS, the disease semantic similarity matrix is denoted by $\mathbf{A}_{\mathrm{d}} \in \mathbb{R}^{n\times n}$ and constructed using

$$[\boldsymbol{A}_d]_{ij} = \begin{cases} \dfrac{dS_1(i,j)+dS_2(i,j)}{2}, & \textit{if } i \textit{ and } j \textit{ has semantic similarity score} \\ dGS(i,j), & \textit{otherwise} \end{cases} \tag{10}$$

2.4 miRNA-disease heterogeneous information network

We combined the miRNA functional similarity network $\mathbf{A}_{\mathrm{m}}$, disease semantic similarity network $\mathbf{A}_{\mathrm{d}}$, and experimentally valid miRNA-disease interactions $\mathbf{T}$ to obtain the whole miRNA-disease heterogeneous information network as illustrated by Figure 1. Note that both the miRNA functional similarity network and the disease semantic similarity network are edge-weighted graphs.

In this study, based on the miRNA–miRNA similarity network, the disease–disease similarity network and the experimentally verified miRNA-disease data, a novel NIMCGCN method was presented to effectively solve the problem related to the prediction of miRNA-disease association.

2.5 Matrix completion and inductive matrix completion

A problem of miRNA-disease association prediction can be considered with m miRNAs and n diseases, and $m \times n$ experimentally verified miRNA-disease association matrix $\mathbf{T} \in \{0,\ 1\}^{m\times n}$. $\mathbf{T}(i,j) = 1$ if a miRNA i is associated with a disease j. $\mathbf{T}(i,j) = 0$ if the association between i and j is unknown or unobserved. Without loss of generality, Ω and $\bar{\Omega}$ were used to denote the set of observed and unobserved or unknown miRNA-disease entries from the known association matrix $\mathbf{T}$. The observation Ω consisted only of positive associations, i.e. if $\forall(i,j) \in \Omega$, $\mathbf{T}(i,j) = 1$. $\bar{\Omega}$ is the set of unknown or unobserved entries if $\forall(i,j) \in \bar{\Omega}$, $\mathbf{T}(i,j) = 0$. A sample of observed entries Ω from a true underlying matrix $\mathbf{Q}$ was considered. The objective was to estimate missing entries under some additional assumptions on the structure of the association matrix $\mathbf{T}$. The most common assumption is that $\mathbf{Q}$ is low rank, i.e. $\mathbf{Q} = \mathbf{F}\mathbf{G}^{\mathrm{T}}$, where $\mathbf{F} \in \mathbb{R}^{m\times k}$ and $\mathbf{G} \in \mathbb{R}^{n\times k}$ are of rank $k \ll m,n$. With these notations, the basic MDA can be formulated as the following matrix completion problems:

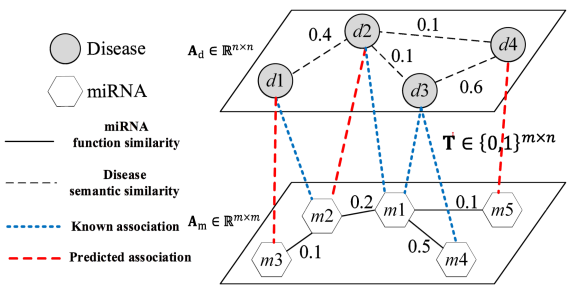

Fig. 1. An Illustration of a miRNAs-diseases heterogeneous information network

$$\min_{\mathbf{F},\mathbf{G}} \frac{1}{2}\left|\left|P_{\Omega}(\mathbf{T}-\mathbf{F}\mathbf{G}^{\mathrm{T}})\right|\right|_F^2+\lambda\left(||\mathbf{F}||_F^2+||\mathbf{G}||_F^2\right) \quad (11)$$

where $P_{\Omega}(\cdot)$ is the projection of the matrix onto the set Ω.

One limitation of the classic matrix completion is that it cannot directly use side information (such as the feature representations of miRNAs and diseases) to predict new and unseen data, or to simplify computation. Inductive matrix completion (IMC) (Natarajan *et al.*, 2014) was proposed to circumvent this limitation. When MDA was solved using IMC, an association predicted rating was modeled as an inner product of the features of a miRNA and a disease projected onto a latent space. IMC assumes that the association matrix is generated by applying feature vectors associated with its row as well as column entities to a projection matrix $\mathbf{Z}$. The goal was to recover $\mathbf{Z}$ using observations from $\mathbf{T}$. Furthermore, to learn parameters effectively from a small number of observed ratings, the latent space was constrained to be low-dimensional, which implies that the parameter matrix is constrained to be low rank.

Specifically, let $\mathbf{x}_i \in \mathbb{R}^{f_m}$ be the feature vector with a dimension f_m for a miRNA i, and $\mathbf{X} \in \mathbb{R}^{m\times f_m}$ be the feature matrix for training miRNAs. Similarly, $\mathbf{y}_j \in \mathbb{R}^{f_d}$ is the feature matrix for a disease j and $\mathbf{Y} \in \mathbb{R}^{n\times f_d}$ is the feature matrix for diseases. The IMC can recover a feature projection matrix $\mathbf{Z} \in \mathbb{R}^{f_m\times f_d}$ using the observed entries from the known miRNA-disease association matrix $\mathbf{T}$ and the feature matrix of $\mathbf{X}$ and $\mathbf{Y}$. To learn the parameters effectively from a small number of observed ratings, the latent space was constrained to be low-dimensional, which implies that the feature projection matrix $\mathbf{Z}$ is constrained to be low rank, i.e. $\mathbf{Z}=\mathbf{Z}_1\mathbf{Z}_2^{\mathrm{T}}$, where $\mathbf{Z}_1 \in \mathbb{R}^{f_m\times k}$ and $\mathbf{Z}_2 \in \mathbb{R}^{f_d\times k}$ are of rank $k \ll f_m, f_d$. With these notions, the IMC for MDA was formulated as the following optimization problem.

$$\min_{\mathbf{Z}_1,\ \mathbf{Z}_2}\left|\left|P_{\Omega}\left(\mathbf{T}-\mathbf{X}\mathbf{Z}_1\mathbf{Z}_2^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\right)\right|\right|_F^2+\left|\left|P_{\bar{\Omega}}\ \left(\mathbf{T}-\mathbf{X}\mathbf{Z}_1\mathbf{Z}_2^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}\right)\right|\right|_F^2 +\lambda\left(||\mathbf{Z}_1||_F^2+||\mathbf{Z}_2||_F^2\right) \quad (12)$$

IMC defined in Eq. (12) predicts a miRNA-disease association rating by an inner product of the features of a miRNA and a disease projected onto a latent space. However, such a bilinear modeling of the ratings simply combines the multiplication of latent features linearly and may not be enough to capture complex and subtle interactions between the features of miRNAs and diseases. To overcome this limitation, we proposed a novel NIMC for miRNA-disease association prediction in this study, where the low-rank feature projection matrix $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are replaced with a nonlinear fully connected layer, which induces nonlinear transformations on top of the miRNA or disease feature representations.

2.6 Neural inductive matrix completion

A classical IMC for MDA can be considered with the feature matrices $\mathbf{X} \in \mathbb{R}^{m\times f_{\mathrm{m}}}$ for miRNAs and $\mathbf{Y} \in \mathbb{R}^{n\times f_d}$ for diseases as inputs of IMC. $\mathbf{Z}_1 \in \mathbb{R}^{f_{\mathrm{m}}\times k}$ and $\mathbf{Z}_2 \in \mathbb{R}^{f_d\times k}$ are of rank k decompositions of the projection matrix $\mathbf{Z}$. In our proposed neural inductive matrix completion (NIMC), the bilinear rating model $\mathbf{X}\mathbf{Z}_1\mathbf{Z}_2^{\mathrm{T}}\mathbf{Y}^{\mathrm{T}}$ in IMC was replaced with a novel method of Nonlinear neural rating model to capture the complex and subtle interactions between the features of miRNAs and diseases.

In NIMC, the feature matrices $\mathbf{X}$ for miRNAs and $\mathbf{Y}$ for diseases are encoded separately by two different nonlinear fully connected layers. In the following sections, the encode process for $\mathbf{X}$ and $\mathbf{Y}$ have been described. $\mathbf{W}_{\mathrm{m}}^{(l)}$ $(l \in \{1,\ 2,\ 3,\ldots\})$ denotes the weight matrix of the l-th layer in the nonlinear fully connected layer, and $\mathbf{X}^l \in \mathbb{R}^{f_{\mathrm{m}}^l\times f_{\mathrm{m}}^{l+1}}$ denotes the input miRNAs' feature matrix of the l-th layer, where f_{m}^l and f_{m}^{l+1} are the input and output dimensions for l-th layer. The total nonlinear transformations of the fully connected layers are described as follows

$$\phi_{\mathrm{m}}^l(\boldsymbol{X})=\mathrm{relu}\left(\mathbf{W}_{\mathrm{m}}^{(l)}\mathrm{relu}\left(\cdots\mathrm{relu}\left(\boldsymbol{W}_m^{(1)}\boldsymbol{X}+\boldsymbol{b}_m^1\right)\cdots\right)+\mathbf{b}_{\mathrm{m}}^l\right) \quad (13)$$

where $\mathbf{b}_{\boldsymbol{m}}^{\boldsymbol{l}}$ is the bias item for the l-th layer and $\mathrm{relu}(\cdot)$ is the rectified linear unit nonlinear activation function.

Similarly, the total nonlinear transformations of fully connected layers for diseases are described as follows

$$\phi_{\mathrm{d}}^l(\boldsymbol{Y})=\mathrm{relu}\left(\mathbf{W}_{\mathrm{d}}^{(l)}\mathrm{relu}\left(\cdots\mathrm{relu}\left(\boldsymbol{W}_d^{(1)}\boldsymbol{Y}+\boldsymbol{b}_d^1\right)\cdots\right)+\mathbf{b}_{\mathrm{d}}^l\right) \quad (14)$$

Let $\boldsymbol{\Psi}_{\mathrm{m}}=\left\{\boldsymbol{W}_m^{(1)},\ldots,\boldsymbol{W}_m^{(l)},\boldsymbol{b}_{\boldsymbol{m}}^{(1)},\ldots,\boldsymbol{b}_{\boldsymbol{m}}^{(l)}\right\}$ and $\boldsymbol{\Psi}_{\mathrm{d}}=\left\{\boldsymbol{W}_d^{(1)},\ldots,\boldsymbol{W}_d^{(l)},\boldsymbol{b}_{\boldsymbol{d}}^{(1)},\ldots,\boldsymbol{b}_{\boldsymbol{d}}^{(l)}\right\}$ be the parameters involved in Eq. (13) and Eq. (14), respectively. Therefore, with the above notations, a neural inductive matrix completion model is defined by the following formulation.

$$\min_{\boldsymbol{\Psi}_{\mathrm{m}},\ \boldsymbol{\Psi}_{\mathrm{d}}}\left|\left|P_{\Omega}\left(\boldsymbol{T}-\phi_m^l(\boldsymbol{X})\boldsymbol{\phi}_{\boldsymbol{d}}^{\boldsymbol{l}}(\boldsymbol{Y})^T\right)\right|\right|_F^2 +\left|\left|P_{\bar{\Omega}}\ \left(\boldsymbol{T}-\phi_m^l(\boldsymbol{X})\boldsymbol{\phi}_{\boldsymbol{d}}^{\boldsymbol{l}}(\boldsymbol{Y})^T\right)\right|\right|_F^2+\lambda\left(||\boldsymbol{\Psi}_m||^2+||\boldsymbol{\Psi}_d||^2\right) \quad (15)$$

As the formulation (15) may yield degenerate results, the following biased neural inductive matrix completion formulation was proposed

$$\min_{\boldsymbol{\Psi}_{\mathrm{m}},\ \boldsymbol{\Psi}_{\mathrm{d}}}\frac{(1-\alpha)}{2}\left|\left|P_{\Omega}\left(\mathbf{T}-\phi_{\mathrm{m}}^l(\mathbf{X})\boldsymbol{\phi}_{\mathbf{d}}^{\boldsymbol{l}}(\mathbf{Y})^{\mathrm{T}}\right)\right|\right|_F^2 +\frac{\alpha}{2}\left|\left|P_{\bar{\Omega}}\ \left(\mathbf{T}-\phi_{\mathrm{m}}^l(\mathbf{X})\boldsymbol{\phi}_{\mathbf{d}}^{\boldsymbol{l}}(\mathbf{Y})^{\mathrm{T}}\right)\right|\right|_F^2+\lambda\left(||\boldsymbol{\Psi}_{\mathrm{m}}||^2+||\boldsymbol{\Psi}_{\mathrm{d}}||^2\right) \quad (16)$$

where the parameter $\alpha \in (0,\ 1)$ is the bias item that appropriately weighs observed and unobserved entries.

It is worth noting that the classical IMC is actually a special case of our proposed NIMC when the number of the layer is set to $l=1$, the nonlinear activation functions are removed, and the weight matrix is $\mathbf{Z}_1=\mathbf{W}_m^{(1)} \in \mathbb{R}^{f_m\times k}$ and $\mathbf{Z}_2=\mathbf{W}_d^{(1)} \in \mathbb{R}^{f_d\times k}$. Therefore, the capability of feature extraction of NIMC is theoretically not worse than that of IMC. In fact, experimental results confirmed that the feature extraction ability of NIMC is significantly better than that of IMC.

When a NIMC method was used to solve an MDA problem, some questions must be answered. For example, what are the appropriate feature representations for miRNAs and diseases? How can the feature matrices $\mathbf{X}$ and $\mathbf{Y}$ be obtained?

A basic assumption for miRNA-disease association prediction is that functionally similar miRNAs are more likely to be associated with phenotypically similar diseases, and vice versa. Therefore, miRNA functional similarity information and disease semantic similarity information are both crucial for effectively predicting the association between a miRNA and a disease. Actually, miRNA functional similarity information can be formulated by a miRNA functional similarity network where the nodes represent miRNAs and the score with an edge describes the functional similarity degree between two miRNAs. Similarly, there are disease semantic similarity networks.

In the following section, a feature representation learning method to obtain the feature matrices $\mathbf{X}$ for miRNAs and $\mathbf{Y}$ for diseases has been proposed. Our method learns the embeddings with GCNs, which completely leverage the structure information encoded in the similarity networks.

2.7 Learning embeddings with graph convolutional networks

Certain recent studies (Su et al., 2018; Zhang et al., 2018) have attempted to automatically learn network topology-preserving node-level vector representations (embedding) from networks. In particular, the GCN and its variants (Defferrard *et al.*, 2016; Kipf *et al.*, 2017) have significantly improved many networks-related prediction tasks, such as predicting the biological activities of small molecules and recommendation.

In this study, GCNs on a miRNA functional similarity network and a disease semantic similarity network were leveraged to learn miRNA and disease latent embedding via supervised learning. These learned embeddings will be used as the input for the downstream NIMC-based rating model to make a final association prediction.

Specifically, let G_m and G_d be the miRNA functional similarity network and disease semantic similarity network, respectively. $\mathbf{A}_m$ denotes the adjacent matrix for G_m and $\mathbf{A}_d$ for G_d. $V_m = m$ and $V_d = n$ denote the size of the node set V_m over G_m and V_d over G_d, respectively. Let $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ be scalar features on the set of nodes of the graph G_m and G_d, respectively. In graph signal processing literature, $\mathbf{x}$ and $\mathbf{y}$ are the graph signals over G_m and G_d, where $\mathbf{x}[i] \in \mathcal{R}$ $(i \in \{1, 2, \ldots, m\})$ is a scalar feature value of the node $i \in V_m$. Next, we introduced the method of learning embeddings for miRNAs over G_m. The way of learning embeddings for diseases over G_d is a similar process.

A graph signal $\mathbf{x}^{(t-1)}$ over G_m at step $t-1$ is transformed into a new graph signal $\mathbf{x}^{(t)}$ using the following defined graph Fourier transform (GFT)

$$\mathbf{x}^{(t)} = \mathbf{D}_m^{-1/2}\mathbf{A}_m\mathbf{D}_m^{-1/2}\mathbf{x}^{(t-1)} \tag{17}$$

where $\mathbf{L}_m = \mathbf{D}_m^{-1/2}\mathbf{A}_m\mathbf{D}_m^{-1/2}$ is the symmetric normalized Laplacian matrix of G_m and $\mathbf{D}_m$ is a diagonal matrix with diagonal entry$[\boldsymbol{D}_m]_{ij} = \sum_j [\boldsymbol{A}_m]_{ij}$. As $\mathbf{L}_m$ is a symmetric matrix, it can be eigen-decomposed as $\mathbf{L}_m = \mathbf{U}_m\mathbf{\Lambda}_m\mathbf{U}_m^{\mathrm{T}}$, where $\mathbf{U}_m$ is the corresponding eigen-vector matrix and $\mathbf{\Lambda}_m = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is the eigen-value matrix. Thus, Eq. (17) is reformulated as

$$\mathbf{x}^{(t)} = \boldsymbol{U}_m\boldsymbol{\Lambda}_m\boldsymbol{U}_m^T\mathbf{x}^{(t-1)} \tag{18}$$

In Eq. (18), the graph signal $\mathbf{x}^{(t-1)}$ has been transformed into the spectral domain by $\hat{\mathbf{x}}^{(t-1)} = \mathbf{U}_m^{\mathrm{T}}\mathbf{x}^{(t-1)}$. $\mathbf{\Lambda}_m$ represents a graph signal filter, which is fixed and determined by G_m. The values in the diagonal matrix $\mathbf{\Lambda}_m$ specify the signal scaling factors and determine the importance of each frequency domain. After filtering by the filter, the graph signal $\hat{\mathbf{x}}^{(t-1)}$ can be transformed inversely into the graph domain using $\mathbf{U}_m\hat{\mathbf{x}}^{(t-1)}$.

It is beneficial to parameterize the filter $\mathbf{\Lambda}_m$ in Eq. (18). Thereby, GFT can dynamically adjust the importance of each frequency domain for graph signal transformations. The advantage of the parametric filter is that we can learn a desirable graph filter in a supervised learning way. Thus, a graph signal $\mathbf{x}^{(t-1)}$ is transformed by a parameterized filter $g_{\boldsymbol{\Theta}}(\mathbf{\Lambda}_m) = \mathrm{diag}(\theta_1\lambda_1, \theta_2\lambda_2, \ldots, \theta_m\lambda_m)$ into $\mathbf{x}^{(t)}$according to the following equation

$$\mathbf{x}^{(t)} = \boldsymbol{U}_m g_{\boldsymbol{\Theta}}(\boldsymbol{\Lambda}_m)\boldsymbol{U}_m^T\mathbf{x}^{(t-1)} \tag{19}$$

However, the GFT defined by Eq. (19) still has two limitations. First, GFT must solve the eigen-decomposition for G_m to obtain the eigen-vector matrix and the eigen-value matrix, which can be expensive for large networks. Second, a graph signal $\mathbf{x}$ over G_m is still a vector, which means that each miRNA is represented by a scalar feature. However, a vector feature for every miRNA is necessary to model the subtle and complex interactions between miRNAs.

The first limitation can be overcome using a first-order Chebyshev approximation (Kipf *et al.*, 2017). Specifically, $\tilde{\mathbf{A}}_m = \mathbf{A}_m + \mathbf{I}_m$ denotes the adjacent matrix of $G_m$with self-loop, where $\mathbf{I}_m$ is the identity matrix. $\tilde{\mathbf{D}}_m$ denotes the diagonal matrix with $[\tilde{\boldsymbol{D}}_m]_{ii} = \sum_j [\tilde{\boldsymbol{A}}_m]_{ij}$. GFT is reformulated as follows,

$$\mathbf{x}^{(t+1)} = \tilde{\boldsymbol{D}}_m^{-1/2}\tilde{\mathbf{A}}_m\tilde{\boldsymbol{D}}_m^{-1/2}\mathbf{x}^{(t)}g_{\theta_1} \tag{20}$$

where g_{θ_1} a scalar parameter, is the first entry of $g_{\boldsymbol{\Theta}}$. This means that when a graph signal over G_m is transformed by GFT, the eigen-decomposition for G_m is not required any more. At the same time, after comparison with Eq. (19), Eq. (20) reduces the size of filter parameters, which speeds up the training process and may avoid overfitting.

The second limitation can be resolved by extending a vector graph signal $\mathbf{x}$ into a f_m-dimensional graph signal $\mathbf{X} \in \mathbb{R}^{m \times f_m}$. Furthermore, the vector filter parameter θ can be extended to a matrix of filter parameters $\boldsymbol{\Theta} \in \mathbb{R}^{f_m \times F}$ with f_m input channels and F filters. As a result, the final spectral convolution operation over miRNA functional similarity network G_m is shown using the following equation:

$$\mathbf{X}^{(t)} = \tilde{\boldsymbol{D}}_m^{-1/2}\tilde{\mathbf{A}}_m\tilde{\boldsymbol{D}}_m^{-1/2}\mathbf{X}^{(t-1)}\boldsymbol{\Theta}_{\mathbf{m}}^{(t-1)} \tag{21}$$

The formulation of Eq. (21) can be considered a linear layer of feed-forward neural networks if $\mathbf{X}^{(t-1)}$ is the input and $\mathbf{X}^{(t)}$ is the feature map convoluted via G_m. Therefore, a natural extension will be to add in a nonlinear activation function and stack multiple layers (LeCun *et al.*, 2015), which will enhance the expressiveness of the model. To simplify notations, $\tilde{\mathbf{D}}_m^{-1/2}\tilde{\mathbf{A}}_m\tilde{\mathbf{D}}_m^{-1/2}$ was denoted using $\tilde{\mathbf{L}}_m$. Then, t-layer GCNs for the miRNA feature extraction are defined as follows

$$\mathbf{X}^{(t)} = \mathrm{GCN}_m^{(t)}(\mathbf{X}) = \mathrm{relu}\left(\tilde{\mathbf{L}}_m\mathrm{relu}\left(\cdots relu\left(\tilde{\boldsymbol{L}}_m\boldsymbol{X}\boldsymbol{\Theta}_{\boldsymbol{m}}^{(1)}\right)\cdots\right)\boldsymbol{\Theta}_{\mathbf{m}}^{(t)}\right) \tag{22}$$

where $\mathbf{X}^{(1)}$ is a randomly initialized embedding. Starting with $\mathbf{X}^{(1)}$, a GCN transforms the embedding $\mathbf{X}^{(t-1)}$ into $\mathbf{X}^{(t)}$ in each layer. Similarly,

$$\mathbf{Y}^{(t)} = \mathrm{GCN}_d^{(t)}(\mathbf{Y}) = \mathrm{relu}\left(\tilde{\mathbf{L}}_d\mathrm{relu}\left(\cdots relu\left(\tilde{\boldsymbol{L}}_d\boldsymbol{Y}\boldsymbol{\Theta}_{\boldsymbol{d}}^{(1)}\right)\cdots\right)\boldsymbol{\Theta}_{\mathbf{d}}^{(t)}\right) \tag{23}$$

where$\tilde{\mathbf{L}}_d = \tilde{\mathbf{D}}_d^{-1/2}\tilde{\mathbf{A}}_d\tilde{\mathbf{D}}_d^{-1/2}$.

Thus, considering a miRNA functional similarity network and a diseases semantic similarity network, starting from the randomly initialized embedding $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(1)}$, a GCN transforms the embeddings in a layer by layer manner and finally outputs $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$. These learned embeddings will be used as the input for the downstream NIMC-based rating model to make final association predictions.

2.8 NIMC with GCN for miRNA-disease association predictions

NIMC and GCN were integrated into a unified end-to-end neural network learning framework NIMCGCN for MDA. The overview of the framework NIMCGCN is shown in Figure 2. The total framework of NIMCGCN includes three following modules. (i) The preprocessing module integrates different sources of information to construct the miRNA-disease heterogeneous network. (ii) In the learning module, GCNs are first leveraged to learn miRNA and disease embeddings over a miRNA functional similarity network and a disease semantic similarity network, respectively. Then, the learned embeddings are input into the neural inductive matrix completion model. The miRNA and disease representations are obtained using nonlinear transformations. The learning module learns the parameters of GCN and NIMC based on the observed known associations in an end-to-end supervised learning way. (iii) The prediction module makes a nonlinear inductive matrix completion based on the well-trained model. The NIMCGCN is formally formulated as follows.

$$\begin{aligned}&\min_{\boldsymbol{\Psi}_m,\boldsymbol{\Psi}_d,\boldsymbol{\Theta}_m,\boldsymbol{\Theta}_d} \frac{(1-\alpha)}{2}\left\|P_\Omega\left(\boldsymbol{T}-\boldsymbol{\phi}_m^l\left(GCN_m^{(t)}(\boldsymbol{X})\right)\boldsymbol{\phi}_d^l\left(\boldsymbol{GCN}_d^{(t)}(\boldsymbol{Y})\right)^T\right)\right\|_F^2\\&+\frac{\alpha}{2}\left\|P_{\bar{\Omega}}\left(\boldsymbol{T}-\boldsymbol{\phi}_m^l\left(GCN_m^{(t)}(\boldsymbol{X})\right)\boldsymbol{\phi}_d^l\left(\boldsymbol{GCN}_d^{(t)}(\boldsymbol{Y})\right)^T\right)\right\|_F^2\\&+\lambda\left(||\boldsymbol{\Theta}_m||^2+||\boldsymbol{\Theta}_d||^2\right)+\lambda\left(||\boldsymbol{\Psi}_m||^2+||\boldsymbol{\Psi}_d||^2\right)\end{aligned} \tag{24}$$

A mini-batch gradient descent with adaptive moment estimation (Kingma *et al.*, 2015) was adopted to optimize the parameters $\boldsymbol{\Psi}_m$, $\boldsymbol{\Psi}_d$, $\boldsymbol{\Theta}_m$, $\boldsymbol{\Theta}_d$ of NIMCGCN. First, for each mini-batch training iteration, equal-sized batches of miRNA-disease pairs from the set of positive association entries Ω and the set of unobserved entries $\bar{\Omega}$ were sampled. Second, in the process of forward propagation, the embeddings for the sampled miRNAs and diseases are learned with graph convolution network according to Eq. (22) and Eq. (23), respectively. Next, graph convolutional embeddings of miRNAs and diseases are separately fed into nonlinear fully connected layers to obtain final embeddings via nonlinear transformations. Then, the miRNA-disease associations were predicted using the inner product

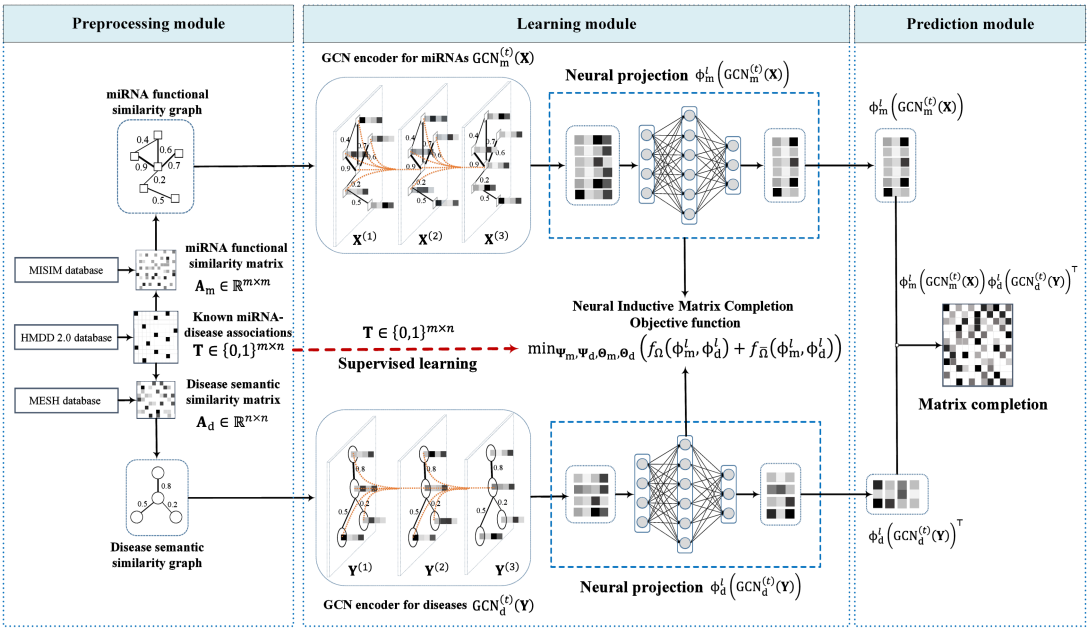

Fig. 2. The framework of NIMCGCN

of two transformed embeddings. The parameters of NIMC and GCN are learned via back propagations.

3 Results and discussion

3.1 Experimental data and settings

Three databases were used in the experiments. The disease semantic similarity data was derived from the MeSH disease descriptor database (http://www.ncbi.nlm.nih.gov/mesh). The miRNA functional similarity data was derived from the MISIM database (http://www.cuilab.cn/files/images/cuilab/misim.zip). The known human miRNA-disease association data was obtained from the HMDD v2.0 database (http://www.cuilab.cn/hmdd). Two miRNA-disease association datasets derived from HMDD v2.0 database were used in our experiments; the first dataset (D1 for short) included 383 diseases and 495 miRNAs, and provided 5430 experimentally verified associations, whereas the second dataset (D2) contained 6313 experimentally verified human miRNA-disease associations between 336 diseases and 577 miRNAs. The experimental code is implemented based on the open source machine learning framework Pytorch (https://pytorch.org). GCN encoders are implemented based on the open source geometric deep learning extension library Pytorch Geometry (https://github.com/rusty1s/pytorch_geometric). All experiments are carried on Windows 10 operation system with a Dell Precision T5820 workstation computer of an intel W-2145 8 cores, 3.7 GHz CPU and 64 G memory.

3.2 Performance evaluation

In the experiments, Leave-One-Out-Cross-Validation (LOOCV) and 5-Fold-Cross Validation (5FCV) were used to evaluate the predictive performance of NIMCGCN. In global LOOCV, all verified miRNA-disease associations were considered prediction objectives, and each known association acted by turn as the test sample, while the other known associations were considered training samples. In global 5FCV, the known miRNA-disease associations were considered positive samples and randomly divided into five subsets. A part of them was considered the testing set and the rest was considered the training set. When the ranking score of the miRNA-disease pair (m_i, d_j) was higher than a specific threshold, the model was considered to successfully predict the (m_i, d_j) miRNA-disease pair. In local 5FCV, while only considering miRNAs for a specific disease, say disease d_i, the miRNAs related to d_i were left out as test samples, while other associations were considered training samples. All the miRNA-disease associations in the test set were set to be 0 in the association matrix $\mathbf{T}$. The receiver operating characteristic (ROC) curves were drawn for performance evaluation. The area under the ROC curve (AUC) was calculated to evaluate the performance of the model.

3.3 Components affecting prediction performance

Three components significantly affected the performance of NIMCGCN: (i) biased item α in the loss function of neural inductive matrix completion formulation defined by Eq. (24), (ii) neural projection layer l for miRNA and disease feature transformation and (iii) GCN encoders layer t for miRNAs and diseases. $\alpha = 0.4$, $l = 3$, and $t = 2$ were identified as the best parameters for NIMCGCN. Specifically, take α as an identification example, we randomly choose several pairs of l and t, and output the AUCs of different $\alpha = 0.1,\ 0.2, \ldots,\ 0.9$ for each pair of l and t. When we set $\alpha = 0.4$, the best performance is obtained for most cases. In a similar way, we can identify $l = 3$ and $t = 2$ as the best performance parameters. In fact, the experimental results show that the effect of α, l, and t on AUCs are almost independent of each other.

In the follows, let other two parameters fixed, we varied a parameter to test its effect on AUCs. All experiments are conducted on the dataset D1 in 5FCV with one randomized division on known miRNA-disease associations.

3.3.1 Effect of the biased item α

The biased item α in the Eq. (24) was introduced to appropriately weigh observed and unobserved entries. The loss function was optimized only using positive samples if $\alpha = 0$ and only using unobserved samples if $\alpha = 1$. Figure 3(a) shows the effects of different α on the prediction performance of NIMCGCN. The performances when we appropriately weighed observed and unobserved entries were superior to the performance when we used only positive samples or unobserved samples.

3.3.2 Effect of neural projection layers

The low-rank feature projection matrix in IMC was replaced with nonlinear neural projection layers, which induced nonlinear transformations on top of the miRNA or disease feature representations in NIMCGCN. In this experiment, the effect of the neural projection

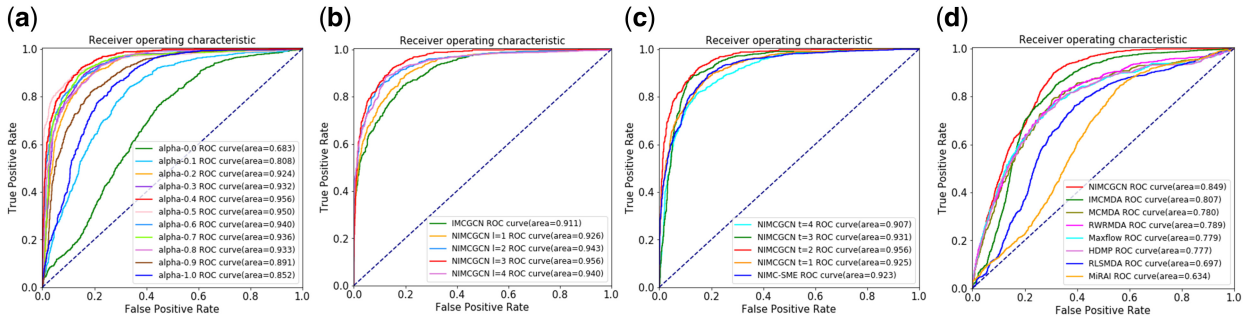

Fig. 3. (**a**) The prediction performance on different α; (**b**) Comparison of predicted performance for different projection layers; (**c**). Comparison of ROC between NIMCGCN and NIMC with similarity matrix encoders; (**d**) Comparison of ROC between NIMCGCN and other existing methods in LOOCV

layer on the performance of NIMCGCN was evaluated under two following situations. In the following situations, GCN encoders with $t = 2$ were assumed to generate the miRNA and disease embeddings. These embeddings were inputted into neural projection layers to make final predictions. The biased item is set as $\alpha = 0.4$

1. *Linear projection layer*: In this situation, we set the layer number of the neural projection layer $l = 1$ and removed nonlinear activation functions. The weight matrix was $\mathbf{Z}_1 = \mathbf{W}_m^{(1)} \in \mathbb{R}^{f_m \times k}$ and $\mathbf{Z}_2 = \mathbf{W}_d^{(1)} \in \mathbb{R}^{f_d \times k}$. It is noteworthy that the neural projection layers of NIMCGCN in this situation were equivalent to the linear projection matrices in the classic IMC.
2. *Neural projection with l-layer*: In these situations, the neural projection layers defined in Eqs. (13) and (14) were adopted to transform the GCN-learned miRNA and disease embeddings. We performed the experiments for $l = 1, 2, 3, 4$ situations.

The comparative results shown in Figure 3(b) indicate that the performance of neural projection was better than that of linear projection when GCN-embeddings were used as feed features in both cases. For the neural projections, in general, the prediction performance improved as the number of layers increased. However, when the number of layers $l > 3$, the predicted performance of NIMCGCN was degraded.

3.3.3 Effect of GCN encoders

One of the characteristics of NIMCGCN is the use of GCNs for encoding miRNAs and diseases. We investigated the capability of GCN encoders by comparing the prediction performance when using the GCN encoder, and the performance when using the similarity matrix encoder used in Chen *et al.* (2018a). $\mathbf{A}_m$ was considered the similarity adjacent matrix for miRNAs and $\mathbf{A}_d$ for diseases. The similarity matrix encoder for miRNAs uses *i*-row of $\mathbf{A}_m$ as the embedding of the miRNA *i*. Disease embeddings were handled similarly.

In this experiment, we compared the performance of NIMC with different convolutional layers GCN encoders and the performance of NIMC with similarity matrix encoders. In the following situations, feature transformations were conducted by the $l = 3$ neural projection layers. The comparative result is shown in Figure 3(c). The results show that the NIMC with $t = 1, 2, 3$ convolutional layers GCN encoders provide better prediction performance than NIMC with similarity matrix encoders. GCN encoders with $t = 2$ provided the best performance. However, we also note that with convolutional layers increasing ($t \geq 4$), the performance of GCN encoders decrease as higher convolutional layers oversmoothed the encoded features.

3.4 Comparisons with existing work

In recent years, researchers have proposed many miRNA-disease association prediction methods. However, the datasets or evaluated methods used in the existing methods are not consistent. In order to compare NIMCGCN fairly with the existing methods, we conducted experiments from the following several aspects.

First, we used the same LOOCV as described by the reference Chen *et al.* (2018a) to compare NIMCGCN (our method) with IMCMDA (Chen *et al.*, 2018a), MCMDA (Li *et al.*, 2017), RWRMDA (Chen *et al.*, 2012), Maxflow (Yu *et al.*, 2017), HDMP (Xuan *et al.*, 2013), RLSMDA (Chen *et al.*, 2014), MiRAI (Pasquier and Gardès 2016) to evaluate the performance of our proposed method. All compared results were carried experiments on D1 dataset. The AUCs of NIMCGCN, IMCMDA, MCMDA, RWRMDA, Maxflow, HDMP, RLSMDA and MiRAI were 0.849, 0.807, 0.780, 0.789, 0.779, 0.777, 0.697, 0.634, respectively. As shown in Figure 3(d), NIMCGCN shows the best prediction performance.

Second, we compared NIMCGCN with other 10 methods in global LOOCV. All these methods conducted global LOOCV on D1 dataset. In global LOOCV, each known miRNA-disease association was choosing as test sample in turn and other known associations were treated as training samples to train the model. Prediction scores of the test sample and all candidate samples (those miRNA-disease pairs without association evidences, i.e. the '0' entries in interaction matrix) could be obtained by NIMCGCN. Then, the test sample was ranked with all candidate samples based on their scores, and if the rank was higher than the specific threshold, the test sample was successfully predicted. Several methods, such as Chen et al. (2018b, c, d, 2019b, c) and Ding et al. (2019), have been published recently and provide the state-of-the-art of computational prediction results for miRNA-disease association predictions. The compared results are shown in Table 1. The global LOOCV AUC of NIMCGCN achieves 0.9387 which is superior to the results of other methods.

Third, we also compared NIMCGCN with other nine methods in global 5FCV. All these methods conducted global 5FCV on D1 dataset. In global 5FCV, 100 randomized divisions on known miRNA-disease associations are implemented to reduce the impact caused by samples division. Finally, the result shows that the mean and the standard deviation AUC of NIMCGCN were respectively 0.9291 and 0.00031, which are obviously superior to the results of other compared methods. The details of compared results are shown in Table 1.

Finally, as some studies, such as MIDP (Xuan *et al.*, 2015), HDMP (Xuan *et al.*, 2013), RLSMDA (Chen et al., 2014), RWRMDA (Chen et al., 2012), Katz-WS (Zhang *et al.*, 2019) and Katz-ML(Zhang *et al.*, 2019), provided local 5FCV results for 15 specific diseases on the D2 dataset, we also performed experiments for the same 15 specific diseases on the same D2 dataset to compare the performance. As shown in Table 2 (in Supplementary document), it is obvious that the performance of our method was better than those of MIDP, HDMP, RLSMDA, RWRMDA Katz-WS and Katz-ML for most diseases, except for urinary bladder neoplasms.

3.5 Case studies

To further demonstrate the prediction accuracy of NIMCGCN, we performed case studies on three important complex human diseases such as colon cancer, lymphoma and kidney cancer by prioritizing candidate miRNAs for the diseases using our model with the

Table 1. Comparisons with other methods in global LOOCV and global 5FCV on dataset D1

Comparison methods	LOOCV	5-FCV
NIMCGCN (our method)	**0.9387**	**0.9291 ± 0.00031**
BNPMDA (Chen *et al.*, 2018c)	0.9025	0.8980±0.0013
ABMDA (Chen *et al.*, 2019b)	0.9175	0.9023±0.0016
EDTMDA (Chen *et al.*, 2019c)	0.9301	0.9192±0.0009
MDHGI (Chen *et al.*, 2018b)	0.8951	0.8794±0.0012
LRSSLMDA (Chen *et al.*, 2017c)	0.9179	0.9181±0.0004
PBMDA (You *et al.*, 2017)	0.9165	0.9172±0.0007
MCMDA (Li *et al.*, 2017)	0.8745	0.8767±0.0011
MaxFlow (Yu *et al.*, 2017)	0.8623	0.8579±0.0010
HDMP (Xuan *et al.*, 2013)	0.8364	0.8342±0.0010
IIMCMP (Ding *et al.*, 2019)	0.9011	—

training dataset from HMDD v2.0. We verified the top 50 predictions made using NIMCGCN with three other miRNA-disease association databases, namely, dbDEMC v2.0 (Yang *et al.*, 2010), miR2Disease (Jiang *et al.*, 2009) and miRCancer (Xie *et al.*, 2013).

The dbDEMC database calculates the differential expression value of miRNAs by acquiring the microarray data to obtain cancer-related miRNAs. The dbDEMC v2.0 (dbDEMC2 for short) updates them and adds more cancer-related miRNAs obtained from expression data.

The miR2Disease database is manually annotated, which is summarized by the staff after surveying many studies. The miRCancer database provides a comprehensive collection of miRNA expression profiles from various human cancers. These expression profiles were extracted automatically from published literature in PubMed. Text mining was used to collect information, and manual correction was adopted. The accuracy achieves 100%.

Once the specific disease prediction results of NIMCGCN were obtained, we removed the miRNAs with value equal to 1 in the original miRNA-disease matrix. Then, the predicted results of the remaining new miRNAs were descending sorted by prediction scores. Thus, we obtained the top 50 associated miRNAs.

The prediction and validation results of the colon cancer-associated miRNAs are shown in Table 3 (in Supplementary document). The predicted top 50 potential colon cancer-associated miRNAs can be verified in dbDEMC2, miR2Disease and miRCancer databases. The prediction accuracy in top 50 was 100%.

The prediction results for lymphoma are shown in Table 4 (in Supplementary document). Forty-seven out of the top predicted 50 potential lymphoma-associated miRNAs were verified in dbDEMC2, miR2Disease and miRCancer databases. Three miRNAs, namely, has-mir-34c, has-mir-378a and has-mir-103a, were not supported by relevant literature or databases. The prediction accuracy reached 94% in top 50.

The prediction results for kidney cancer are shown in Table 5 (in Supplementary document). Forty-eight out of the top predicted 50 potential kidney cancer-related miRNAs were verified in dbDEMC2, miR2Disease and miRCancer databases. Two miRNAs, hsa-let-7e, hsa-mir-196a, were not supported by relevant literatures or databases. The prediction accuracy was 96% in the top 50.

3.6 Prediction of unknown disease

NIMCGCN can be used to predict the potential miRNAs associated with an unknown disease. Unknown diseases are those that have not been shown to associate with any miRNAs. We used breast cancer as a case study in this experiment.

First, the known miRNAs associated with breast cancer were removed. Then, the changed miRNA-disease matrix was used as input in training our model. Finally, the top predicted 50 miRNAs were verified in four different databases, namely HMDD v2.0 (Li *et al.*, 2014), dbDEMC2 (Yang *et al.*, 2010), miR2Disease (Jiang *et al.*, 2009) and miRCancer (Xie *et al.*, 2013). As shown in Table 6 (in Supplementary document), the top 50 miRNAs were present in all databases, and the prediction accuracy of the top 50 was 100%.

4 Conclusion

Identification of potential miRNA-disease associations using computational approaches is important as it will improve our understanding of the pathogenesis of diseases and guide treatment. In this study, we extended the classical inductive matrix completion model and developed a novel model called NIMCGCN for miRNA-disease association prediction. NIMCGCN was compared in 5FCV and LOOCV with the existing computational prediction methods, including BNPMDA (Chen *et al.*, 2018c), ABMDA (Chen *et al.*, 2019b), IMCMDA (Chen *et al.*, 2018a), EDTMDA (Chen *et al.*, 2019c), PBMDA (You *et al.*, 2017) and so on, which show excellent performance for prediction of miRNA-disease associations. The experimental results show that NIMCGCN has significantly high accuracy in both 5FCV and LOOCV. Three high-risk human diseases, colon cancer, lymphoma and kidney cancer, were used as case studies to further evaluate the performance of NIMCGCN. Fifty, 47 and 48 out of the top 50 predicted miRNAs were verified using published experimental studies. Finally, 100% associated miRNAs were detected using NIMCGCN for breast cancer, which is considered a case study for verifying the ability of NIMCGCN to predict unknown diseases.

However, NIMCGCN has certain limitations, which require further investigations. First, the structural information regarding miRNA and disease similarity networks significantly affect the learned feature representations, which further affect the final prediction results. Methods of gathering different valuable biological information to effectively construct miRNA and disease similarity networks are worth investigating in the future. Second, in NIMCGCN, feature representations of miRNAs and diseases are learned with GCNs from a disease semantic similarity network and a miRNA functional similarity network. Other feature representation learning approaches, such as multiple similarity kernel-based learning or even multi-layer multiple similarity kernel-based learning, warrant further investigation.

Acknowledgements

We thank anonymous reviewers for valuable suggestions.

Funding

This work has been supported by the National Natural Science Foundation of China (61562091 and 91631305).

Conflict of Interest: none declared.

References

Bandyopadhyay,S. *et al.* (2010) Development of the human cancer microRNA network. *Silence*, **1**, 6.

Chen,M. *et al.* (2016) Uncover miRNA-disease association by exploiting global network similarity. *PLoS One*, **11**, e0166509.

Chen,M. *et al.* (2018a) A novel information diffusion method based on network consistency for identifying disease related microRNAs. *RSC Adv.*, **8**, 36675–36690.

Chen,M. *et al.* (2018b) Global similarity method based on a two-tier random walk for the prediction of microRNA-disease association. *Sci. Rep.*, **8**, 6481.

Chen,M. *et al.* (2019) Bipartite heterogeneous network method based on co-neighbor for MiRNA-disease association prediction. *Front. Genet.*, **10**, 385.

Chen,X. *et al.* (2012) RWRMDA: predicting novel human microRNA-disease associations. *Mol. Biosyst.*, **8**, 2792–2798.

Chen,X. *et al.* (2014) Semi-supervised learning for potential human miRNA-disease associations inference. *Sci. Rep.*, **4**, 5501.

Chen,X. *et al*. (2017a) Long non-coding RNAs and complex disease: from experimental results to computational models. *Brief. Bioinform*., **18**, 558–576.

Chen X et al. (2017b) NRDTD: a database for clinically or experimentally supported non-coding RNAs and drug targets associations. *Database (Oxford)*, **2017**, 1–6.

Chen,X. *et al*. (2017c) LRSSLMDA: Laplacian regularized sparse subspace learning for MiRNA-disease association prediction. *PLoS Comput. Biol*., **13**, e1005912.

Chen,X. *et al*. (2018a) Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics*, **34**, 4256–4265.

Chen,X. *et al*. (2018b) MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput. Biol*., **14**, e1006418.

Chen,X. *et al*. (2018c) BNPMDA: bipartite network projection for MiRNA-disease association prediction. *Bioinformatics*, **34**, 3178–3186.

Chen,X. *et al*. (2018d) EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. *Cell Death Dis*., **9**, 3.

Chen,X. *et al*. (2019a) MicroRNAs and complex diseases: from experimental results to computational models. *Brief. Bioinform*, **20**, 515–539.

Chen,X. *et al*. (2019b) Adaptive boosting-based computational model for predicting potential miRNA-disease associations. *Bioinformatics*, **35**, 4730–4738.

Chen,X. *et al*. (2019c) Ensemble of decision tree reveals potential miRNA-disease associations. *PLoS Comput. Biol*., **15**, e1007209.

Cheng,L. *et al*. (2019) Adaptive multi-view multi-label learning for identifying disease-associated candidate miRNAs. *PLoS Comput. Biol*., **15**, e1006931.

Defferrard,M. *et al*. (2016) Convolutional neural networks on graphs with fast localized spectral filtering. In Advances in Neural Information Processing Systems 2016, pp. 3844–3852.

Ding,X. *et al*. (2019) Improved inductive matrix completion method for predicting microRNA-disease associations. In *ICIC 2019:* Intelligent Computing Theories and Application, Vol. **11644**, pp. 247–255.

Goh,K.-I. *et al*. (2007) The human disease network. *Proc. Natl. Acad. Sci. USA*, **104**, 8685–8690.

Jiang,Q. *et al*. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*., **37**, D98–D104.

Kingma,D. P. *et al*. (2015) Adam: a method for stochastic optimization. In *The 3rd International Conference on Learning Representations (ICLR 2015)*, pp. 1–13.

Kipf,T.-N. *et al*. (2017) Semi-supervised classification with graph convolutional networks. In *The 5th International Conference on Learning Representations (ICLR 2017)*, pp. 1–14.

LeCun,Y. *et al*. (2015) Deep learning. *Nature*, **521**, 436–444.

Li,G. *et al*. (2018) Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity. *J. Biomed. Inform*., **82**, 169–177.

Li,J.-Q. *et al*. (2017) MCMDA: matrix completion for MiRNA-disease association prediction. *Oncotarget*, **8**, 21187–21199.

Li,Y. *et al*. (2014) HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*., **42**, D1070–D1074.

Lu,M. *et al*. (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.

Luo,J. *et al*. (2017) Collective prediction of disease-associated miRNAs based on transduction learning. *IEEE/ACM Trans. Comput. Biol. Bioinform*., **14**, 1468–1475.

Natarajan,N. *et al*. (2014) Inductive matrix completion for predicting gene-disease associations. *Bioinformatics*, **30**, i60–i68.

Pasquier,C. and Gardès,J. (2016) Prediction of miRNA-disease associations with a vector space model. *Sci. Rep*., **6**, 27036.

Su C. et al. (2018) Network embedding in biomedical data science. *Brief. Bioinform*., 2018, 1–16.

Wang,D. *et al*. (2010) Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*, **26**, 1644–1650.

Xie,B.-Y. *et al*. (2013) miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics*, **29**, 638–644.

Xuan,P. *et al*. (2013) Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One*, **8**, e70204.

Xuan,P. *et al*. (2015) Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics*, **31**, 1805–1815.

Xuan,P. *et al*. (2018) Dual convolutional neural network based method for predicting disease-related miRNAs. *Int. J. Mol. Sci*., **19**, 3732.

Xuan,P. *et al*. (2019) Inferring the disease-associated miRNAs based on network representation learning and convolutional neural networks. *Int. J. Mol. Sci*., **20**, 3648.

Yang,Z. *et al*. (2010) dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics*, **11**, S5.

You,Z.-H. *et al*. (2017) PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol*., **13**, e1005455.

Yu,H. *et al*. (2017) Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm. *Sci. Rep*., **7**, 43792.

Zhang,D. *et al*. (2018) Network representation learning: a survey. *IEEE Trans. Big Data*., doi: 10.1109/TBDATA.2018.2850013.

Zhang,X. *et al*. (2019) Meta-path methods for prioritizing candidate disease miRNAs. *IEEE/ACM Trans. Comput. Biol. Bioinform*., **16**, 283–291.

Zeng,X. *et al*. (2016) Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform*., **17**, 193–203.