2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-0126-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIBM52615.2021.9669673

Graph Attention Mechanism-based Deep Tensor Factorization for Predicting disease-associated miRNA-miRNA pairs

Jiawei Luo, Zihan Lai, Cong Shen, Pei Liu
College of Computer Science and Electronic Engineering
Hunan University
Changsha, China
luojiawei@hnu.edu.cn

Heyuan Shi*
KLISS, BNRist
Tsinghua University
Beijing, China
hey.shi@foxmail.com

***Abstract*—MicroRNAs (miRNAs) play a significant role in regulating gene transcription and tend to act in a combinatorial way, which provides great insights to explore disease-related miRNA pairs or modules for comprehending the synergistic roles of miRNAs in complex diseases. As wet experiments are often laborious and costly, computational methods offer great convenience for predicting potential associations between miRNAs and diseases. Existing methods focus on either the 'one miRNA-one disease' paradigm, or merely the synergetic miRNA network about specific diseases, which may lead to the incomplete understanding of the synergistic effect of miRNAs on the pathogenesis of complex diseases. In this work, we present a novel tensor-based framework, named GraphTF[1], to predict disease-associated miRNA-miRNA pairs. GraphTF exploits graph attention network to effectively capture node features over multi-source biological network. Then, the learned miRNA and disease representations are used to reconstruct the association tensor for predicting potential disease-associated miRNA-miRNA pairs. Empirical results showed that the proposed method outperformed all other state-of-the-art methods under five-fold cross-validation. Robustness experiments also indicated the stability of GraphTF. Moreover, case studies for Breast Neoplasms and Lung Neoplasms further demonstrated the effectiveness of GraphTF in identifying potential disease-related miRNA-miRNA pairs.**

***Index Terms*—disease associated miRNA-miRNA pairs, graph attention network, tensor factorization**

I. Introduction

MiRNAs are a class of endogenous non-coding RNAs with about 22-25 nucleotides. MiRNAs tend to co-regulate target genes in a combinatorial way and influence the occurrence and development of complex diseases [1]. Therefore, identifying disease-related miRNA modules is of great help for the pathological study and biomarker detection of diseases.

To dig out more about miRNA synergistic mechanism, Zhao et al. [2] constructed a miRNA-miRNA synergistic relationships of differentially expressed miRNAs of colorectal cancer via computational analysis of miRNA and gene expression microarray and uncovered the functional synergistic miRNA pairs in colorectal cancer tissues. Cilek et al. [3] explored functionally relevant miRNAs by constructing pathway-based miRNA-miRNA networks in trastuzumab treated breast cancer cell lines. Although these studies give a certain comprehension of the functional synergism of miRNAs, they only study the effect of synergetic miRNA combinations on specific diseases, which is difficult to extend to large-scale prediction of miRNA combinations for multiple diseases.

Consequently, developing an effective computational framework to discover disease-associated miRNA combinations is a promising mission. However, how to model the miRNA-miRNA-disease associations still remains challenging. High-order tensors (i.e., N-way arrays with $N \geq 3$) is ubiquitous in social network analysis and have shown to successfully represent data that have more than two modes of variation [4]. Huang et al. [5] present Tensor Decomposition with Relational Constraints (TDRC) which combined miRNA functional similarity and disease semantic similarity to constrain the factorization of tensor to predict multi-type miRNA-disease associations. Liu et al. [6] integrated multi-view miRNAs and diseases information to discover potential miRNA-miRNA-disease associations based on a tensor completion framework. Although conventional tensor-based factorization method is efficient to represent multi-way data, it ignores the complex interactions between each other. As machine learning-based method can better exploit the relationship using contextual information [7], Sun et al. [8] proposed a Deep Tensor Factorization model (DTF) to predict drug synergy. Julkunen et al. [9] present comboFM, a machine learning framework for predicting the responses of drug combinations by modeling the drug interactions through higher-order tensors. Nevertheless, existing approaches cannot fully explore the rich structural information contained in the obtained similarity matrices, which may affect the quality of feature representations and thus influence the predictive performance.

To address the above issues, we propose a novel tensor factorization framework combined with graph attention network (GAT), namely GraphTF to predict new diseases related miRNA-miRNA pairs. First, we model the triple miRNA-miRNA-disease relationships as an association tensor, where

*Corresponding author: hey.shi@foxmail.com (Heyuan Shi).
[1]Availability and implementation: https://github.com/ZihanLai/GraphTF.

its three dimensions representing miRNA, miRNA and disease, respectively. Then we use graph attention network to extract the feature representation of miRNAs and diseases from multi-source data. After that, GraphTF applies a tensor-based procedure to reconstruct the association tensor using the learned feature representation for prediction. We use five-fold cross-validation to evaluate the performance of GraphTF. Experimental results demonstrate that our proposed GraphTF model outperforms existing state-of-the-art methods. Case studies on Breast Neoplasms and Lung Neoplasms further confirm the powerful predictive ability of GraphTF.

II. Materials

A. Construction of known third-order association tensor

Previous studies have shown that miRNAs from the same family tend to regulate common gene expression [10] [11] and members from the same miRNA family are related to similar diseases [12] [13]. Therefore, we collect the miRNA-miRNA synergetic links using the miRNA family information from miRbase v22 [14] and download the experimentally verified human miRNA-disease associations from HMDD v3.2 [15]. To construct the miRNA-miRNA-disease associations, we choose shared miRNAs that appear in both miRNA-disease and miRNA-miRNA relations. Finally, we get 964 known miRNA-miRNA associations and 5187 known miRNA-disease associations, covering 325 miRNAs in 78 families and 283 diseases. The third-order tensor X is constructed as follows:

$$X\left(m_i, m_j, d_k\right) = \begin{cases} 1, \langle m_i, m_j \rangle = 1, \\ \quad \langle m_i, d_k \rangle = 1 \ and \ \langle m_j, d_k \rangle = 1 \\ 0, otherwise. \end{cases} \tag{1}$$

where $X\left(m_i, m_j, d_k\right) = 1$ indicates that miRNA i and miRNA j are jointly associated with disease k, which may form a miRNA pair and have a synergistic or similar function to disease. The tensor is extremely sparse containing only 15484 known interactions.

B. MiRNA functional similarity

To avoid existing reliance between miRNAs and diseases, we measure the miRNA functional similarity using associations between miRNAs and genes in the same way as GRNMF [16]. The gene interaction score (IS) is downloaded from the HumanNet database. Then the normalized interaction scores (NIS) of genes are obtained using min-max normalization as follows:

$$NIS\left(g_i, g_j\right) = \frac{IS\left(g_i, g_j\right) - IS_{min}}{IS_{max} - IS_{min}} \tag{2}$$

where $IS\left(g_i, g_j\right)$ represents the raw interaction scores of gene i and j. IS_{max} and IS_{min} denote the maximum and minimum IS in HumanNet. Consequently, the similarity between gene i and j is given as follows:

$$S\left(g_i, g_j\right) = \begin{cases} 1, \ g_i = g_j \\ 0, \ e\left(g_i, g_j\right) \notin HumanNet \\ NIS\left(g_i, g_j\right), \ e\left(g_i, g_j\right) \in HumanNet. \end{cases} \tag{3}$$

where $e\left(g_i, g_j\right)$ indicates the association between gene i and j. Afterward, the similarity between gene m and gene set $GS = \{g_1, \ \ldots, g_k, \ \ldots, g_n\}$ is calculated as follows:

$$S\left(g_m, GS\right) = \min_{1 \le k \le n} S\left(g_m, g_k\right) \tag{4}$$

The functional similarity of miRNA i and j is calculated by:

$$MFS\left(m_i, m_j\right) = \frac{\sum\limits_{g \in G_i} S\left(g, G_j\right) + \sum\limits_{g \in G_j} S\left(g, G_i\right)}{|G_i| + |G_j|} \tag{5}$$

where G_i and G_j are miRNA i and j associated gene sets respectively; and $|G|$ is the number of genes in G.

C. MiRNA sequence similarity

The miRNA sequence data is downloaded from miRbase v22 [14]. We calculate the sequence similarity by applying the Needleman-Wunsch global alignment algorithm [17]. Let $MS \in R^{q \times q}$ be the matrix of miRNA sequence similarity score. The normalized score $NMS(m_i, m_j)$ between miRNA i and j can be obtained as follows:

$$NMS\left(m_i, m_j\right) = \frac{MS\left(m_i, m_j\right) - MS_{min}}{MS_{max} - MS_{min}} \tag{6}$$

where MS_{max} and MS_{min} denote the maximum and minimum score in the similarity matrix MS.

D. Disease semantic similarity

The MeSH database can be described as a hierarchical Directed Acyclic Graph (DAG), we then obtain the disease DAGs to calculate the disease semantic similarities as the same way in MISIM [18]. For a disease d_i, the DAG_i consists not only the ancestor node of d_i and d_i itself, but the direct edges from parent nodes to child nodes. That is, $DAG_i = (d_i, A_i, \ E_i)$, where A_i is the set of all nodes in DAG_i; E_i denotes the corresponding links. Define the semantic contribution of disease $d_j \in A_i$ to disease d_i as $DC(d_i, \ d_j)$, which can be formulated as follows:

$$DC\left(d_i, d_j\right) = \begin{cases} 1, if \ d_i = d_j \\ \max\left\{\Delta * DC\left(d_i, d_j^c\right)\right\}, \ otherwise. \end{cases} \tag{7}$$

where $d_j^c \in children \ of \ d_j$, Δ denotes the semantic contribution factor ($\Delta = 0.5$). Based on Eq.(7), the semantic similarity score between disease i and j can be calculated as:

$$DS\left(d_i, d_j\right) = \frac{\sum\limits_{t \in DAG_i \cap DAG_j} \left(DC\left(d_i, d_t\right) + DC\left(d_j, d_t\right)\right)}{\sum\limits_{t \in DAG_i} DC\left(d_i, d_t\right) + \sum\limits_{t \in DAG_j} DC\left(d_j, d_t\right)} \tag{8}$$

where $DC\left(d_i, \ d_t\right)$ and $DC\left(d_j, \ d_t\right)$ denote the semantic values of disease t related to disease i and j, respectively. It can be seen that the more diseases shared by disease i and j, the more similar the two diseases are.

III. METHODS

A. Model Overview

In this work, we propose a new tensor factorization model based on GAT, called GraphTF, which consists of three parts (Figure 1). First, we calculate the miRNA functional similarity, miRNA sequence similarity and disease semantic similarity via multi-source information. Second, we use GAT to capture the complex interaction from similarity graphs and obtain the representation of miRNA and disease. GAT assigns a new weight for each edge, which can not only fully mine the similarity information, but also better capture the network topology. Finally, the learned latent embeddings are put into the deep neural network and the output embeddings are used to reconstruct the miRNA-miRNA-disease tensor for predicting disease-related miRNA-miRNA pairs.

B. Deep Tensor Factorization with Graph Attention Network

1) Tensor Factorization: Tensor factorization can effectively extract potential high-order relationships for multi-dimensional data and has been widely used in studying biological relationships. Thus, we attempt to transform the triple association prediction problem into a three-order tensor factorization problem. In general, a tensor factorization model first learns the latent low rank factors from the observed entries, then reconstructs the tensor with missing entries based on the learned factors. The CANDECOMP/PARAFAC (CP) method is one of the most commonly used tensor decomposition models. A traditional CP model decomposes a tensor into a sum of component rank-one tensors. Given a third-order tensor $X \in R^{n_1 \times n_2 \times n_3}$, the CP model can be expressed as:

$$\begin{aligned} &X \approx [\![P, B, D]\!] = \sum_{r=1}^{R} p_r \circ b_r \circ d_r, \\ &s.t. p_r \in R^{n_1 \times 1}, b_r \in R^{n_2 \times 1}, d_r \in R^{n_3 \times 1} \end{aligned} \tag{9}$$

where R denotes the number of rank-one tensors and the minimal R is the rank of tensor X and $P \in \mathbb{R}^{n_1 \times R}, B \in \mathbb{R}^{n_2 \times R}, D \in \mathbb{R}^{n_3 \times R}$ represent the decomposed factor matrices concerning the two miRNA modes and the disease mode, respectively; $\circ$ denotes the outer product. The goal to best approximate X is formulated as follow:

$$\min_{P,B,D} \|X - [\![P, B, D]\!]\|_F^2 \tag{10}$$

2) GraphTF model: Generally, the traditional tensor factorization models use different updating rules such as alternating least squares (ALS) method to optimize the objective function, which cannot fully extract the complicated interactions between miRNAs and diseases. Inspired by NIMCGCN [19], in GraphTF, the feature matrices P and B for miRNAs and D for diseases are encoded separately by three different nonlinear fully connected layers. The nonlinear transformations of the fully connected layers for miRNAs can be expressed as follows:

$$f_{m_1}^l(P) = W_{m_1}^{(l)} relu\left(\cdots relu\left(W_{m_1}^{(1)} P + b_{m_1}^1\right) + b_{m_1}^l\right) \tag{11}$$

$$f_{m_2}^l(B) = W_{m_2}^{(l)} relu\left(\cdots relu\left(W_{m_2}^{(1)} B + b_{m_2}^1\right) + b_{m_2}^l\right) \tag{12}$$

where m_1 and m_2 represent the feature matrix of miRNAs in 1-mode and 2-mode; $W_{m_1}^{(l)}$ and $W_{m_2}^{(l)}$ are the weight matrices of the l-th layer in the nonlinear fully connected layer; $b_{m_1}^l$ and $b_{m_2}^l$ are the bias items for the l-th layer and $relu(\cdot)$ denotes the rectified linear unit nonlinear activation function.

Similarly, the final nonlinear transformations for diseases are described as follows:

$$f_d^l(D) = W_d^{(l)} relu\left(\cdots relu\left(W_d^{(1)} D + b_d^1\right) + b_d^l\right) \tag{13}$$

With the above definitions, the predicted tensor can be obtained by $\left[\!\left[f_{m_1}^1(P), f_{m_2}^1(B), f_d^1(D)\right]\!\right]$ and the deep tensor factorization model is defined as:

$$\begin{aligned} &\min_{P,B,D} \left\| A_\Omega \left(X - \left[\!\left[f_{m_1}^1(P), f_{m_2}^1(B), f_d^1(D)\right]\!\right]\right)\right\|_F^2 \\ &+ \left\| A_{\overline{\Omega}} \left(X - \left[\!\left[f_{m_1}^1(P), f_{m_2}^1(B), f_d^1(D)\right]\!\right]\right)\right\|_F^2 \\ &+ \lambda \left(\|\Gamma_{m_1}\|_F^2 + \|\Psi_{m_2}\|_F^2 + \|\Theta_d\|_F^2\right) \end{aligned} \tag{14}$$

where Ω and $\overline{\Omega}$ denote the set of observed and unobserved or unknown ternary association entries in the association tensor X i.e., if $\forall X_{ijk} \in \Omega, X_{ijk} = 1$; $\forall X_{ijk} \in \overline{\Omega}, X_{ijk} = 0$. $\Gamma_{m_1} = \left\{W_{m_1}^{(1)}, ..., W_{m_1}^{(l)}, b_{m_1}^{(1)}, ..., b_{m_1}^{(l)}\right\}$, $\Psi_{m_2} = \left\{W_{m_2}^{(1)}, ..., W_{m_2}^{(l)}, b_{m_2}^{(1)}, ..., b_{m_2}^{(l)}\right\}$, $\Theta_d = \left\{W_d^{(1)}, ..., W_d^{(l)}, b_d^{(1)}, ..., b_d^{(l)}\right\}$ are the parameters involved in Eq.(11), Eq.(12) and Eq.(13).

C. Embedding Propagation Layer

GAT is a powerful neural architecture for learning with graph, which can assign different attention scores to each neighbor and learn the preferences of nodes adaptively. To capture fine-grained features over similarity graphs, we aggregate the neighborhood information by introducing multi-head attention mechanism adopted in graph attention network.

The embedding propagation layers are motivated as adaptation layer of GNNs, which perform aggregation operations based on the node's neighborhoods. Let $p \in \mathbb{R}^{f_m}$, $b \in \mathbb{R}^{f_m}$ and $d \in \mathbb{R}^{f_d}$ be the feature vectors of miRNAs and diseases, where f_m, f_d are the feature dimensions for miRNAs and diseases, the miRNA functional, sequence similarity matrix can be $MF \in \mathbb{R}^{X \times f_m}$, $MS \in \mathbb{R}^{X \times f_m}$ and disease semantic similarity matrix $DS \in \mathbb{R}^{Y \times f_d}$, where X, Y denote the number of miRNAs and diseases, respectively. For each similarity graph, we perform self-attention on the nodes to compute attention coefficients c_{ij}^l:

$$c_{ij}^l = a\left(W_1 h_i \| W_1 h_j\right), \quad j \in N_i \tag{15}$$

where $h = \{p, b, d\}$, $a(\cdot)$ denotes a single-layer feed-forward neural network which maps the input $\mathbb{R}^{K'} \times \mathbb{R}^{K'}$ to $\mathbb{R}$, $\|$ is the concatenation operation. $W_1 \in \mathbb{R}^{f \times k}$ is a learnable weight matrix, which transforms input raw features with dimension f into high-level features of dimension k (which is set as 128 in this paper) for each node; N_i indicates

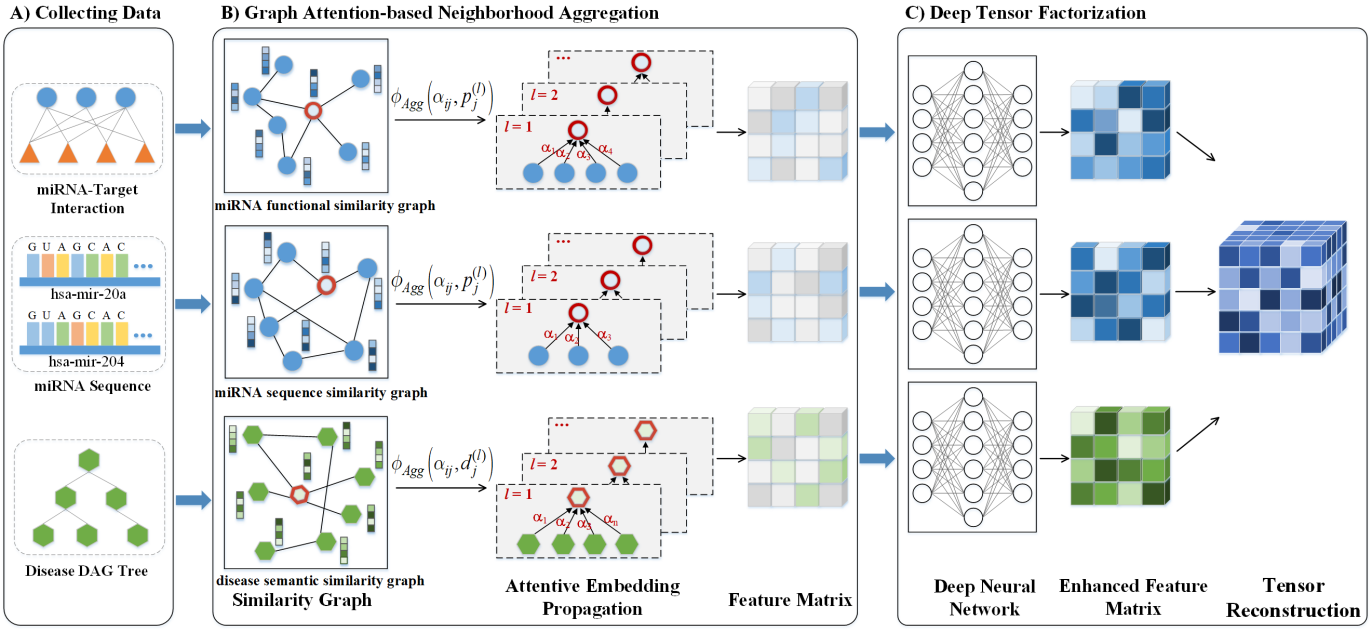

Fig. 1. Overview of the Deep Tensor Factorization with Graph Attention Network. A) Data collection and preprocessing. B) GraphTF conducts graph attention mechanism-based neighborhood information aggregation over miRNA functional similarity graph, miRNA sequence similarity graph and disease semantic similarity graph, respectively. C) The learned latent features from Step B are put into a deep tensor factorization model for link prediction.

the neighborhoods of node i. Hereafter, we normalize the coefficients across all choices of j using the softmax function:

$$\alpha_{ij}^{l} = softmax\left(c_{ij}^{l}\right) = \frac{exp\left(c_{ij}^{l}\right)}{\sum_{t\in N_i} exp\left(c_{it}^{l}\right)} \tag{16}$$

Following GAT [20], we apply the LeakyReLU nonlinearity on c_{ij}^{l} and the coefficients α_{ij}^{l} are then expressed as follows:

$$\alpha_{ij}^{l} = \frac{exp\left(LeakyReLU\left(c_{ij}^{l}\right)\right)}{\sum\limits_{t\in N_i} exp\left(LeakyReLU\left(c_{ij}^{l}\right)\right)} \tag{17}$$

Then the neighborhood aggregation for node i can be performed as:

$$h_i^{l} = \sigma\left(\sum_{j\in N_i} \alpha_{ij}^{l} W_1 h_j\right) \tag{18}$$

where $\sigma(\cdot)$ denotes the ReLU activation function.

Similar to Transformer [21], we extend our mechanism to employ multi-head attention to stabilize the learning process of self-attention. Specifically, with U attention heads executing the transformation of Eq.(18), we concatenate the multi-head features and obtain the output feature representation in the attention layer l:

$$h_i^{l} = \Big\Vert_{u=1}^{U} \sigma\left(\sum\nolimits_{j\in N_i} \alpha_{ij}^{(l,u)} W_1^{u} h_j^{(l,u)}\right) \tag{19}$$

where $\|$ is concatenation, $\alpha_{ij}^{(l,u)}$ are normalized attention score computed by the u-th head, and W_1^{u} is the corresponding linear transformation matrix.

Thereby, with the randomly initialized node embeddings $P^{(0)}$, $B^{(0)}$ and $D^{(0)}$, the feature embedding module learns the final node representation $P^{(L)}$, $B^{(L)}$ and $D^{(L)}$ by integrating the neighborhood information with attention mechanism over the similarity networks. The learned embeddings are then taken as the input for the downstream GraphTF-based model to make final association predictions.

D. Optimization for miRNA-miRNA-disease tensor reconstruction

After deriving the representation matrices P, B, D for miRNA and disease, we can then reconstruct the miRNA-miRNA-disease association tensor $\widehat{X}$ as follows:

$$\left[\left[\Phi_{m_1}^{l}\left(GAT_{m_1}^{t}\left(P^{(0)}\right)\right)\Phi_{m_2}^{l}\left(GAT_{m2}^{t}\left(B^{(0)}\right)\right)\Phi_{d}^{l}\left(GAT_{d}^{t}\left(D^{(0)}\right)\right)\right]\right] \tag{20}$$

where $\Phi_{m_1}^{l}(\cdot)$, $\Phi_{m_2}^{l}(\cdot)$ and $\Phi_{d}^{l}(\cdot)$ represent the final feature matrices of miRNA and disease after neural projection, respectively. $GAT_{m_1}^{t}\left(P^{(0)}\right)$, $GAT_{m_2}^{t}\left(B^{(0)}\right)$ and $GAT_{d}^{t}\left(D^{(0)}\right)$ denote the learned feature matrices from GATs.

For the elements in $\widehat{X}$, the higher X_{ijk} is, the more likely miRNA pair $\langle m_i, m_j\rangle$ is associated with disease d_k, otherwise, the less likely miRNA pair $\langle m_i, m_j\rangle$ is associated with disease d_k. We use mean square error to minimize the Frebious norm of the difference between preference tensor X and label tensor $\widehat{X}$ as the loss function of our model. The loss is formally formulated as follows:

$$\begin{aligned}\min_{P,B,D} &\left\|A_{\Omega}\left(X-\widehat{X}\right)\right\|_F^2 + \left\|A_{\overline{\Omega}}\left(X-\widehat{X}\right)\right\|_F^2 \\ &+\lambda\left(\|\Gamma_{m_1}\|_F^2 + \|\Psi_{m_2}\|_F^2 + \|\Theta_d\|_F^2\right)\end{aligned} \tag{21}$$

We tested our model on a machine equipped with one NVIDIA 2060 GPU and one 2.90 GHz AMD Ryzen 7 4800H

with Radeon Graphics CPU with 16GB memory. For GraphTF, the time complexity is $O\left(|V|\,FF' + |E|\,F' + X^2Y\right)$, where F is the number of input features, F' is the number of output features, $|V|$ and $|E|$ are the numbers of nodes and edges in the graph, X, Y denote the number of miRNAs and diseases in tensor factorization, respectively.

IV. Experiments

A. Baselines

Currently, there are many outstanding tensor completion methods for predicting triple associations in bioinformatics, we take the following models as baselines. Noted that as they are not developed for miRNA-miRNA disease associations prediction, we only use the algorithms for comparison. The baselines can be divided into two categories, i.e., traditional tensor factorization model and machine learning-based models. CANDECOMP/PARAFAC(CP) [22]: A classical tensor factorization model without any auxiliary information which decomposes a tensor as a sum of rank-one tensors via alternating least squares (ALS) rules. DrugCom [23]: A tensor-based framework for computing drug combinations over multiple information of drugs and diseases. TDRC [5]: A tensor decomposition method to predict multi-type miRNA-disease associations by constraining the factor matrices using auxiliary information of miRNAs and diseases. DeepSynergy [24]: A deep learning approach for predicting drug combinations, which combines the feature representation of drug combinations and cancer cell line, and treats the prediction task as a classification problem. DTF [8]: A deep tensor factorization model integrating a tensor factorization method and a deep neural network for predicting drug synergy.

B. Experimental Setup

To systematically evaluate the performance of GraphTF, we conduct five-fold cross-validation experiments and compare it with five state-of-the-art methods. The original observed tensor is very sparse with many unobserved entries, to deal with the imbalance of the positive and negative samples, we randomly select the same size of negative samples as positive samples from the missing entries of the association tensor. Then the positive and negative samples are divided into five parts. For every subset, the four parts of positive and negative samples are used as training set, the rest one part is treated as testing set. We apply grid search for hyper-parameters fining and for each fold, embedding size D is selected in $\{32, 64, 128, 256, 512\}$, projection layers L is set from $\{1, 2, 3, 4\}$, the number of attention heads U is selected from $\{1, 2, 3, 4\}$, the dropout ratio ρ is chosen in $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. We mainly choose metrics that are typical for classification task: area under the receiver operator characteristics curve (ROC-AUC), area under the precision recall curve (PR-AUC).

In our model, the training epoch is set to 300 and the parameters are optimized by Adam solver [25] with a learning rate 0.001, the dropout ratio ρ is set to 0.4. We perform neighborhood aggregation via one-layer-GAT with 3 attention heads and use three-layer MLP to enhance feature representation. Specifically, we randomly initialize the features which satisfy the standard normal distribution and set the embedding size D for miRNAs and diseases as 128.

C. Experimental Results

We compare our model with the five baseline methods as shown in Figure 2. The sharp decrease of precision in the AUPR curve is probably because there is outlier in our constructed data. From the results, we can see: (1) GraphTF outperforms in both AUC and AUPR compared with the other five computational methods. This may attribute to the attention mechanism in GAT, which not only preserves the topology of similarity networks but also considers different weights of the neighborhood nodes. (2) Compared with CP, TDRC integrates miRNA functional and disease semantic similarity to constrain model optimization, while DrugCom merges miRNA functional similarity, miRNA sequence similarity and disease semantic similarity for training, which proves that adding multi-source information for tensor-based models can effectively improve the prediction performance. (3) DrugCom outperforms other deep learning models, which may be because DrugCom captures the consensus information between multiple auxiliary data and tensor to constrain factors, while the deep learning models simply take multiple data as input features. (4) For GraphTF, DeepSynergy and DTF, the GAT used in GraphTF can more efficiently capture the complex nonlinear relations well on both miRNA network and disease network than simple deep neural network.

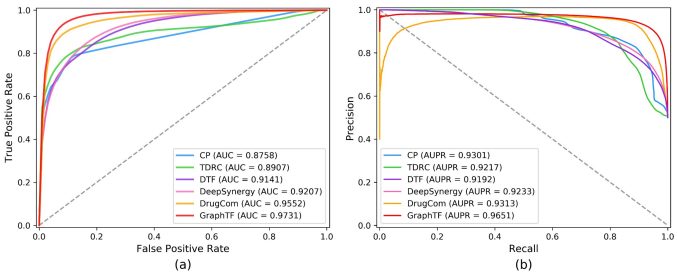

Fig. 2. (a)The ROC curve of GraphTF and compared methods. (b) The AUPR curve of GraphTF and compared methods.

D. Robustness of GraphTF

To verify the stability of GraphTF, we perform comparison experiments under different size of negative samples and known links, respectively. First, we randomly generate the negative samples with a size of $[n, 2n, 4n, 6n, 8n, 10n]$ to make comparison. The results are averaged over ten different negative samples of the same size. As Figure 3 shows, with the size of negative samples increasing, the AUPR of all methods decrease at a certain degree. However, the drop rate of our GraphTF is significantly lower than other methods, which indicates GraphTF is robust to imbalanced data.

To further test the robustness of GraphTF, we randomly remove the known triple links by 5%, 10%, 20%, 30% and 40%, i.e., the known miRNA-miRNA-disease associations are

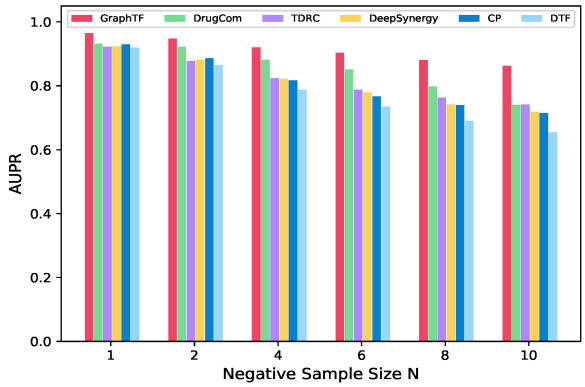

Fig. 3. The AUPR values with different negative sample sizes.

decreased to 14710, 13969, 10839 and 9291, the unknown triple associations are 16258, 16999, 20129 and 21677. From Table I, all the methods perform steadily under different data distributions, and GraphTF still outperforms all other algorithms, which may because of the advantages of sparse tensor factorization. In summary, GraphTF is resilient to both negative sample size and missing data. Therefore, it is more generalizable for real-world applications.

TABLE I
THE AUPR VALUES FOR ALL APPROACHES WHEN REMOVING ASSOCIATION DATA AT RATE OF 5%, 10%, 20%, 30% AND 40% COMPARED WITH ORIGINAL RESULTS.

Ratio	CP	TDRC	DeepSynergy	DTF	DrugCom	GraphTF
0%	0.9301	0.9217	0.9233	0.9192	0.9313	**0.9651**
5%	0.9326	0.9262	0.9189	0.9355	0.9355	**0.9653**
10%	0.9305	0.9258	0.9194	0.9563	0.9563	**0.9638**
20%	0.9386	0.9170	0.9203	0.9483	0.9483	**0.9665**
30%	0.9316	0.9202	0.9185	0.9593	0.9593	**0.9649**
40%	0.9388	0.9206	0.9193	0.9514	0.9514	**0.9635**

E. Ablation Study

To analyze the importance of multi-source data, we compare the results of our model leveraging different similarity graphs. Specifically, GraphTF-nseq means we only use miRNA functional similarity to learn miRNA features, GraphTF-nfunc denotes we only use miRNA sequence similarity to learn miRNA features, and GraphTF learns miRNA features exploiting both miRNA functional and sequence similarity graphs. As shown in Table II, GraphTF achieves the best performance using all multi-source data. Moreover, we can find that miRNA functional similarity contributes a little more than miRNA sequence similarity to GraphTF.

TABLE II
PERFORMANCE OF GRAPHTF WITH DIFFERENT SIMILARITY DATA.

Models	GraphTF-nseq	GraphTF-nfunc	GraphTF
AUC	0.9676	0.9660	**0.9731**
AUPR	0.9600	0.9586	**0.9651**

F. Parameters Analysis

In this section, we explore the impact of four crucial hyper-parameters for GraphTF. The results are shown in Figure 4.

- Effect of embedding size D. In this experiment, we fixed other parameters, where the neural projection layer $L = 3$, number of attention heads $U = 3$, dropout ratio $\rho = 0.4$, and changed the embedding size in $\{32, 64, 128, 256, 512\}$. From Figure 4(a), we can see that different embedding sizes would lead to different final results. A large size of embedding may incur overfitting problem and a small size of embedding may lead to underfitting. When $D = 128$, GraphTF achieves the best performance, which indicates that our model exploits much information when the embedding size is set as 128.
- Effect of neural projection layer L. The neural projection layers defined in Eq.(11), Eq.(12), Eq.(13) are adopted to enhance the learned embeddings of miRNAs and disease from GAT. We fixed other parameters and performed the experiments for $L = \{1, 2, 3, 4\}$ situations. Figure 4(b) shows that GraphTF achieves a peak with $L = 3$ and as the number of layers increases, the performance of GraphTF goes down. This may because of overfitting problem caused by information loss.
- Effect of attention heads U. Multi-head attention can jointly capture information from different subspaces. In our comparison, we keep all other parameters fixed and change the number of attention heads in $\{1, 2, 3, 4\}$. Figure 4(c) indicates that with the increase number of attention heads, the performance of GraphTF first improves and reaches a peak with $U = 3$ then decreases, which indicates that a proper number of attention heads can enhance the model performance.
- Effect of dropout ratio ρ. Dropout is an effective module in deep learning model to avoid overfitting. From the results in Figure 4(d), we can see that dropout can significantly help improve the performance of GraphTF. We choose $\rho = 0.4$ to achieve the best AUC.

G. Case Studies

To demonstrate the capability of GraphTF to predict novel miRNA-miRNA-disease associations, we carried out case studies on two common human complex diseases Breast Neoplasms and Lung Neoplasms. In particular, all known miRNA-miRNA-disease associations as training set for GraphTF to make predictions, and the unknown associations are treated as candidate set for validation. For a given triple association $< m_1,\ m_2,\ d >$, we validated it by verifying the decomposed pairwise associations i.e., $< m_1, m_2 >$, $< m_1, d >$, $< m_2,\ d >$. In particular, we verified the top 20 predictions with the prominent miRNA-disease association database dbDEMC 2.0 [26], and validated the miRNA-miRNA pairs by enrichment analysis. Table III, IV show the top-20 Breast Neoplasms and Lung Neoplasms related miRNA-miRNA pairs predicted by GraphTF, from which we can see that most disease related miRNA pairs are confirmed by dbDEMC or literature.

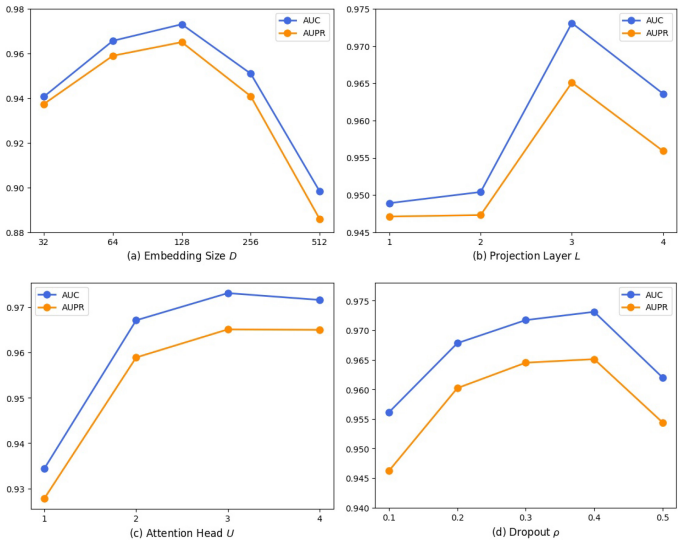

Fig. 4. (a) The performance on different embedding size D; (b) Comparison of predicted performance with different projection layers L; (c) Impact of number of attention heads U; (d) Influence of dropout ratio ρ.

TABLE III
TOP-20 BREAST NEOPLASMS RELATED MIRNA PAIRS AND THE CONFIRMATION FOR THEIR ASSOCIATIONS BY DBDEMC 2.0 AND LITERATURES. MIRNA I AND MIRNA J IN THE SAME ROW FORM A DISEASE-RELATED MIRNA PAIR.

miRNA i	Evidence	miRNA j	Evidence
hsa-mir-520h	dbDEMC	hsa-mir-302d	dbDEMC
hsa-mir-520h	dbDEMC	hsa-mir-323b	dbDEMC
hsa-mir-323a	dbDEMC	hsa-mir-323b	dbDEMC
hsa-mir-520h	dbDEMC	hsa-mir-409	PMID: 26631969
hsa-mir-323a	dbDEMC	hsa-mir-302d	dbDEMC
hsa-mir-323a	dbDEMC	hsa-mir-181d	dbDEMC
hsa-mir-323a	dbDEMC	hsa-mir-409	PMID: 26631969
hsa-mir-520f	**unconfirmed**	hsa-mir-302d	dbDEMC
hsa-mir-323a	dbDEMC	hsa-mir-509-2	PMID: 25659578
hsa-mir-382	PMID: 31635836	hsa-mir-302d	dbDEMC
hsa-mir-520h	dbDEMC	hsa-mir-509-2	PMID: 25659578
hsa-mir-382	PMID: 31635836	hsa-mir-323b	dbDEMC
hsa-mir-382	PMID: 31635836	hsa-mir-181d	dbDEMC
hsa-mir-519d	dbDEMC	hsa-mir-323b	dbDEMC
hsa-mir-323a	dbDEMC	hsa-mir-487b	**unconfirmed**
hsa-mir-519d	dbDEMC	hsa-mir-181d	dbDEMC
hsa-mir-520h	dbDEMC	hsa-mir-181d	dbDEMC
hsa-mir-382	PMID: 31635836	hsa-mir-409	PMID: 26631969
hsa-mir-302d	dbDEMC	hsa-mir-409	PMID: 26631969
hsa-mir-520h	dbDEMC	hsa-mir-1302-2	**unconfirmed**

To validate the biological significance of the predicted miRNA pairs, we perform the GO and KEGG enrichment analysis for the gene set which contains common genes among each miRNA pair and disease. From Table V, we find that most gene sets of the top-20 predicted triple associations are enriched either in GO-BP terms or KEGG pathway for Breast Neoplasms and Lung Neoplasms, respectively. Take <hsa-mir-1-1, hsa-mir-9-2> pair related to Lung Neoplasms

TABLE IV
TOP-20 LUNG NEOPLASMS RELATED MIRNA PAIRS AND THE CONFIRMATION FOR THEIR ASSOCIATION BY DBDEMC 2.0 AND LITERATURES. MIRNA I AND MIRNA J IN THE SAME ROW FORM A DISEASE-RELATED MIRNA PAIR.

miRNA i	Evidence	miRNA j	Evidence
hsa-mir-1-1	dbDEMC	hsa-mir-9-2	dbDEMC
hsa-mir-29c	dbDEMC	hsa-mir-19a	dbDEMC
hsa-mir-29b-2	dbDEMC	hsa-mir-19a	dbDEMC
hsa-mir-29c	dbDEMC	hsa-mir-1-1	dbDEMC
hsa-mir-29b-2	dbDEMC	hsa-mir-19b-1	dbDEMC
hsa-mir-29c	dbDEMC	hsa-mir-19b-1	dbDEMC
hsa-mir-222	dbDEMC	hsa-mir-1-1	dbDEMC
hsa-mir-520h	dbDEMC	hsa-mir-302d	dbDEMC
hsa-mir-29b-2	dbDEMC	hsa-mir-9-2	dbDEMC
hsa-mir-323a	dbDEMC	hsa-mir-181d	PMID: 30548184
hsa-mir-29b-2	dbDEMC	hsa-mir-1-1	dbDEMC
hsa-mir-519d	dbDEMC	hsa-mir-181d	PMID: 30548184
hsa-mir-17	dbDEMC	hsa-mir-19b-1	dbDEMC
hsa-mir-19b-1	dbDEMC	hsa-mir-9-2	dbDEMC
hsa-mir-17	dbDEMC	hsa-mir-19a	dbDEMC
hsa-mir-382	dbDEMC	hsa-mir-181d	PMID: 30548184
hsa-mir-222	dbDEMC	hsa-mir-19a	dbDEMC
hsa-mir-17	dbDEMC	hsa-mir-1-1	dbDEMC
hsa-mir-29b-1	dbDEMC	hsa-mir-1-1	dbDEMC
hsa-mir-382	dbDEMC	hsa-mir-302d	dbDEMC

as an example, we enumerate the significant GO-BP terms and KEGG pathways involved in this pair (see Figure 5). The majority of GO-BP terms are related to the cellular component, molecular function, and biological process, illustrating <hsa-mir-1-1, hsa-mir-9-2> pair is related to the normal development and complex disease. Furthermore, the 'Human T-cell leukemia virus 1 infection' and 'MicroRNAs in cancer' are the most significant KEGG pathways, which are closely associated with lung adenocarcinoma. Research [27] also demonstrated that miR-1 and miR-9 collectively impinged the epithelial-mesenchymal transition process. All these results show the synergism between hsa-mir-1-1 and hsa-mir-9-2, which proves the effectiveness of GraphTF in discovering disease-related miRNA-miRNA pairs.

TABLE V
THE ENRICHMENT RESULTS OF PREDICTED TOP-20 MIRNA PAIRS FOR BREAST NEOPLASMS AND LUNG NEOPLASMS.

GO/KEGG	Confirmed	Unconfirmed
Breast Neoplasms	16	4
Lung Neoplasms	18	2

V. CONCLUSION

Identification of potential disease-associated miRNA-miRNA pairs is a challenging but promising strategy for comprehensively understanding the pathogenesis of human diseases and can be a great alternative for disease treatment. In this paper, we propose a novel tensor-based model

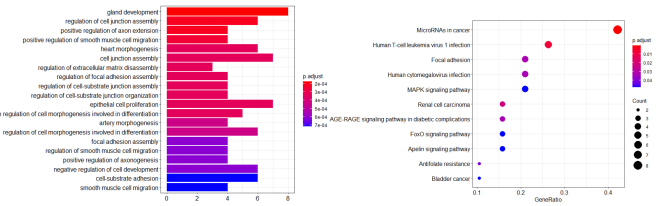

Fig. 5. (a) The GO-BP terms (top 20) of miRNA pair <hsa-mir-1-1, hsa-mir-9-2> related to Lung Neoplasms. (b) The KEGG pathways of <hsa-mir-1-1, hsa-mir-9-2> related to Lung Neoplasms.

GraphTF, which integrates a deep learning module for predicting miRNA-miRNA-disease ternary associations. First, we extract the features of miRNAs and diseases via a multi-head attention mechanism aggregating neighborhood information over multiple similarity network, then we take the obtained feature matrices as factor matrices decomposed from tensor to reconstruct the association tensor, and finally make predictions. Different from traditional tensor factorization framework, GraphTF exploits graph attention network to explore high-order nonlinear relationships between miRNAs and diseases. The experimental results through five-fold cross-validation show the effectiveness and robustness of GraphTF. Case studies also verify the powerful predictive ability of GraphTF in discovering disease-related miRNA pairs.

In the future, we plan to integrate more biological auxiliary information and explore more efficient methods to fuse the learned feature. More effort will be put into investigating other negative sample approaches for improving the model performance.

ACKNOWLEDGMENT

This work has been supported by the Natural Science Foundation of China (Grant No. 61873089, 62032007).

REFERENCES

[1] J Skommer, I Rana, FZ Marques, W Zhu, Z Du, and FJ Charchar. Small molecules, big effects: the role of micrornas in regulation of cardiomyocyte death. *Cell death & disease*, 5(7):e1325–e1325, 2014.

[2] Xuhai Zhao, Hongjiang Song, Zhigui Zuo, Yu Zhu, Xinshu Dong, and Xiangshi Lu. Identification of mirna–mirna synergistic relationships in colorectal cancer. *International journal of biological macromolecules*, 55:98–103, 2013.

[3] Emine Ezel Cilek, Hakime Ozturk, and Bala Gur Dedeoglu. Construction of mirna-mirna networks revealing the complexity of mirna-mediated mechanisms in trastuzumab treated breast cancer cell lines. *PLoS One*, 12(10):e0185558, 2017.

[4] Fumikazu Miwakeichi, Eduardo Martınez-Montes, Pedro A Valdés-Sosa, Nobuaki Nishiyama, Hiroaki Mizuhara, and Yoko Yamaguchi. Decomposing eeg data into space–time–frequency components using parallel factor analysis. *NeuroImage*, 22(3):1035–1045, 2004.

[5] Feng Huang, Xiang Yue, Zhankun Xiong, Zhouxin Yu, Shichao Liu, and Wen Zhang. Tensor decomposition with relational constraints for predicting multiple types of microrna-disease associations. *Briefings in bioinformatics*, 22(3):bbaa140, 2021.

[6] Pei Liu, Jiawei Luo, and Xiangtao Chen. mircom: Tensor completion integrating multi-view information to deduce the potential disease-related mirna-mirna pairs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–1, 2020.

[7] Cong Shen, Jiawei Luo, Wenjue Ouyang, Pingjian Ding, and Xiangtao Chen. Iddkin: network-based influence deep diffusion model for enhancing prediction of kinase inhibitors. *Bioinformatics*, 36(22-23):5481–5491, 2020.

[8] Zexuan Sun, Shujun Huang, Peiran Jiang, and Pingzhao Hu. Dtf: Deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics*, 36(16):4483–4489, 2020.

[9] Heli Julkunen, Anna Cichonska, Prson Gautam, Sandor Szedmak, Jane Douat, Tapio Pahikkala, Tero Aittokallio, and Juho Rousu. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nature communications*, 11(1):1–11, 2020.

[10] Peter S Linsley, Janell Schelter, Julja Burchard, Miho Kibukawa, Melissa M Martin, Steven R Bartz, Jason M Johnson, Jordan M Cummins, Christopher K Raymond, Hongyue Dai, et al. Transcripts targeted by the microrna-16 family cooperatively regulate cell cycle progression. *Molecular and cellular biology*, 27(6):2240–2252, 2007.

[11] Ewelina Perdas, Robert Stawski, Dariusz Nowak, and Maria Zubrzycka. The role of mirna in papillary thyroid cancer in the context of mirna let-7 family. *International journal of molecular sciences*, 17(6):909, 2016.

[12] Ahmet Ucar, Shashi K Gupta, Jan Fiedler, Erdem Erikci, Michal Kardasinski, Sandor Batkai, Seema Dangwal, Regalla Kumarswamy, Claudia Bang, Angelika Holzmann, et al. The mirna-212/132 family regulates both cardiac hypertrophy and cardiomyocyte autophagy. *Nature communications*, 3(1):1–11, 2012.

[13] Yang Wang, Xinwei Zhang, Hui Li, Jinpu Yu, and Xiubao Ren. The role of mirna-29 family in cancer. *European journal of cell biology*, 92(3):123–128, 2013.

[14] Ana Kozomara, Maria Birgaoanu, and Sam Griffiths-Jones. mirbase: from microrna sequences to function. *Nucleic acids research*, 47(D1):D155–D162, 2019.

[15] Zhou Huang, Jiangcheng Shi, Yuanxu Gao, Chunmei Cui, Shan Zhang, Jianwei Li, Yuan Zhou, and Qinghua Cui. Hmdd v3. 0: a database for experimentally supported human microrna–disease associations. *Nucleic acids research*, 47(D1):D1013–D1017, 2019.

[16] Qiu Xiao, Jiawei Luo, Cheng Liang, Jie Cai, and Pingjian Ding. A graph regularized non-negative matrix factorization method for identifying microrna-disease associations. *Bioinformatics*, 34(2):239–248, 2018.

[17] Saul B Needleman and Christian D Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, 1970.

[18] Dong Wang, Juan Wang, Ming Lu, Fei Song, and Qinghua Cui. Inferring the human microrna functional similarity and functional network based on microrna-associated diseases. *Bioinformatics*, 26(13):1644–1650, 2010.

[19] Jin Li, Sai Zhang, Tao Liu, Chenxi Ning, Zhuoxuan Zhang, and Wei Zhou. Neural inductive matrix completion with graph convolutional networks for mirna-disease association prediction. *Bioinformatics*, 36(8):2538–2546, 2020.

[20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

[22] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[23] Huiyuan Chen and Jing Li. Drugcom: synergistic discovery of drug combinations using tensor decomposition. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 899–904. IEEE, 2018.

[24] Kristina Preuer, Richard PI Lewis, Sepp Hochreiter, Andreas Bender, Krishna C Bulusu, and Günter Klambauer. Deepsynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics*, 34(9):1538–1546, 2018.

[25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[26] Zhen Yang, Liangcai Wu, Anqiang Wang, Wei Tang, Yi Zhao, Haitao Zhao, and Andrew E Teschendorff. dbdemc 2.0: updated database of differentially expressed mirnas in human cancers. *Nucleic acids research*, 45(D1):D812–D818, 2017.

[27] Francesca Citron, Joshua Armenia, Giovanni Franchin, Jerry Polesel, Renato Talamini, Sara D'Andrea, Sandro Sulfaro, Carlo M Croce, William Klement, David Otasek, et al. An integrated approach identifies mediators of local recurrence in head and neck squamous carcinoma. *Clinical Cancer Research*, 23(14):3769–3780, 2017.