Information Fusion 63 (2020) 153–165

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Attributed heterogeneous network fusion via collaborative matrix tri-factorization

Guoxian Yu[a,b,c], Yuehui Wang[c], Jun Wang[a,*], Carlotta Domeniconi[d], Maozu Guo[e], Xiangliang Zhang[b]

[a] School of Software, Shandong University, Jinan, China
[b] CEMSE, King Abdullah University of Science and Technology, Thuwal, SA
[c] College of Computer and Information Sciences, Southwest University, Chongqing, China
[d] Department of Computer Science, George Mason University, Fairfax, VA, USA
[e] College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China

ARTICLE INFO

Keywords:
Attributed heterogeneous networks
Matrix factorization
Data fusion
Insufficient relations
LncRNA-disease associations

ABSTRACT

Heterogeneous network based data fusion can encode diverse inter- and intra-relations between objects, and has been sparking increasing attention in recent years. Matrix factorization based data fusion models have been invented to fuse multiple data sources. However, these models generally suffer from the widely-witnessed insufficient relations between nodes and from information loss when heterogeneous attributes of diverse network nodes are transformed into ad-hoc homologous networks for fusion. In this paper, we introduce a general data fusion model called Attributed Heterogeneous Network Fusion (AHNF). AHNF firstly constructs an attributed heterogeneous network composed with different types of nodes and the diverse attribute vectors of these nodes. It uses indicator matrices to differentiate the observed inter-relations from the latent ones, and thus reduces the impact of insufficient relations between nodes. Next, it collaboratively factorizes multiple adjacency matrices and attribute data matrices of the heterogeneous network into low-rank matrices to explore the latent relations between these nodes. In this way, both the network topology and diverse attributes of nodes are fused in a coordinated fashion. Finally, it uses the optimized low-rank matrices to approximate the target relational data matrix of objects and to effectively accomplish the relation prediction. We apply AHNF to predict the lncRNA-disease associations using diverse relational and attribute data sources. AHNF achieves a larger area under the receiver operating curve 0.9367 (by at least 2.14%), and a larger area under the precision-recall curve 0.5937 (by at least 28.53%) than competitive data fusion approaches. AHNF also outperforms competing methods on predicting *de novo* lncRNA-disease associations, and precisely identifies lncRNAs associated with breast, stomach, prostate, and pancreatic cancers. AHNF is a comprehensive data fusion framework for universal attributed multi-type relational data. The code and datasets are available at http://mlda.swu.edu.cn/codes.php?name=AHNF.

1. Introduction

The rapid development and wide application of high-throughput techniques lead to the production of different types of data that are directly related to the main task, and also other data sources indirectly related to the task but still helpful for the completion of this task. For example, in Fig. 1, the association prediction tasks between lncRNAs (long non-coding RNAs) and diseases, between genes and Gene Ontology functional terms can be more reliably made by fusing the gene-level, transcript-level and the diverse interactions among molecules, which convey complementary biological knowledge about the functions of lncRNAs or genes. Given that, data fusion approaches have been increasing popular over the past decades, they have been widely applied in various domains to collectively explore interesting patterns from heterogeneous data sources, to remedy the insufficient information of individual data sources, and to reduce the impact to noisy or irrelevant data sources [1–3].

Existing data fusion-based models can be generally classified into three categories: *early*, *intermediate*, and *late*. The early integration models generally concatenate the original (or transformed) attributes into a single long attribute vector before fitting a unsupervised or supervised model [4–6]. But this concatenation ignores the intrinsic characteristics of different attribute types, and may suffer from the curse of dimensionality. Late integration models firstly learn individual models for different

* Corresponding author.
E-mail addresses: yhwang@email.swu.edu.cn (Y. Wang), kingjun@sdu.edu.cn (J. Wang), carlotta@cs.gmu.edu (C. Domeniconi), guomaozu@bucea.edu.cn (M. Guo), xiangliang.zhang@kaust.edu.sa (X. Zhang).

https://doi.org/10.1016/j.inffus.2020.06.012
Received 17 September 2019; Received in revised form 23 May 2020; Accepted 21 June 2020
Available online 26 June 2020

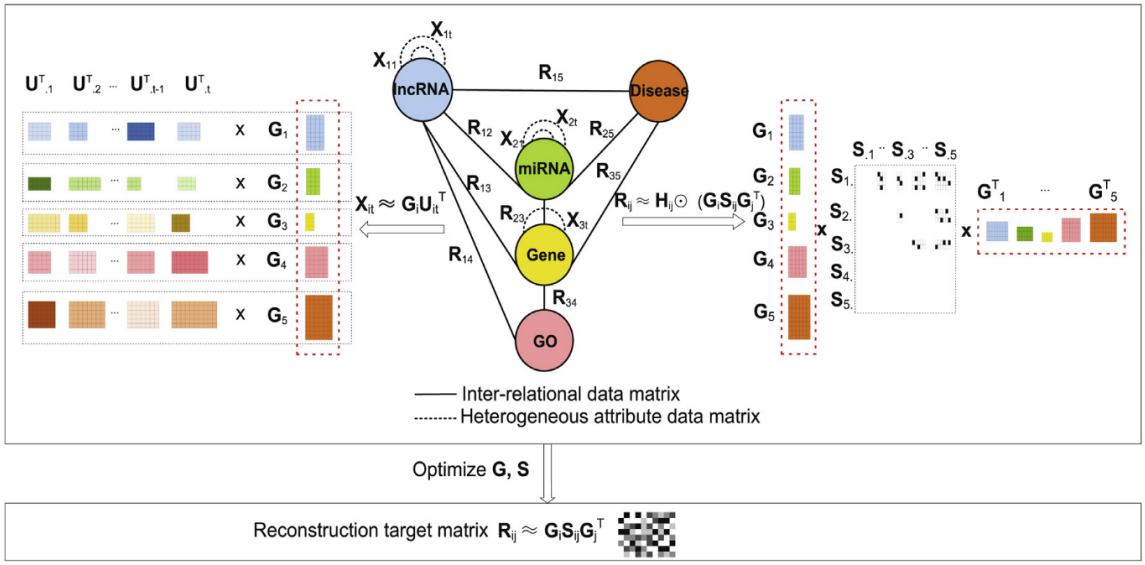


Fig. 1. The operating principle of AHNF. $\mathbf{R}_{ij}$ is the relational data matrix between object types i and j, $\mathbf{X}_{it}$ is the tth attribute data matrix of the ith object type, $\mathbf{G}_i$ is the low-rank representation matrix of the ith object type. The inter-associations between objects can be predicted by $\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T$.

data sources, and then combine the outputs of these models for ensemble prediction [7–10]. This ensemble paradigm, though easy to implement, may be compromised by low-quality base models independently trained on single data sources. For this reason, intermediate models are more popular in various data fusion tasks [1,11–13]. Intermediate models can combine different data sources in a single model, and simultaneously perform the prediction.

Multi-view learning, as an important intermediate data fusion paradigm, generally focuses on a single type of objects and requires the same number of objects across multiple data sources (views) [14–16]. As such, it can not sufficiently make use of other indirectly related data sources. On the other hand, heterogeneous network based intermediate approaches can encode the inter-relations between different object types, and the intra-relational subnetworks of objects derived from different data sources, they can accommodate diverse types of objects and indirectly related data sources [3,11,17]. These approaches generally project heterogeneous data sources onto homologous networks (kernels), and then follow the principle of multiple kernel (network) learning [18,19] to fuse the networks and to make prediction [20,21]. Based on this network, and the assumption that similar objects are associated with similar labels and vice versa, different network-based inference techniques have been developed to infer associations between nodes [20–22]. However, these models typically have to project multi-relational data with objects onto the homologous intra-relational networks of respective objects for data fusion. This *hand-crafted* projection may enshroud the intrinsic structure of multi-relational data, and thus does not make the best usage of them [1].

Matrix factorization based data fusion models factorize the relational data matrices of the heterogeneous network into low-rank matrices to explore latent relationships between network nodes, and to fuse multiple relational data sources, without projecting them onto homogeneous networks [11]. For this advantage, they have been extensively studied in recent years. To name a few, Wang et al. [23] proposed a symmetric nonnegative matrix tri-factorization approach (S-NMTF) to achieve clustering on multi-type relational data sources. Zitnik and Zupan [11] developed a penalized matrix tri-factorization based model (DFMF) to fuse multiple relational data matrices for predicting gene functions and pharmacologic actions. Fu et al. [24] extended matrix tri-factorization (MFLDA) to predict lncRNA-disease associations by selectively fusing inter-relational data matrices. Yu et al. [25] improved MFLDA to predict lncRNA-disease associations by further differentiating the relevance of multiple intra-relational data matrices. Wang et al. [26] proposed a selective matrix tri-factorization (SelMFDF) to avoid the preference toward sparse relational data matrices. Lu et al. [27] introduced an inductive matrix completion based approach with the fusion of different attribute kernels of lncRNAs (or disease). Biswas et al. [28] proposed a robust inductive matrix completion strategy using an $\ell_{2,1}$ norm penalty function to fuse data sources.

However, all these network-based data fusion techniques [11,26–28] still have *two* major limitations: (i) The fusion of the diverse attributes of network nodes is achieved by converting the attribute vectors into homologous networks via various similarity metrics. This leads to suboptimal results (as will show in our experiments). (ii) They implicitly and optimistically assume that the observed associations between objects are 'complete'. On the contrary, they are usually incomplete. As such, their performance is compromised by insufficient observed associations. To address these limitations, we propose the Attributed Heterogeneous Network Fusion (AHNF) approach and illustrate the overall procedure of AHNF in Fig. 1. AHNF firstly constructs an attributed heterogeneous network composed with different types of nodes (i.e., lncRNAs and Disease Ontology terms), along with the diverse attribute vectors of these nodes. It differentiates the observed relations from the other ones by means of indicator matrices, thus remedying the impact of insufficient relations between nodes. Next, it collaboratively factorizes multiple inter(intra)-relational adjacency matrices of the heterogeneous network, along with the indicator matrices, into low-rank matrices to explore the latent relations between these nodes. In addition, the low-rank matrices are also collaboratively factorized with respect to the heterogeneous attribute data matrices of multi-type nodes. In this way, both the network topology and diverse features of nodes are fused, and both the impact of insufficient relations and information loss are accounted for. Finally, it uses the optimized low-rank matrices to approximate the relational data matrix, and thus accomplishes the association prediction.

We apply AHNF to fuse multiple heterogeneous biological data sources for predicting lncRNA-disease associations, which is a practical, important and challenging topic in biomedical data mining [29,30]. We compare it with related and competitive methods, including S-NMTF [23], DFMF [11], MFLDA [24], WMFLDA [25], SelMFDF [26], SIMCLDA

[27], and RIMC [28]. AHNF obtains an AUROC (area under the receiver-operating-characteristic curve) of 0.9367 and an AUPRC (area under the precision-recall curve) of 0.5937, which are significantly better than those of comparing methods. In the *de novo* lncRNA-disease association prediction, AHNF again achieves higher values of AUROC and AUPRC than the competitive comparing methods. The case studies on breast, stomach, prostate and pancreatic cancers confirm that our predictions are supported by evidence from the bio-medical literature. In addition, the experiments confirm that both the consideration of insufficient relations and of information loss caused by data transformation contributes to a better performance. In summary, our diverse experiments confirm the effectiveness and the potential value of AHNF in identifying potential lncRNA-disease associations. AHNF is a comprehensive data fusion framework for universal attributed multi-type relational data and readily available for various prediction tasks, such as drug re-purposing and user-item recommendation.

The rest of the paper is organized as follows. We elaborate on the formulation of AHNF in Section 2, and detail the experimental protocol and analyze the experimental results in Section 3. Section 4 concludes our work along with directions for future pursue.

2. Attributed heterogeneous networks fusion

Given a heterogeneous network (as shown in Fig. 1) composed with m types of nodes (i.e., lncRNAs or diseases), the data information contained in this heterogeneous network typically presents in two forms as follows:

- **Relational data matrices** characterize the relations between data objects from different types. A collection of relational data matrices $\mathcal{R}$, $\mathbf{R}_{ij} \in \mathcal{R}$ $\left(\mathbf{R}_{ij} \in \mathbb{R}^{n_i \times n_j}\right)$ is an association matrix which relates n_i objects of type i with n_j objects of type j. Without loss of generality, if $s \in \{1, 2, \cdots, n_i\}$ and $t \in \{1, 2, \cdots, n_j\}$ have a known relation, then $\mathbf{R}_{ij}(s, t) > 0$; otherwise, $\mathbf{R}_{ij}(s, t) = 0$.
- **Attribute data matrices** characterize the attribute information of objects. Diverse attribute data matrices for the ith type of objects are encoded as $\mathbf{X}_{it} \in \mathbb{R}^{n_i \times d_{it}}$, $t \in \{1, 2, \cdots, t_i\}$ where t_i is the number of attribute matrices for the ith object type, and d_{it} is the number of attributes for the tth attribute data matrix.

The rich structures of multi-type heterogeneous network provide a potential opportunity to improve the prediction accuracy, which, however, also present a new challenge on how to take advantage of all available information. In this paper, we address this problem via collaborative matrix tri-factorization. Our goal is to make full use of the structure and attribution information of a multi-type heterogeneous network.

2.1. Matrix factorization based data fusion

Matrix factorization techniques recently have been widely used for integrating multi-type relational data [1,11,15,24,27]. Among them, matrix tri-factorization can accommodate multi-type relational data matrices. Matrix tri-factorization does not require the matrices having the same type (and number) of objects, nor it requires projecting them into homologous ones for fusion [31]. Zitnik and Zupan [11] introduced a representative data fusion model based on matrix tri-factorization (DFMF) as follows:

$$\min_{\mathbf{G}\geq 0} \mathcal{Z}(\mathbf{G}, \mathbf{S}) = \sum_{i,j=1}^{m} \left\|\mathbf{R}_{ij} - \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T\right\|_F^2 + \sum_{t=1}^{\max_i t_i} tr\left(\mathbf{G}^T \mathbf{\Theta}^{(t)} \mathbf{G}\right) \tag{1}$$

where $\mathbf{G}_i \in \mathbb{R}^{n_i \times k_i}$, $\mathbf{G}_j \in \mathbb{R}^{n_j \times k_j}$, $\mathbf{S}_{ij} \in \mathbb{R}^{k_i \times k_j}$, $(k_i \ll n_i,\ k_j \ll n_j)$, $\mathbf{G} = diag\left(\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_m\right)$. $tr(\cdot)$ and $\|\cdot\|_F^2$ are the matrix trace operator and the Frobenius norm, respectively. $\mathbf{S}_{ij}$ is much smaller than $\mathbf{R}_{ij}$ and it can be viewed as a compressed matrix that encodes latent inter-relations between two object types. $\mathbf{G}_i$ can be viewed as the low-rank representation of the ith type objects. To accommodate diverse attribute data matrices $\mathbf{X}_{it}$, Eq. (1) handcrafts them into multiple homologous networks $\mathbf{\Theta}_i^{(t)} \in \mathbb{R}^{n_i \times n_i}$. $\mathbf{\Theta}^{(t)}$ collectively stores all the following block diagonal matrices: $\mathbf{\Theta}^{(t)} = diag(\mathbf{\Theta}_1^{(t)}, \mathbf{\Theta}_2^{(t)}, \cdots, \mathbf{\Theta}_m^{(t)})$, $t \in (1, 2, \cdots, \max_i t_i)$, and the ith block matrix along the main diagonal of $\mathbf{\Theta}^{(t)}$ is zero if $t > t_i$.

Eq. (1) works directly on the heterogeneous network without performing any projection of subnetworks. This is in contrast with other network based data fusion approaches [21,22,27,28,32], which first project target object (i.e., lncRNAs or diseases) related data sources onto homologous (intra-relational) networks, and then merge the homologous networks for predicting the inter (or intra) relations between objects. Eq. (1) can explore and employ the intrinsic relation of multi-type objects. In addition, it optimizes $\mathbf{R}_{ij}$ with respect to $\mathbf{G}_i$ and $\mathbf{G}_j$, which are also coordinated by other data sources. As a result, multi-type relational data matrices are fused in a collaborative fashion.

Although the extensions of DFMF [24–26] can reduce the impact of irrelevant relational data matrices, alike matrix completion based solutions [27,28], they implicitly assume the observed relations ($\mathbf{R}_{ij}(s, t) = 1$) are 'complete' by *equally* treating all elements of $\mathbf{R}_{ij}$. But, in fact, $\mathbf{R}_{ij}$ is typically quite incomplete. Furthermore, these solutions need a handcraft projection to convert attribute data matrices into homologous similarity networks $\mathbf{\Theta}_i^{(t)}$, which causes information loss. In practice, these issues are suffered by most matrix factorization or network (kernel) based data fusion solutions [17–19,32–34]. As such, both issues reduce the reliability of data fusion based prediction. As a result, these approaches have a compromised performance and do not make full use of the *attributed heterogeneous network*, which can universally represent diverse interconnected heterogeneous data sources.

To make full use of the attributed heterogeneous network, we introduce a framework called AHNF and illustrate its operating principle in Fig. 1. AHNF firstly constructs an attributed heterogeneous network composed of different types of object nodes and diverse attributes associated to the nodes. It then collaboratively factorizes relational adjacency matrices and diverse attribute data matrices, along with the indicator matrices, into low-rank matrices to explore the latent relations between these nodes. After this, it approximates the target relational data matrix using the optimized low-rank matrices, and thus completes the prediction of associations between nodes.

2.2. Attributed heterogeneous network fusion

To avoid the information loss caused by projecting diverse attribute data matrices and to differentiate the observed relations from other unobserved ones (entries with zero values in $\mathbf{R}_{ij}$), we formulate the attributed heterogeneous network based data fusion as follows:

$$\begin{aligned}\min_{\mathbf{G}\geq 0} \mathcal{J}\left(\mathbf{G}, \mathbf{S}, \boldsymbol{\omega}^r, \boldsymbol{\omega}^h\right) = &\sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r \left\|\mathbf{H}_{ij} \odot \left(\mathbf{R}_{ij} - \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T\right)\right\|_F^2 \\ &+ \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h \left\|\mathbf{X}_{it} - \mathbf{G}_i \mathbf{U}_{it}^T\right\|_F^2 \\ s.t.\quad &\boldsymbol{\omega}^r \geq 0, \boldsymbol{\omega}^h \geq 0, \sum_{i,j=1}^{m} \omega_{ij}^r = 1, \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \omega_{it}^h = 1\end{aligned} \tag{2}$$

where $\odot$ is the Hadamard product (element-wise product), $\mathbf{H}_{ij}$ is the indicator matrix for $\mathbf{R}_{ij}$ and with the same size as $\mathbf{R}_{ij}$. $\mathbf{H}_{ij}(s, t) = 1$ if $\mathbf{R}_{ij}(s, t) > 0$, and $\mathbf{H}_{ij}(s, t) = 0$ otherwise. $\mathbf{G}_i$ is the network structure and attribute information co-guided low-rank representation of n_i objects, and $\mathbf{U}_{it} \in \mathbb{R}^{k_i \times d_{it}}$. Each of $\boldsymbol{\omega}^r \in \mathbb{R}^{m \times m}$ and $\boldsymbol{\omega}^h \in \mathbb{R}^{m \times \max_i t_i}$ includes the weights assigned to $|\mathcal{R}|$ inter-relational data matrices and to diverse attribute data matrices. For $\mathbf{R}_{ij} \notin \mathcal{R}$, $\omega_{ij}^r = 0$. For $\mathbf{X}_{it}$, if $t > \max_i t_i$, $\omega_{it}^h = 0$. $\left\|\mathbf{H}_{ij} \odot \left(\mathbf{R}_{ij} - \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T\right)\right\|_F^2$ is incorporated to differentiate the observed relations from others. It can enforce the preservation of the observed associations in the approximated $\mathbf{R}_{ij}$, and can be viewed as the approximation loss. Unlike Eqs. (1), (2) directly factorizes diverse attribute data matrices without converting them into intra-relational matrices.

However, Eq. (2) prefers to set $\boldsymbol{\omega}_{ij}^r = 1$ to $\mathbf{R}_{ij}$, when $\mathbf{R}_{ij}$ has the smallest approximation loss among all the relational matrices, and the other relational matrices will be disregarded. Eq. (2) also likes to assign $\boldsymbol{\omega}_{it}^h = 1$ to $\mathbf{X}_{it}$, if it has the smallest approximation loss among all the attribute matrices. In other words, the sparser the relational (or attribute) matrix is, the larger the weight assigned to it. As a result, the contribution of other relational (attribute) data matrices is disregarded.

Different relational (attribute) data matrices generally provide complementary information, and their fusion can give a more comprehensive view of the target objects. To remedy the trivial weight assignment and to selectively fuse multiple data sources, we add two l_2-norm based regularization terms on $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$, and update Eq. (2) as follows:

$$\begin{aligned}\min_{\mathbf{G}\geq 0} \mathcal{J}\left(\mathbf{G},\mathbf{S},\boldsymbol{\omega}^r,\boldsymbol{\omega}^h\right) &= \sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r \left\| \mathbf{H}_{ij} \odot \left(\mathbf{R}_{ij} - \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T\right)\right\|_F^2 \\ &+ \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h \left\| \mathbf{X}_{it} - \mathbf{G}_i \mathbf{U}_{it}^T \right\|_F^2 + \alpha \|vec(\boldsymbol{\omega}^r)\|_F^2 + \beta \left\| vec\left(\boldsymbol{\omega}^h\right)\right\|_F^2 \\ s.t. \quad \boldsymbol{\omega}^r \geq 0, \boldsymbol{\omega}^h \geq 0, &\sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r = 1, \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h = 1 \end{aligned} \tag{3}$$

where $vec(\boldsymbol{\omega}^r)$ is the vectorization operator that stacks the rows of $\boldsymbol{\omega}^r$; $\alpha > 0$ and $\beta > 0$ are used to control the complexity of $vec(\boldsymbol{\omega}^r)$ and $vec(\boldsymbol{\omega}^h)$. α and β can also help to selectively integrate relational matrices and attribute data matrices.

The sparse indicator matrices $\mathbf{H}_{ij}$ differentiate the observed relations from others, but they leave other unobserved relations at random to some extent and thus result in noisy relations of the target matrix. Given that, we further add an l_2-norm based regularization and formalize the final objective function of AHNF as follows:

$$\begin{aligned}\min_{\mathbf{G}\geq 0} \mathcal{J}\left(\mathbf{G},\mathbf{S},\boldsymbol{\omega}^r,\boldsymbol{\omega}^h\right) &= \sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r \left\| \mathbf{H}_{ij} \odot \left(\mathbf{R}_{ij} - \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T\right)\right\|_F^2 \\ &+ \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h \left\| \mathbf{X}_{it} - \mathbf{G}_i \mathbf{U}_{it}^T \right\|_F^2 \\ &+ \alpha \|vec(\boldsymbol{\omega}^r)\|_F^2 + \beta \left\| vec\left(\boldsymbol{\omega}^h\right)\right\|_F^2 + \gamma \sum_{i,j=1}^{m} \left\| \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T \right\|_F^2 \\ s.t.\ \boldsymbol{\omega}^r \geq 0, \boldsymbol{\omega}^h \geq 0, &\sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r = 1, \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h = 1 \end{aligned} \tag{4}$$

In Eq. (4), $\left\| \mathbf{G}_i \mathbf{S}_{ij} \mathbf{G}_j^T \right\|_F^2$ is the added constraint to control the noisy associations, and $\gamma > 0$ is the scalar parameter to balance its effect.

The objective function of AHNF is non-convex in $\mathbf{G}$, $\mathbf{S}$, $\boldsymbol{\omega}^h$, and $\boldsymbol{\omega}^r$ altogether. We can optimize Eq. (4) according to the idea of ADMM (alternating direction method of multipliers) [35], which was used to compute the approximate matrix tri-factorization [11,24]. For $\mathbf{G}$, $\mathbf{S}$, $\boldsymbol{\omega}^h$, and $\boldsymbol{\omega}^r$, we alternatively consider three of them as constant and optimize the fourth one. The iterative optimization procedure is provided in Appendix A.

The time complexity of optimizing low-rank representation $\mathbf{G}_i$ of the ith type objects is $\mathcal{O}(n_i k^2)$. In each iteration, there are m types of nodes to be integrated, so the time complexity for optimizing $\mathbf{G}$ is $\mathcal{O}(\sum_{i=1}^{m} n_i k^2)$. The time complexity of optimizing $\mathbf{S}$ is $\mathcal{O}(\sum_{i,j=1}^{m} n_i n_j k)$. As to $\mathbf{U}$, its time complexity is $\mathcal{O}(\sum_{i=1}^{m} \sum_{t=1}^{max_i t_i} n_i d_{it} k)$. The complexity for seeking weight $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$ are both $\mathcal{O}(m^2)$. Therefore, the overall time complexity of AHNF is

$$\mathcal{O}(T(\sum_{i=1}^{m} n_i k^2 + \sum_{i,j=1}^{m} n_i n_j k + \sum_{i=1}^{m} \sum_{t=1}^{max_i t_i} n_i d_{it} k + m^2))$$

T is the number of iterations to reach the convergence. In our study, since the adopted multi-type relational data matrices and the diverse attribute data matrices are all sparse, the actual time costs of the above operations can be further reduced.

3. Results and discussion

3.1. Experimental setup

To study the performance of AHNF, we consider 5 object types: lncRNAs (Type 1), miRNAs (Type 2), genes (Type 3), Gene Ontology (Type 4), and diseases (Type 5) and apply it to predict lncRNA-disease associations, a fundamental and challenging topic in functional genome [29,30,36]. We collect eight relational matrices between these objects from public databases and list the details of the multi-type relational data sources in Table 1. We also collect 36 attribute data sources of lncRNAs, miRNAs, and genes, and gives the details of these attribute data of objects in Table 2. Particularly, we collect 12 attributed data matrices for lncRNAs, including lncRNA expression profiles, lncRNA methylation profiles, lncRNA-single nucleotide polymorphisms (SNPs) associations and so on. We also collect 5 attributed data matrices for miRNAs, including miRNAs expression profiles and miRNA-transcription factor (TF) interactions. We further collect 9 attributed data matrices for genes, including gene expression profiles in different experiments and gene interactions. All these relational and attribute data sources are shared at http://mlda.swu.edu.cn/codes.php?name=AHNF.

We compare AHNF against seven related and representative data fusion methods, including S-NMTF [23], DFMF [11], MFLDA [24], WMFLDA [25], SelMFDF [26], SIMCLDA [27], and RIMC [28]. The first five comparing methods predict target associations by integrating multiple inter(intra)-relational data matrices and matrix tri-factorization. The last two adopts matrix completion to complete the potential lncRNA-disease associations with the handcrafted attribute data matrix of lncRNAs and diseases. The input parameters of these methods are set as specified by the authors in their code, or optimized in the suggested ranges. $\alpha = 10^5$, $\beta = 10^3$ and $\gamma = 10^2$ are used for AHNF. These parameters will be analyzed in the next subsections. We adopt the widely-used area under the receiver-operating-characteristics curve (AUROC) and area under the precision-recall curve (AUPRC) to quantify the prediction results of the methods.

3.2. LncRNA-disease association prediction with cross validation

We perform five-fold cross validation (with 10 repetitions) on experimentally confirmed lncRNA-disease associations to study the performance of AHNF. Particularly, we randomly divide known lncRNA-disease associations ($\mathbf{R}_{15}$) into five folds; the associations in four folds are used as training samples and the remaining associations of the other fold are alternatively used as testing samples for evaluation. The ROC (PR) curves of the comparing methods, along with their AUROC and AUPRC values are given in Fig. 2. AHNF manifests the highest TPRs under the same FPRs, and has the highest AUROC (0.9367) among the comparing methods. The AUROC values of S-NMTF, DFMF, MFLDA, WMFLDA, SelMFDF, SIMCLDA, and RIMC are 0.7862, 0.8236, 0.8398, 0.8852, 0.9171, 0.8212, and 0.8079, respectively. AHNF improves the AUROC by at least 2.14%. As for the PR curves and AUPRC values, AHNF again has the highest AUPRC (0.5937), and it improves the AUPRC by at least 28.53%. The performance margin between AHNF and other comparing methods with respect to AUPRC is more prominent. This is because the known lncRNA-disease associations are scanty; predicting the lncRNA-disease associations is a class-imbalanced data mining task, and AUPRC is more sensitive to class-imbalance than AUROC.

SelMFDF, WMFLDA, MFLDA, DFMF and S-NMTF all aim at integrating multi-type relational data matrices, but they are statistically (pairwise t-test with p-value=0.01) outperformed by AHNF. The main reason is that they implicitly assume the observed relations are 'complete', and need to convert heterogeneous attribute vectors of nodes into homologous networks before fusion. AHNF, SelMFDF, and WMFLDA account for the different importance levels of inter(intra)-relational matrices and achieve a better performance than other competitive methods, which equally treat all the inter(intra)-relational data matrices.

Table 1
Details of the collected multi-type relational data sources.

Datasets	Size	#Associations		Sources
LncRNA-miRNA	240 × 495	1002	$\mathbf{R}_{12}$	http://starbase.sysu.edu.cn/mirLncRNA.php/
LncRNA-Gene	240 × 15527	6186	$\mathbf{R}_{13}$	http://www.lncrna2target.org/
LncRNA-GO	240 × 6428	3094	$\mathbf{R}_{14}$	ftp.ncbi.nih.gov/gene/GeneRIF/
LncRNA-Disease	240 × 412	2697	$\mathbf{R}_{15}$	http://www.rnanut.net/lncrnadisease/
miRNA-Gene	495 × 15527	135,852	$\mathbf{R}_{23}$	http://mirtarbase.mbc.nctu.edu.tw/
miRNA-Disease	495 × 412	13,562	$\mathbf{R}_{25}$	http://www.cuilab.cn/hmdd/
Gene-GO	15527 × 6428	1,191,503	$\mathbf{R}_{34}$	http://geneontology.org/
Gene-Disease	15527 × 412	115,317	$\mathbf{R}_{35}$	http://www.disgenet.org/

Table 2
Details on the collected features matrices from different data sources.

Object Type	Size	#nonzero entries		Feature Type	Sources
LncRNAs	240 × 53	844	$\mathbf{X}_{11}$	lncRNA expression	http://bigd.big.ac.cn/lncbook/index
	240 × 32	3057	$\mathbf{X}_{12}$	profiles	
	240 × 7	1253	$\mathbf{X}_{13}$	lncRNA methylation	
	240 × 6	1080	$\mathbf{X}_{14}$	profiles	
	240 × 3	1253	$\mathbf{X}_{15}$		
	240 × 7	1263	$\mathbf{X}_{16}$		
	240 × 6	1078	$\mathbf{X}_{17}$		
	240 × 5	901	$\mathbf{X}_{18}$		
	240 × 3	540	$\mathbf{X}_{19}$		
	240 × 5	899	$\mathbf{X}_{110}$		
	240 × 6	1079	$\mathbf{X}_{111}$		
	240 × 68851	69,751	$\mathbf{X}_{112}$	lncRNA-SNP interactions	http://bioinfo.life.hust.edu.cn/lncRNASNP/
miRNAs	495 × 474	46,843	$\mathbf{X}_{21}$	miRNA similarity	http://www.lirmed.com/misim/
	495 × 112	999	$\mathbf{X}_{22}$	miRNA-TF interactions	https://doi.org/10.1371/journal.pone.0152792
	495 × 135	29,028	$\mathbf{X}_{23}$	miRNA expression	http://guanlab.ccmb.med.umich.edu/mirmine/
	495 × 39	9321	$\mathbf{X}_{24}$	profiles	http://doi.org/10.1186/1471-2164-8-166
	495 × 22	32	$\mathbf{X}_{25}$	miRNA-miRNA interactions	10.1016/j.gene.2012.09.066
	15527 × 13	18,068	$\mathbf{X}_{31}$	gene expression	https://www.ebi.ac.uk/gxa/experiments/E-MTAB-4344/
	15527 × 53	7504	$\mathbf{X}_{32}$	profiles	https://www.ebi.ac.uk/gxa/experiments/E-MTAB-5214/
	15527 × 83	173,264	$\mathbf{X}_{33}$		https://www.ebi.ac.uk/gxa/experiments/E-PROT-3/
	15527 × 2055	4551	$\mathbf{X}_{34}$	protein-protein	http://dip.doe-mbi.ucla.edu/dip/Main.cgi
Genes	15527 × 6170	32,097	$\mathbf{X}_{35}$	interactions	http://hprd.org/index_html
	15527 × 12826	283,306	$\mathbf{X}_{36}$		http://ophid.utoronto.ca/ophidv2.204/index.jsp
	15527 × 10559	113973	$\mathbf{X}_{37}$		http://www.ebi.ac.uk/intact
	15527 × 6170	32,097	$\mathbf{X}_{38}$		http://mint.bio.uniroma2.it
	15527 × 12471	223,546	$\mathbf{X}_{39}$		http://thebiogrid.org

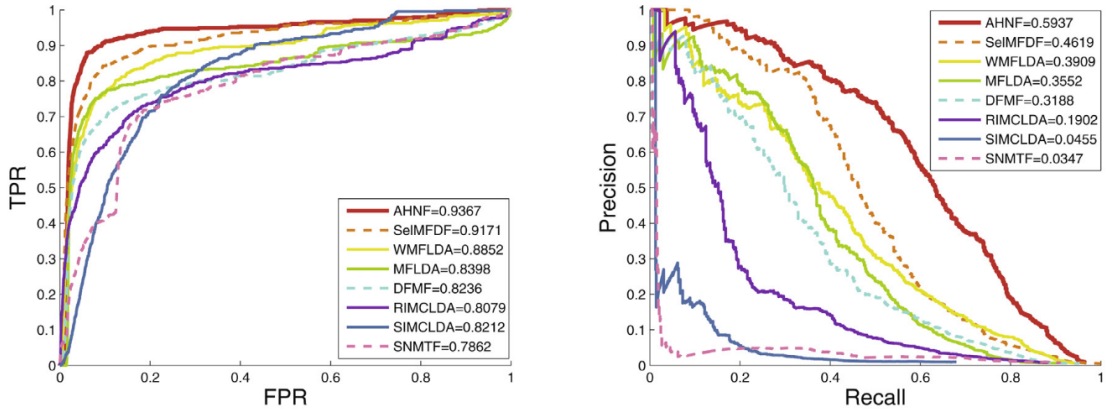
These results demonstrate the effectiveness of separately weighting the inter-relational data matrices and attribute data matrices in AHNF. Both SelMFDF and WMFLDA lose to AHNF, since they also suffer from the information loss when handcrafting attribute data matrices into homologous networks, and do not differentiate the observed associations from other ones. Both SIMCLDA and RIMC complete lncRNA-disease associations via matrix completion. Although they can accommodate attribute data matrices by concatenating and projecting various heterogeneous attribute matrices into a homologous one, they are still outperformed by matrix tri-factorization based solutions (except S-NMTF). That is because they implicitly deem the collected associations as 'complete'; they can only utilize the biological data directly related with lncRNAs or diseases; and do not fuse multi-type relational biological data in a coordinated fashion.

In addition, we introduce two variants, AHNF(nH) and AHNF(Homo), to study the effect of considering the insufficiently observed associations and directly fusing multiple attribute data, respectively. AHNF(nH) excludes the Hadamard product $\mathbf{H}$ in the approximate matrix factorization, so it still assumes the observed associations are complete'. AHNF(Homo) handcrafts multiple heterogeneous attribute matrices into homologous networks derived from the b nearest neighbors network based on the Euclidean distance. Fig. 3 gives the results of AHNF and its variants. AHNF obtains higher AUROC and AUPRC than AHNF(nH), since it accounts for the fact that the observed associations are insufficient, and models them. Although both AHNF and AHNF(Homo) account for the incomplete associations and use the same data sources, AHNF outperforms AHNF(Homo). That is because AHNF(Homo) suffers from information loss when handcrafting heterogeneous attribute data into homologous similarity matrices.

We further study other distance metrics (i.e., 'Euclidean', 'Standard Euclidean', 'Cosine', "', and 'Spearman') with different choices of b, and report the results of AHNF(Homo) in five-fold cross validation in Fig. 4. AHNF does not require hand-crafting different attribute data matrices into homologous networks, so its results are independent from the distance metrics and b. We can clearly see that no matter what distance metric is used, both AUROC and AUPRC of AHNF(Homo) are decreased as the increase of b and the results are stable after $b > 5$. AHNF(Homo) obtains the best performance using 'Euclidean' distance metric with $b = 1$. Even though, AHNF(Homo) is still always outperformed by AHNF, which does not handcraft heterogeneous attribute data matrices into homologous networks. This study confirms the information loss when handcrafting heterogeneous data into homologous networks for fusion, and also supports our motivation for attributed heterogeneous network fusion.

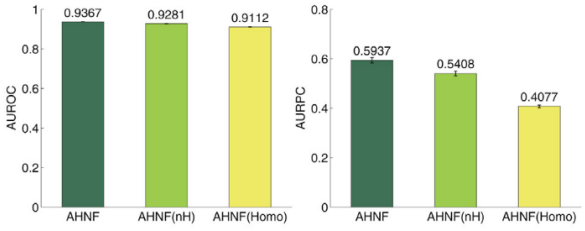
3.3. Predicting masked lncRNA-disease associations

Disease Ontology (DO) [37] has a hierarchical structure among diseases via a direct acyclic graph. If a given lncRNA is associated with a DO term, then the same lncRNA is also associated with ancestor terms of



(a) ROC curves and AUROCs (b) PR curves and AUPRCs

Fig. 2. Prediction results of AHNF, SelMFDF, WMFLDA, MFLDA, DFMF, SIMCLDA, RIMC, and S-NMTF.



(a) AUROCs (b) AUPRCs

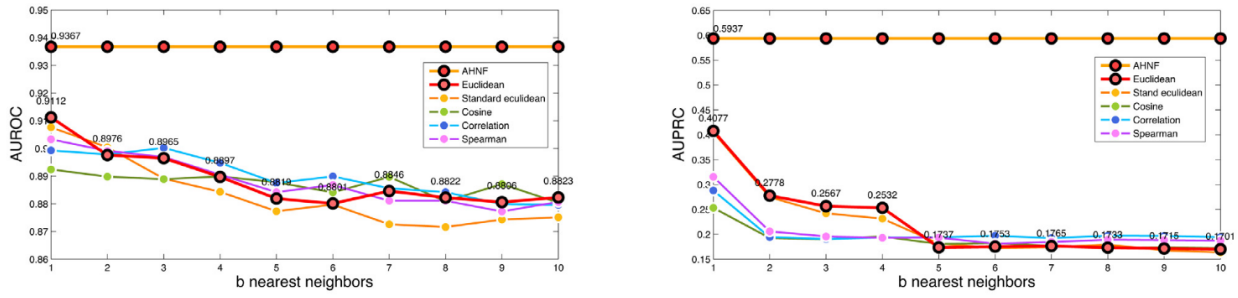
Fig. 3. Prediction results of variants of AHNF. AHNF(nH) disregards the indicator matrix $\mathbf{H}_{ij}$, and treats equally the observed relations and others; ANHF(Homo) projects heterogeneous attribute data matrices onto homologous networks.

the DO term, and may also be associated with its descendant terms in the hierarchy. We may further find that the discovered diseases associated with a certain lncRNA become more and more specific as the biomedical technique develops [38,39]. Given this, we perform another type of simulated experiment to further study the capability of AHNF of predicting specific lncRNA-disease associations. We assume that the disease terms related to a specific lncRNA are complete, and these terms form a direct acyclic graph. We observe that, once a particular leaf disease node is masked in the direct acyclic graph and its direct parent disease node has no descendants associated with the specific lncRNA, then the direct parent disease nodes can also be masked. Following this principle, we randomly mask q particular diseases corresponding to leaf nodes in the direct acyclic graph. We consider the masked lncRNA-disease associations for prediction. We repeat the mask operation for each lncRNA under each $q \in \{1, 3, 5\}$ in 5 independent rounds, and show the average results with standard deviations in Table 3. If q is larger than the number of disease terms associated with an lncRNA, we ensure that the lncRNA is associated with at least one disease. The average numbers of masked specific diseases and associations under different input values of q are also counted. • indicates the best results in the same setting, with significance checked by a pairwise t-test at 95% confidence level.

From Table 3, we can clearly see that AHNF outperforms the competing methods across different input values of q. All the comparing methods show a decreasing trend as q increase; this is because, as the number of masked associations increases, the prediction task becomes more difficult. This pattern suggests that the number of known associations can affect the prediction performance, and that the known associations should be specifically used. In other words, the insufficient associations between lncRNAs and diseases impact the prediction.

3.4. Predicting De Novo lncRNAs-disease associations

We perform another type of experiments to investigate the potential of AHNF in predicting *de novo* associations between an lncRNA (or disease), whose associated diseases (or lncRNAs) are *completely* unknown. Particularly, we randomly divide the lncRNAs into five equal folds and then remove all known lncRNA-disease associations for the queried lncRNAs in each fold, rather than randomly masking 20% known lncRNA-disease associations. Similarly, we perform disease oriented five-fold cross validation by removing all the associated lncRNAs for the queried diseases in each fold. Table 4 reports the AUROC and



(a) AUROCs (b) AUPRCs

Fig. 4. AUROC and AUPRC of AHNF(Homo) with different distance metrics and b nearest neighbors network.

Table 3
AUROC and AUPRC of the competitive methods for predicting (q) masked lncRNA-disease associations of each lncRNA.

q	1		3		5	
Masked diseases/associations	24.6/158		59.4/467		80.8/756	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
AHNF	0.9650 ± 0.0015•	0.7859 ± 0.0004•	0.9537 ± 0.0000•	0.7856 ± 0.0037•	0.9444 ± 0.0007•	0.7566 ± 0.0002•
SelMFDF	0.9591 ± 0.0010	0.7398 ± 0.0007	0.9520 ± 0.0003	0.7288 ± 0.0012	0.9308 ± 0.0014	0.7070 ± 0.0002
WMFLDA	0.9581 ± 0.0002	0.7328 ± 0.0000	0.9453 ± 0.0000	0.7009 ± 0.0003	0.9124 ± 0.0002	0.6552 ± 0.0003
MFLDA	0.9474 ± 0.0006	0.7237 ± 0.0009	0.9368 ± 0.0000	0.7085 ± 0.0002	0.9125 ± 0.0001	0.6477 ± 0.0010
DFMF	0.9457 ± 0.0012	0.6968 ± 0.0021	0.9359 ± 0.0000	0.6577 ± 0.0002	0.9093 ± 0.0004	0.5671 ± 0.0009
SIMCLDA	0.8748 ± 0.0000	0.1026 ± 0.0000	0.8700 ± 0.0000	0.0892 ± 0.0001	0.8608 ± 0.0001	0.0772 ± 0.0000
RIMC	0.8225 ± 0.0000	0.3966 ± 0.0001	0.8210 ± 0.0023	0.3850 ± 0.0070	0.8195 ± 0.0000	0.3745 ± 0.0000
S-NMTF	0.8036 ± 0.0000	0.1287 ± 0.0010	0.8017 ± 0.0000	0.1274 ± 0.0002	0.8014 ± 0.0002	0.1257 ± 0.0000

Table 4
AUROC and AUPRC of the competitive methods for predicting de novo lncRNA-disease associations.

	LncRNA		Disease	
	AUROC	AUPRC	AUROC	AUPRC
AHNF	0.766 ± 0.016•	0.212 ± 0.001•	0.848 ± 0.001•	0.478 ± 0.006•
SelMFDF	0.642 ± 0.001	0.177 ± 0.001	0.815 ± 0.002	0.420 ± 0.013
WMFLDA	0.612 ± 0.003	0.156 ± 0.000	0.776 ± 0.001	0.400 ± 0.002
MFLDA	0.584 ± 0.003	0.134 ± 0.004	0.755 ± 0.006	0.387 ± 0.006
DFMF	0.555 ± 0.002	0.155 ± 0.002	0.707 ± 0.002	0.298 ± 0.001
SIMCLDA	0.604 ± 0.004	0.046 ± 0.000	0.731 ± 0.006	0.034 ± 0.001
RIMC	0.647 ± 0.003	0.078 ± 0.006	0.783 ± 0.001	0.204 ± 0.004
S-NMTF	0.528 ± 0.007	0.072 ± 0.003	0.774 ± 0.000	0.083 ± 0.000

AUPRC values for AHNF and the competing methods. AHNF achieves the highest AUROC (0.7661 and 0.8487), and the highest AUPRC (0.2119 and of 0.4785) in both the lncRNA-oriented and disease-oriented de novo association prediction.

We observe that the AUROC and AUPRC values are lower than those of the previous five fold cross validation. The reason is that it is more difficult to infer the association between a particular disease and lncRNA, when the related lncRNAs are completely unknown. RIMC performs better than MFLDA and DFMF; this is because RIMC completes the potential lncRNA-disease associations using attribute data matrices, which provide additional information for identifying potential associations between lncRNAs and diseases. Another possible cause is that RIMC induces a sparse regularization to reduce the false positive rate. These results suggest that leveraging multi-type relational data and diverse attribute data can boost the identification of novel lncRNA-disease associations. AHNF can more effectively integrate heterogeneous data sources, and also includes a sparsity regularization; as such, it manifests superior results than the competitive solutions.

3.5. Case studies

To further study whether the lncRNA-disease associations predicted by AHNF can be confirmed by biological experiments, we apply AHNF to prioritize all the candidate lncRNAs for *breast* cancer, *stomach* cancer, *prostate* cancer, and *pancreatic* cancer, respectively. Here, besides the associations between lncRNAs and the studied cancer, all the other known associations are used as training samples. We then select the top 20 plausible associations as the predicted lncRNA-disease associations for each cancer. After this, we check the predicted associations by referring to available associations in LncRNADisease 2.0 [40] and Lnc2Cancer v2.0 [41]. For the predicted associations that cannot be found in the above two databases, we further do manual text retrieval on PubMed and list the supportive PubMed IDs (not included in the aforementioned two databases) with **boldface**.

Table 5 lists the currently supported and un-supported associations for breast cancer. We find that 19 prioritized lncRNAs have supportive evidence for breast cancer. For example, the expression of lncRNA 'PVT1' was significantly up-regulated in breast cancer tissues, compared with adjacent normal tissues (ANTs), so lncRNA 'PVT1' may be a prognostic predictive biomarker for breast cancer [42]. Furthermore, lncRNA 'HCP5' was observed to be significantly up-regulated in breast cancer [43], which provides first time evidence that lncRNA 'HCP5' is associated with breast cancer prognosis. Therefore, lncRNA 'HCP5' probably contributes to the comprehensive treatment and diagnose of breast cancer after experimental verification. These case studies further confirm the capability of AHNF in identifying novel lncRNA-disease associations with confidence. The un-validated association 'MIR17HG' may be further verified when more experimental evidences become available.

Table 5
AHNF predicted lncRNAs associated with *breast cancer* (top 20 in ranking list), and the corresponding evidence.

Rank	lncRNA	Confirmed	Evidence(PubMed ID)
1	PVT1	√	28882595;26889781;25122612
2	BCYRN1	√	27277684;18006640;15240511
3	CDKN2B-AS1	√	28580310;27102007;26835415
4	H19	√	29693231;28544374;28102845
5	HOTAIR	√	29630518;29423075;29222472
6	NEAT1	√	28338194;28034643;27556296
7	MEG3	√	28635399;28051255;27166155
8	GAS5	√	29180320;27034004;26862727
9	MALAT1	√	29574704;29416769;29386907
10	UCA1	√	29408336;27977766;27831634
11	KCNQ1OT1	√	26323944;21304052
12	HOTTIP	√	29415429;28036281
13	MIR17HG	?	Not yet found
14	BANCR	√	29805676;29565494
15	HCP5	√	**31215169;31028999;30618123**
16	XIST	√	29550489;27362246;27248326
17	DANCR	√	27716745
18	TUG1	√	28950664;28053623;27848085
19	HULC	√	27986124
20	LSINCT5	√	29785740;21532345;20214974

Table 6 lists 20 prioritized lncRNAs with supportive evidence for stomach cancer. For example, lncRNA 'MEG3' expression was signifi-

Table 6
AHNF predicted lncRNAs associated with *stomach cancer* (top 20 in ranking list), and the corresponding evidence.

Rank	lncRNA	Confirmed	Evidence(PubMed ID)
1	MEG3	√	26718650;26253106;26233544
2	HOTAIR	√	29683069;29417297;29080815
3	CDKN2B-AS1	√	27121324;27027260;25636450
4	H19	√	29207111;28105222;27835600
5	PVT1	√	28258379;28122299;27756785
6	GAS5	√	29098549;27827524;27466992
7	UCA1	√	29723509;29516678;29212166
8	MALAT1	√	29162158;28396617;28276823
9	NEAT1	√	29544562;29363783;28401449
10	KCNQ1OT1	√	25765901
11	HULC	√	28356873;27781386;27246953
12	TUG1	√	29182008;27983921;27261864
13	SPRY4-IT1	√	26238992;25835973
14	BANCR	√	29483646;28042329;26054683
15	HOTTIP	√	29486794;28693275;27546609
16	CCAT2	√	29435046;28248065;27904778
17	WT1-AS	√	26449525
18	CCAT1	√	28535628;28239816;27134049
19	LSINCT5	√	25694351;25526476
20	ZFAS1	√	29424266;28285404;27654478

Table 7
AHNF predicted lncRNAs associated with *prostate cancer* (top 20 in ranking list), and the corresponding evidence.

Rank	lncRNA	Confirmed	Evidence(PubMed ID)
1	PVT1	√	29452232;29050519;27794184
2	H19	√	24988946;24063685
3	MEG3	√	27507663;26610246;23728290
4	NEAT1	√	25415230;23728290
5	HOTAIR	√	29436234;28259691;27922078
6	CDKN2B-AS1	√	28621612;27507663;22664915
7	GAS5	√	29416676;27743383;27507663
8	MALAT1	√	29633510;27922078;27600237
9	UCA1	√	28337266;27902466;27686228
10	TUG1	√	26975529
11	MIR17HG	√	27556357
12	KCNQ1OT1	√	23728290
13	HOTTIP	√	27064878;23728290
14	BANCR	?	Not yet found
15	HULC	√	23728290
16	DANCR	√	27191265;23728290
17	LSINCT5	?	Not yet found
18	CCAT2	√	27558961
19	CCAT1	√	29863242;29765457;29694502
20	TP53COR1	√	25999983

Table 8
AHNF predicted lncRNAs associated with *pancreatic cancer* (top 20 in ranking list), and the corresponding evidence.

Rank	lncRNA	Confirmed	Evidence(PubMed ID)
1	MALAT1	√	29215734;28034748;27777857
2	MEG3	√	29328401;26850851
3	HOTAIR	√	28476883;27895308;27028998
4	H19	√	24920070
5	GAS5	√	29112934;24026436
6	CDKN2B-AS1	√	28344092
7	PVT1	√	28657147;28355965;27028998
8	UCA1	√	29510195;28315290;27562722
9	NEAT1	√	27888106
10	TUG1	√	28617552
11	HULC	√	27781386;25412939
12	HOTTIP	√	28947139;27546609;26447755
13	BANCR	?	Not yet found
14	SPRY4-IT1	√	29489909
15	CCAT2	?	Not yet found
16	HCP5	√	**31061236**
17	XIST	√	29393501;28295543
18	WT1-AS	√	19196508
19	AFAP1-AS1	√	25910082
20	PCGEM1	√	27682980;27507663;25744782

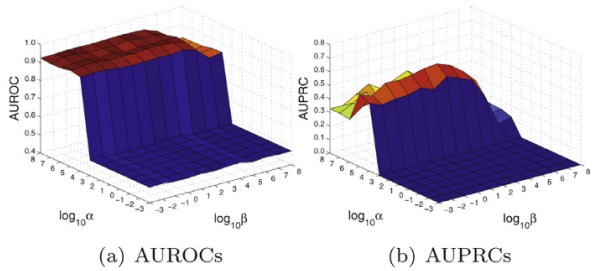


Fig. 5. AUROC and AUPRC of AHNF under different input values of α and β.

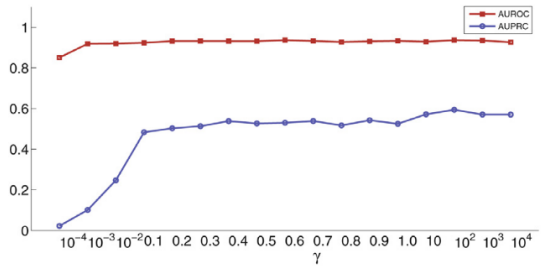


Fig. 6. AUROC and AUPRC of AHNF under different input values of the parameter γ.

cantly lower in 20 early stage stomach cancer patient tumor tissues than adjacent non-tumor tissues, and it may be a promising biomarker for the early detection and early screening of high risk populations [44]. Furthermore, plasma level of lncRNA 'HOTAIR' was significantly higher in stomach cancer patients compared with healthy controls and it also can be a potential non-invasive biomarker for diagnosis of stomach cancer [45].

For prostate cancer, as shown in Table 7, 18 prioritized lncRNAs have found supportive evidences. To name a few, lncRNA 'PVT1' has been reported that it can promote prostate cancer invasion and metastasis by modulating the mechanism of endothelialmesenchymal transition (EMT) and it further promotes EMT by up-regulation of 'Twist1', a transcription factor associated with EMT [46]. Since the increased expression of lncRNA 'CCAT1' was significantly related to lymph node metastasis in prostate cancer; low-expression of 'CCAT1' could suppress cell proliferation; knockdown of 'CCAT1' inhibited the migration of prostate cancer PC-3 cells; down-regulation of 'CCAT1' attenuated the invasion of PC-3 cells, lncRNA 'CCAT1' can promote the growth and the metastasis of prostate cancer and it might be a potential target for the diagnosis and treatment of prostate cancer [47]. As for the un-validated associations 'BANCR' and 'LSINCT5', they may be further verified once more experimental evidences become available.

As reported in Table 8, we also find 18 prioritized lncRNAs have supportive evidences for pancreatic cancer. For instance, lncRNA 'MALAT1' presented an extremely high expression level in pancreatic cancer tissues and cells [48]. Loss of MEG3 expression was observed in both the cancerous tissues and cancer cell lines compared to the expression in adjacent non-cancerous tissues and a human pancreatic normal epithelial cell line, so lncRNA 'MEG3' functions as a tumour suppressor in human pancreatic cancer [49]. In addition, a novel sub-network, mRNA 'MMP9'-miRNA 'miR-29b-3p'-lncRNA 'HCP5', established by [50] was first demonstrated that it was linked to prognosis of pancreatic cancer according to the results from expression, survival and correlation analysis. The lncRNA 'BANCR' and 'CCAT2' might be further verified once more experimental evidence becomes available. These case studies fur-

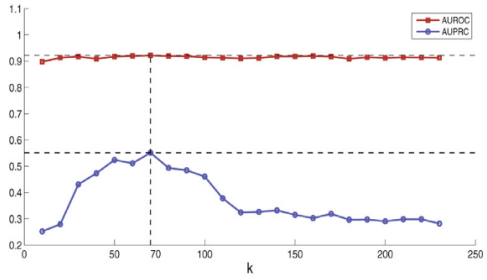


Fig. 7. AUROC and AUPRC of AHNF under different input values of low-rank size k.

ther confirm the capability of AHNF in identifying novel lncRNA-disease associations with confidence.

3.6. Parameter analysis

According to the explicit solution of $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$ in Appendix A, we can clearly see that once the value of α is specified, the weight $\boldsymbol{\omega}^r_{ij}$ assigned to $\mathbf{R}_{ij} \in \mathcal{R}$ can be computed based on the reconstruction loss of that matrix. In addition, we can easily see that once the value of β is specified, the weight $\boldsymbol{\omega}^h_{it}$ assigned to attribute data $\mathbf{X}_{it}$ can be determined based on the reconstruction loss for the attribute data matrix. Given this, both α and β play important roles for the performance of AHNF. To find a feasible value of α and β, following the experimental settings in Section 3.2, we conduct five-fold cross validation to predict lncRNA-disease associations by varying α and β in $\{10^{-3}, \ldots, 10^{7}, 10^{8}\}$, and report the AUROC and AUPRC under each combination of α and β in Fig. 5.

From Fig. 5, we observe that when $\alpha = 10^5$ and $\beta = 10^3$, AHNF achieves the highest AUPRC. The input value of α significantly affects the performance; the AUROC value increases as α rises, and reaches a plateau when $\alpha \geq 10^5$; the AUPRC reaches the highest value when $\alpha = 10^5$, and then it slightly decreases and stays stable. This is because a too small α value assigns little emphasis to the relational adjacency matrices, and a too large α will treat all the relational adjacency matrices equally, and the target relational matrix is underrated as a result. The input value of β also affects the performance; the AUROC is stable when $\beta \leq 10^5$, and then slightly decreases; the AUPRC value becomes larger as β increases, it reaches the highest value when $\beta = 10^3$, and then it significantly decreases. This observation shows that both the input values of α and β have an impact on the performance of AHNF.

The regularization parameter γ is used to balance the reconstruction constraint and approximation error. Intuitively, if the magnitude of γ is too large, the reconstructed association matrix will be too sparse to complete the missing associations. On the other hand, a too small γ can not enforce a sufficient regularization of $\left\|\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right\|_F^2$ and result in more false positive associations in the reconstructed association matrix. Following the experimental settings in Section 3.2, we conduct five-fold cross validation to predict lncRNA-disease associations by varying γ in $\{10^{-4}, 10^{-3}, \ldots, 10^{4}\}$, and report the average AUROC and AUPRC in Fig. 6. We can find that the AUPRC value gets larger as γ increases. When $\gamma \leq 10^{-2}$, the AUPRC manifests a significant decrease. AUROC holds a relatively stable value as $\gamma \geq 0.2$. As a result, we adopt $\gamma = 10^2$ for AHNF.

The low-rank size k_i is an important parameter for low-rank matrix approximation based solutions. To study the sensitivity of k_i, we fix all $k_i = k$ across these five types of objects for simplicity, and then increase k from 10 to 230. Fig. 7 reports the AUROC and AUPRC under different input values of k in predicting lncRNA-disease associations in five-fold cross validation. We observe that AUROC values are much stable than AUPRC values, this is mainly because Precision-Recall curves are much sensitive than ROC curves for the class imbalance problem [51,52]. We can see that AUROC value has a slight fluctuation as k increase. AUPRC increases as the increase of k when $k \leq 70$, and then it has a consistent decrease after $k > 70$. This is because the dimensionality of some attribute data matrices is small, and a too large rank size imports redundancy (and even noise) into the low-rank representation matrices $\mathbf{G}_i$ and consequently increases false positive associations. Given these observations, we adopt $k = 70$ for experiments.

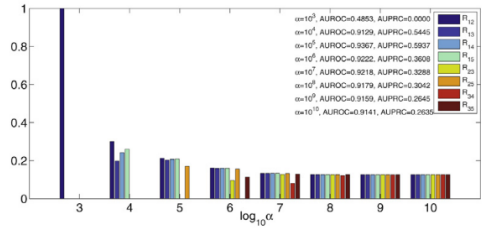


Fig. 8. Weights ($\boldsymbol{\omega}^r_{ij}$) assigned to 8 relational adjacency matrices $\mathbf{R}_{ij}$ when $\alpha \in \{10^3, 10^4, \cdots, 10^{10}\}$.

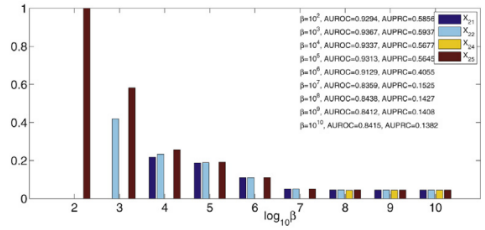


Fig. 9. Weights ($\boldsymbol{\omega}^h_{2t}$) assigned to 5 heterogeneous feature matrices $\mathbf{X}_{2t}$ of miRNAs when $\beta \in \{10^3, 10^4, \cdots, 10^{10}\}$.

To show the capability of AHNF in selectively combining multiple inter-relational matrices and attribute matrices, we further report the weights ($\boldsymbol{\omega}^r_{ij}$) assigned to 8 relational adjacency matrices under different input values of α with $\beta = 10^3$ in Fig. 8, and the weights ($\boldsymbol{\omega}^h_{2t}$) assigned to 5 heterogeneous attribute matrices of miRNAs under different input values of β with $\alpha = 10^5$ in Fig. 9.

We can clearly observe that when $\alpha = 10^3$, only the relational adjacency matrix $\mathbf{R}_{12}$ is selected. $\mathbf{R}_{12}$ has the fewest known associations (1002) among all the relational data matrices in Table 1. When $\beta \geq 10^7$, all the 8 relational data matrices are selected and assigned nearly equal weights. This is because a (too) small α value does not have a sufficient regularization effect on the weights assigned to different relational adjacency matrices. On the other hand, a (too) large α value results in a strong regularization effect, and forces similar weight assignments to all matrices. When $\alpha = 10^4$, 10^5 or 10^6, some relational matrices ($\mathbf{R}_{35}$) are excluded for fusion, and AHNF has the highest AUROC and AUPRC when $\alpha = 10^5$. The exclusion of some inter-relational matrices ($\mathbf{R}_{23}$ and $\mathbf{R}_{35}$) is possible because these matrices may contain too much noisy relational associations than the selected sources, and the selected inter-relational matrices ($\mathbf{R}_{25}$) have more reliable inter-associations for an accurate lncRNA-disease associations prediction.

We further observe that when $\beta = 10^2$, only the attribute matrix $\mathbf{X}_{25}$ for miRNAs is selected. When $\beta = 10^3$, $\mathbf{X}_{22}$ and $\mathbf{X}_{25}$ are selected for fusion, and AHNF has the highest AUROC and AUPRC. Similar to the parameter α, a (too) small or a (too) large β value can not provide the different weights to distinguish the multiple attribute matrices. When $\beta \geq 10^4$, nearly all the selected attribute matrices are assigned nearly equal weights and the performance of AHNF decreases as the β value increases. The mainly reason may be the more noisy attribute matrices have been selected as β grows. These experiments confirm that AHNF indeed can selectively integrate different inter-relational matrices and heterogeneous attribute data matrices, which contribute to an improved performance.

4. Conclusion

We introduced an attributed heterogeneous network fusion framework based on collaborative matrix tri-factorization. Unlike existing matrix factorization based data fusion approaches, AHNF can not only conquer the negative impact of insufficient relations between nodes, but also avoid the information loss when transforming attribute data of diverse network nodes into homologous networks for fusion. Extensive experimental results show that AHNF achieves a superior performance compared to the state-of-the-art solutions in predicting lncRNA-disease associations. AHNF is a general data fusion framework to comprehensively integrate attributed multi-type relational data for diverse tasks. There are some avenues for future work: (i) our collaborative matrix tri-factorization framework is not limited to the data types we used in the experiments, it can be applied for various link prediction problems, such as drug repurposing, user recommendation and so on. (ii) The framework can be extended by merging with deep learning techniques to improve the efficiency and effectiveness.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Guoxian Yu: Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft. **Yuehui Wang:** Data curation, Validation, Formal analysis. **Jun Wang:** Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Carlotta Domeniconi:** Formal analysis, Writing - review & editing. **Maozu Guo:** Formal analysis, Writing - review & editing. **Xiangliang Zhang:** Writing - review & editing.

Acknowledgment

This work is supported by Natural Science Foundation of China (61872300 and 61873214).

Appendix A. Objective Function Optimization for AHNF

This appendix elaborates on how to iteratively optimize **G, S,** $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$ in the objective function (Eq. (4) of the main text) of AHNF. Before elaborating on the updating rule, we introduce the Lagrangian multipliers $\{\boldsymbol{\lambda}_i\}_{i=1}^m$ for $\mathbf{G}_i \geq 0$, and reformulate the objective function of AHNF as follows:

$$
\begin{aligned}
&\min_{\mathbf{G}\geq 0} \mathcal{J}\left(\mathbf{G},\mathbf{S},\mathbf{U},\boldsymbol{\omega}^r,\boldsymbol{\omega}^h\right) = \\
&\sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r tr\Big[\mathbf{H}_{ij} \odot \left(\mathbf{R}_{ij}\mathbf{R}_{ij}^T - 2\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T\mathbf{G}_i^T + \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\mathbf{G}_i^T\right)\Big] \\
&+ \sum_{i=1}^{m}\sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h tr\left(\mathbf{X}_{it}\mathbf{X}_{it}^T - 2\mathbf{X}_{it}\mathbf{U}_{it}\mathbf{G}_i^T + \mathbf{G}_i\mathbf{U}_{it}^T\mathbf{U}_{it}\mathbf{G}_i^T\right) \\
&+ \alpha\|vec(\boldsymbol{\omega}^r)\|_F^2 + \beta\left\|vec\left(\boldsymbol{\omega}^h\right)\right\|_F^2 \\
&+ \gamma\sum_{i,j=1}^{m} \gamma tr\left(\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\mathbf{G}_i^T\right) \\
&+ \sum_{i=1}^{m} tr(\lambda_i\mathbf{G}_i^T) \\
&s.t. \qquad \boldsymbol{\omega}^r \geq 0, \boldsymbol{\omega}^h \geq 0, \sum_{i,j=1}^{m}\boldsymbol{\omega}_{ij}^r = 1, \sum_{i=1}^{m}\sum_{t=1}^{\max_i t_i}\boldsymbol{\omega}_{it}^h = 1
\end{aligned} \tag{A.1}
$$

Our optimization of **G, S** and **U** follows the ideas of multiplicative factor update rules used in [11,24,53,54].

Optimizing S: Suppose **G, U,** $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$ are known, to obtain the optimal $\mathbf{S}_{ij}$ (if $\mathbf{R}_{ij} \in \mathcal{R}$), we can take the partial derivative of $\mathcal{J}(\mathbf{G},\mathbf{S},\mathbf{U},\boldsymbol{\omega}^r,\boldsymbol{\omega}^h,\boldsymbol{\lambda})$ with respect to $\mathbf{S}_{ij}$ as follows:

$$
\begin{aligned}
\frac{\partial\mathcal{J}}{\partial\mathbf{S}_{ij}} &= \boldsymbol{\omega}_{ij}^r\mathbf{H}_{ij} \odot \left(-2\mathbf{G}_i^T\mathbf{R}_{ij}\mathbf{G}_j + 2\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\right) \\
&\quad + 2\gamma\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j \\
&= -2\boldsymbol{\omega}_{ij}^r\mathbf{G}_i^T\left(\mathbf{H}_{ij}\odot\mathbf{R}_{ij}\right)\mathbf{G}_j + 2\boldsymbol{\omega}_{ij}^r\mathbf{G}_i^T\left(\mathbf{H}_{ij}\odot\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right)\mathbf{G}_j \\
&\quad + 2\gamma\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j
\end{aligned} \tag{A.2}
$$

By letting $\frac{\partial\mathcal{J}}{\partial\mathbf{S}_{ij}} = 0$ for $\forall i,j \in \{1,2,\ldots,m\}$, we can obtain:

$$
\mathbf{S}_{ij} = \mathbf{S}_{ij} \odot \frac{\boldsymbol{\omega}_{ij}^r\mathbf{G}_i^T\left(\mathbf{H}_{ij}\odot\mathbf{R}_{ij}\right)\mathbf{G}_j}{\boldsymbol{\omega}_{ij}^r\mathbf{G}_i^T\left(\mathbf{H}_{ij}\odot\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right)\mathbf{G}_j + \gamma\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j} \tag{A.3}
$$

Optimizing U: Similarly, suppose **G, S,** $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$ are known, the partial derivative of $\mathcal{J}(\mathbf{G},\mathbf{S},\mathbf{U},\boldsymbol{\omega}^r,\boldsymbol{\omega}^h,\boldsymbol{\lambda})$ with respect to $\mathbf{U}_{it}$ is:

$$
\frac{\partial\mathcal{J}}{\mathbf{U}_{it}} = \boldsymbol{\omega}_{it}^h\left(-2\mathbf{X}_{it}^T\mathbf{G}_i + 2\mathbf{U}_{it}\mathbf{G}_i^T\mathbf{G}_i\right) \tag{A.4}
$$

By letting $\frac{\partial\mathcal{J}}{\partial\mathbf{U}_{it}} = 0$ for $\forall i,j \in \{1,2,\ldots,m\}, t < \max_i t_i$, we can obtain:

$$
\mathbf{U}_{it} = \mathbf{U}_{it} \odot \frac{\mathbf{X}_{it}^T\mathbf{G}_i}{\mathbf{U}_{it}\mathbf{G}_i^T\mathbf{G}_i} \tag{A.5}
$$

Optimizing G: Suppose **S, U,** $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$ are known, the partial derivative of $\mathcal{J}(\mathbf{G},\mathbf{S},\mathbf{U},\boldsymbol{\omega}^r,\boldsymbol{\omega}^h,\boldsymbol{\lambda})$ with respect to $\mathbf{G}_i$ is:

$$
\begin{aligned}
\frac{\partial\mathcal{J}}{\mathbf{G}_i} &= \sum_{j:\mathbf{R}_{ij}\in\mathcal{R}} \boldsymbol{\omega}_{ij}^r tr\Big[\mathbf{H}_{ij}\odot\left(-2\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T + 2\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)\Big] \\
&\quad + \sum_{j:\mathbf{R}_{ji}\in\mathcal{R}} \boldsymbol{\omega}_{ji}^r tr\Big[\mathbf{H}_{ji}^T\odot\left(-2\mathbf{R}_{ji}^T\mathbf{G}_j\mathbf{S}_{ji} + 2\mathbf{G}_i\mathbf{S}_{ji}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ji}\right)\Big] \\
&\quad + \sum_{j:\mathbf{R}_{ij}\in\mathcal{R}} \gamma 2\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T + \sum_{j:\mathbf{R}_{ji}\in\mathcal{R}} \gamma 2\mathbf{G}_i\mathbf{S}_{ji}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ji} \\
&\quad + \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h\left(-2\mathbf{X}_{it}\mathbf{U}_{it} + 2\mathbf{G}_i\mathbf{U}_{it}^T\mathbf{U}_{it}\right) - \boldsymbol{\lambda}_i
\end{aligned} \tag{A.6}
$$

Multipliers λ_i can be obtained from Eq. (A.6) by letting $\frac{\partial\mathcal{J}}{\mathbf{G}_i} = 0$. The KKT (Karush-Kuhn-Tucker) complementary condition [55] for nonnegativity of $\mathbf{G}_i$ is:

$$
\begin{aligned}
0 &= \boldsymbol{\lambda}_i \circ \mathbf{G}_i \\
&= \Big[\sum_{j:\mathbf{R}_{ij}\in\mathcal{R}} \boldsymbol{\omega}_{ij}^r tr\Big[\mathbf{H}_{ij}\odot\left(-2\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T + 2\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)\Big] \\
&\quad + \sum_{j:\mathbf{R}_{ji}\in\mathcal{R}} \boldsymbol{\omega}_{ji}^r tr\Big[\mathbf{H}_{ji}^T\odot\left(-2\mathbf{R}_{ji}^T\mathbf{G}_j\mathbf{S}_{ji} + 2\mathbf{G}_i\mathbf{S}_{ji}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ji}\right)\Big] \\
&\quad + \sum_{j:\mathbf{R}_{ij}\in\mathcal{R}} \gamma 2\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T + \sum_{j:\mathbf{R}_{ji}\in\mathcal{R}} \gamma 2\mathbf{G}_i\mathbf{S}_{ji}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ji} \\
&\quad + \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h\left(-2\mathbf{X}_{it}\mathbf{U}_{it} + 2\mathbf{G}_i\mathbf{U}_{it}^T\mathbf{U}_{it}\right)\Big] \odot \mathbf{G}_i
\end{aligned} \tag{A.7}
$$

where $\odot$ denotes the Hadamard product. Eq. (A.7) is a fixed point equation and the solution must satisfy it at convergence. We can let

$$
\begin{aligned}
\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T &= \left(\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T\right)^+ - \left(\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T\right)^- \\
\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T &= \left(\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)^+ - \left(\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)^- \\
\mathbf{R}_{ji}^T\mathbf{G}_j\mathbf{S}_{ji} &= \left(\mathbf{R}_{ji}^T\mathbf{G}_j\mathbf{S}_{ji}\right)^+ - \left(\mathbf{R}_{ji}^T\mathbf{G}_j\mathbf{S}_{ji}\right)^- \\
\mathbf{G}_i\mathbf{S}_{ji}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ji} &= \left(\mathbf{G}_i\mathbf{S}_{ji}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ji}\right)^+ - \left(\mathbf{G}_i\mathbf{S}_{ji}^T\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ji}\right)^- \\
\mathbf{X}_{it}\mathbf{U}_{it} &= \left(\mathbf{X}_{it}\mathbf{U}_{it}\right)^+ - \left(\mathbf{X}_{it}\mathbf{U}_{it}\right)^-
\end{aligned}
$$

$$\mathbf{G}_i\mathbf{U}_{it}^T\mathbf{U}_{it} = \left(\mathbf{G}_i\mathbf{U}_{it}^T\mathbf{U}_{it}\right)^+ - \left(\mathbf{G}_i\mathbf{U}_{it}^T\mathbf{U}_{it}\right)^- \tag{A.8}$$

For $\mathbf{R}_{ij} \in \mathcal{R}$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} +&= \boldsymbol{\omega}_{ij}^r\mathbf{H}_{ij} \odot \left[\left(\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T\right)^+ + \left(\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)^-\right] \\
&\quad + \gamma\mathbf{G}_i\left(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)^- \\
\mathbf{G}_i^{(d)} +&= \boldsymbol{\omega}_{ij}^r\mathbf{H}_{ij} \odot \left[\left(\mathbf{R}_{ij}\mathbf{G}_j\mathbf{S}_{ij}^T\right)^- + \left(\mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)^+\right] \\
&\quad + \gamma\mathbf{G}_i\left(\mathbf{S}_{ij}\mathbf{G}_j^T\mathbf{G}_j\mathbf{S}_{ij}^T\right)^+ \\
\mathbf{G}_j^{(e)} +&= \boldsymbol{\omega}_{ij}^r\mathbf{H}_{ij}^T \odot \left[\left(\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij}\right)^+ + \left(\mathbf{G}_j\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\right)^-\right] \\
&\quad + \gamma\mathbf{G}_j\left(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\right)^- \\
\mathbf{G}_j^{(d)} +&= \boldsymbol{\omega}_{ij}^r\mathbf{H}_{ij}^T \odot \left[\left(\mathbf{R}_{ij}^T\mathbf{G}_i\mathbf{S}_{ij}\right)^- + \left(\mathbf{G}_j\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\right)^+\right] \\
&\quad + \gamma\mathbf{G}_j\left(\mathbf{S}_{ij}^T\mathbf{G}_i^T\mathbf{G}_i\mathbf{S}_{ij}\right)^+
\end{aligned} \tag{A.9}$$

For $t = 1, 2, \ldots, \max_i t_i$:

$$\begin{aligned}
\mathbf{G}_i^{(e)} +&= \boldsymbol{\omega}_{it}^h\left(\mathbf{X}_{it}\mathbf{U}_{it}\right)^+ + \boldsymbol{\omega}_{it}^h\mathbf{G}_i\left(\mathbf{U}_{it}^T\mathbf{U}_{it}\right)^- \quad \text{for} \quad i = 1, 2, \ldots, m \\
\mathbf{G}_i^{(d)} +&= \boldsymbol{\omega}_{it}^h\left(\mathbf{X}_{it}\mathbf{U}_{it}\right)^- + \boldsymbol{\omega}_{it}^h\mathbf{G}_i\left(\mathbf{U}_{it}^T\mathbf{U}_{it}\right)^+ \quad \text{for} \quad i = 1, 2, \ldots, m
\end{aligned} \tag{A.10}$$

where the matrices with positive and negative symbols are defined as $\mathbf{Z}^+ = \frac{|\mathbf{Z}|+\mathbf{Z}}{2}$ and $\mathbf{Z}^- = \frac{|\mathbf{Z}|-\mathbf{Z}}{2}$, respectively. We then update $\mathbf{G}$ as follows:

$$\mathbf{G} \leftarrow \mathbf{G} \odot diag\left(\sqrt{\frac{\mathbf{G}_1^{(e)}}{\mathbf{G}_1^{(d)}}}, \sqrt{\frac{\mathbf{G}_2^{(e)}}{\mathbf{G}_2^{(d)}}}, \ldots, \sqrt{\frac{\mathbf{G}_m^{(e)}}{\mathbf{G}_m^{(d)}}}\right) \tag{A.11}$$

Optimizing $\boldsymbol{\omega}^r$ and $\boldsymbol{\omega}^h$: After updating $\mathbf{S}$, $\mathbf{U}$ and $\mathbf{G}$, we view them as known and take the partial derivative of $\mathcal{J}(\mathbf{G}, \mathbf{S}, \mathbf{U}, \boldsymbol{\omega}^r, \boldsymbol{\omega}^h, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\omega}^r$. In this case, the second, the fourth and the fifth terms on the right of Eq. (4) in the main text are irrelevant to $\boldsymbol{\omega}^r$, and can be ignored. Then we can obtain:

$$\begin{aligned}
\widetilde{\mathcal{J}}(\mathbf{G}, \mathbf{S}, \boldsymbol{\omega}^r) &= \sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r \left\|\mathbf{H}_{ij} \odot \left(\mathbf{R}_{ij} - \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right)\right\|_F^2 + \alpha\|vec(\boldsymbol{\omega}^r)\|_F^2 \\
s.t. \quad & \boldsymbol{\omega}_{ij}^r \geq 0, \sum vec(\boldsymbol{\omega}^r) = 1
\end{aligned} \tag{A.12}$$

Let $\mathbf{L}_{ij} = \left\|\mathbf{H}_{ij} \odot \left(\mathbf{R}_{ij} - \mathbf{G}_i\mathbf{S}_{ij}\mathbf{G}_j^T\right)\right\|_F^2$ be the reconstruction loss for $\mathbf{R}_{ij}$, then Eq. (A.12) can be updated as:

$$\begin{aligned}
\widetilde{\mathcal{J}}(\mathbf{L}, \boldsymbol{\omega}^r) &= vec(\boldsymbol{\omega}^r)^T vec(\mathbf{L}) + \alpha vec(\boldsymbol{\omega}^r)^T vec(\boldsymbol{\omega}^r) \\
s.t. \quad & \boldsymbol{\omega}_{ij}^r \geq 0, \sum vec(\boldsymbol{\omega}^r) = 1
\end{aligned} \tag{A.13}$$

Eq. (A.13) is a quadratic optimization problem with respect to $vec(\boldsymbol{\omega}^r)$. By introducing the Lagrangian multipliers ($\boldsymbol{\delta}$ and $\boldsymbol{\eta}$) for the constraints of $\boldsymbol{\omega}^r$, Eq. (A.13) is formulated as:

$$\begin{aligned}
\widetilde{\mathcal{J}}(\mathbf{L}, \boldsymbol{\omega}^r, \boldsymbol{\delta}, \boldsymbol{\eta}) &= vec(\boldsymbol{\omega}^r)^T vec(\mathbf{L}) + \alpha vec(\boldsymbol{\omega}^r)^T vec(\boldsymbol{\omega}^r) \\
&\quad - \sum_{i,j=1}^{m} \boldsymbol{\delta}_{ij}\boldsymbol{\omega}_{ij}^r - \boldsymbol{\eta}\left(\sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r - 1\right) \\
s.t. \quad & \boldsymbol{\omega}_{ij}^r \geq 0, \sum vec(\boldsymbol{\omega}^r) = 1
\end{aligned} \tag{A.14}$$

Base on the KKT conditions, the optional $\boldsymbol{\omega}^r$ should satisfy the following four conditions:

(i) Stationary condition: $\frac{\partial\widetilde{\mathcal{J}}}{\partial\boldsymbol{\omega}^r} = \mathbf{L} + 2\alpha\boldsymbol{\omega}^r - \delta - \eta = 0$
(ii) Feasible condition: $\boldsymbol{\omega}_{ij}^r \geq 0, \sum_{i,j=1}^{m} \boldsymbol{\omega}_{ij}^r - 1 = 0$
(iii) Dual feasibility: $\boldsymbol{\delta}_{ij} \geq 0, \forall\mathbf{R}_{ij} \in \mathcal{R}$
(iv) Complementary slackness: $\boldsymbol{\delta}_{ij}\boldsymbol{\omega}_{ij}^r = 0, \forall\mathbf{R}_{ij} \in \mathcal{R}$

From the stationary condition, $\boldsymbol{\omega}_{ij}^r$ can be computed as follows:

$$\boldsymbol{\omega}_{ij}^r = \frac{\boldsymbol{\delta}_{ij} + \boldsymbol{\eta} - \mathbf{L}_{ij}}{2\boldsymbol{\alpha}} \tag{A.15}$$

We can find that $\boldsymbol{\omega}_{ij}^r$ depends on the specification of $\boldsymbol{\delta}_{ij}$ and $\boldsymbol{\eta}$, and the specification of $\boldsymbol{\delta}_{ij}$ and $\boldsymbol{\eta}$ can be analyzed in the following cases:

(i) If $\boldsymbol{\eta} > \mathbf{L}_{ij}$, then $\boldsymbol{\omega}_{ij}^r > 0$: because of the complementary slackness $\boldsymbol{\delta}_{ij}\boldsymbol{\omega}_{ij}^r = 0, \boldsymbol{\delta}_{ij} = 0$ and $\boldsymbol{\omega}_{ij}^r = \frac{\boldsymbol{\eta}-\mathbf{L}_{ij}}{2\boldsymbol{\alpha}}$
(ii) If $\boldsymbol{\eta} = \mathbf{L}_{ij}$, because of $\boldsymbol{\delta}_{ij}\boldsymbol{\omega}_{ij}^r = 0$ and $\boldsymbol{\omega}_{ij}^r = \frac{\boldsymbol{\delta}_{ij}}{2\boldsymbol{\alpha}}$, then $\boldsymbol{\delta}_{ij} = 0$ and $\boldsymbol{\omega}_{ij}^r = 0$
(iii) If $\boldsymbol{\eta} < \mathbf{L}_{ij}$, since $\boldsymbol{\omega}_{ij}^r \geq 0$, it requires $\boldsymbol{\delta}_{ij} > 0$; because $\boldsymbol{\delta}_{ij}\boldsymbol{\omega}_{ij}^r = 0$, then $\boldsymbol{\omega}_{ij}^r = 0$

From the above analysis, we can set $\boldsymbol{\omega}_{ij}^r$ as:

$$\boldsymbol{\omega}_{ij}^r = \begin{cases} \frac{\boldsymbol{\eta}-\mathbf{L}_{ij}}{2\boldsymbol{\alpha}} & \text{if} \boldsymbol{\eta} > \mathbf{L}_{ij} \text{ and} \mathbf{R}_{ij} \in \mathcal{R} \\ 0 & \text{if} \boldsymbol{\eta} \leq \mathbf{L}_{ij} \text{ or} \mathbf{R}_{ij} \notin \mathcal{R} \end{cases}, \tag{A.16}$$

Let $\mathbf{v}_L \in \mathbb{R}^{|\mathcal{R}|}$ store the entries of vector $vec(\mathbf{L})$ in ascending order with entries corresponding to $\mathbf{R}_{ij} \notin \mathcal{R}$ removed. Accordingly, $\mathbf{v}^r \in \mathbb{R}^{|\mathcal{R}|}$ stores the corresponding entries of $vec(\boldsymbol{\omega}^r)$ with entries corresponding to $\mathbf{R}_{ij} \notin \mathcal{R}$ removed. For a not too big predefined $\boldsymbol{\alpha}$, there exists $p \in \{1, 2, \ldots, |\mathcal{R}|\}$ with $\mathbf{v}_L(p) < \boldsymbol{\eta}$ and $\mathbf{v}_L(p+1) \geq \eta$, satisfying $\sum\mathbf{v}_L = \sum_{\mathbf{v}_L(p)<\boldsymbol{\eta}} \frac{\boldsymbol{\eta}-\mathbf{v}_L(p)}{2\boldsymbol{\alpha}} = 1$. Then $\mathbf{v}^r(p')$ has the following explicit solution:

$$\mathbf{v}^r(p') = \begin{cases} \frac{\boldsymbol{\eta}-\mathbf{v}_L(p')}{2\boldsymbol{\alpha}} & \text{if} p' \leq p \\ 0 & \text{if} p' > p \end{cases}, \tag{A.17}$$

From $\sum_{p'=1}^{|\mathcal{R}|} \mathbf{v}^r(p') = \sum_{p'=1}^{p} \frac{\boldsymbol{\eta}-\mathbf{v}_L(p')}{2\boldsymbol{\alpha}} = 1$, we can get the value for $\boldsymbol{\eta}$ as:

$$\boldsymbol{\eta} = \frac{2\boldsymbol{\alpha} + \sum_{p'=1}^{p} \mathbf{v}_L(p')}{p} \tag{A.18}$$

Then similar to the solution of $\boldsymbol{\omega}^r$, we find that the first, the third and the fifth terms on the right of Eq. (4) in the main manuscript are irrelevant to $\boldsymbol{\omega}^h$, and can be ignored. Then we can obtain:

$$\begin{aligned}
\widetilde{\mathcal{J}}(\mathbf{G}, \mathbf{S}, \boldsymbol{\omega}^h) &= \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h \left\|\mathbf{X}_{it} - \mathbf{G}_i\mathbf{U}_{it}^T\right\|_F^2 + \beta\left\|vec\left(\boldsymbol{\omega}^h\right)\right\|_F^2 \\
s.t. \quad & \boldsymbol{\omega}_{it}^h \geq 0, \sum vec\left(\boldsymbol{\omega}^h\right) = 1
\end{aligned} \tag{A.19}$$

Let $\mathbf{K}_{it} = \left\|\mathbf{X}_{it} - \mathbf{G}_i\mathbf{U}_{it}^T\right\|_F^2$ be feature matrix reconstruction loss, then Eq. (A.19) can be update as:

$$\begin{aligned}
\widetilde{\mathcal{J}}(\mathbf{K}, \boldsymbol{\omega}^h) &= vec\left(\boldsymbol{\omega}^h\right)^T vec(\mathbf{K}) + \beta vec\left(\boldsymbol{\omega}^h\right)^T vec\left(\boldsymbol{\omega}^h\right) \\
s.t. \quad & \boldsymbol{\omega}_{it}^h \geq 0, \sum vec\left(\boldsymbol{\omega}^h\right) = 1
\end{aligned} \tag{A.20}$$

Eq. (A.20) is a quadratic optimization problem with respect to $vec(\boldsymbol{\omega}^h)$. By introducing the Lagrangian multipliers ($\boldsymbol{\varphi}$ and $\boldsymbol{\mu}$) for the reconstruction loss of $\boldsymbol{\omega}^h$, Eq. (A.20) is formulated as:

$$\begin{aligned}
\widetilde{\mathcal{J}}(\mathbf{K}, \boldsymbol{\omega}^h, \boldsymbol{\varphi}, \boldsymbol{\mu}) &= vec\left(\boldsymbol{\omega}^h\right)^T vec(\mathbf{K}) + \beta vec\left(\boldsymbol{\omega}^h\right)^T vec\left(\boldsymbol{\omega}^h\right) \\
&\quad - \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\varphi}_{it}\boldsymbol{\omega}_{it}^h - \boldsymbol{\mu}\left(\sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h - 1\right) \\
s.t. \quad & \boldsymbol{\omega}_{it}^h \geq 0, \sum vec\left(\boldsymbol{\omega}^h\right) = 1
\end{aligned} \tag{A.21}$$

Base on the KKT conditions, the optional $\boldsymbol{\omega}^h$ should be satisfy the following four conditions:

(i) Stationary condition: $\frac{\partial\widetilde{\mathcal{J}}}{\partial\boldsymbol{\omega}^h} = \mathbf{K} + 2\beta\boldsymbol{\omega}^h - \varphi - \mu = 0$
(ii) Feasible condition: $\boldsymbol{\omega}_{it}^h \geq 0, \sum_{i=1}^{m} \sum_{t=1}^{\max_i t_i} \boldsymbol{\omega}_{it}^h - 1 = 0$
(iii) Dual feasibility: $\boldsymbol{\varphi}_{it} \geq 0, \forall\mathbf{X}_{it} \in \mathcal{R}$
(iv) Complementary slackness: $\boldsymbol{\varphi}_{it}\boldsymbol{\omega}_{it}^h = 0, \forall\mathbf{X}_{it} \in \mathcal{R}$

From the stationary condition, $\boldsymbol{\omega}_{it}^h$ can be computed as follows:

$$\boldsymbol{\omega}_{it}^h = \frac{\boldsymbol{\varphi}_{it} + \boldsymbol{\mu} - \mathbf{K}_{it}}{2\boldsymbol{\beta}} \tag{A.22}$$

We can find that $\boldsymbol{\omega}_{it}^h$ depends on the specification of $\boldsymbol{\varphi}_{it}$ and $\boldsymbol{\mu}$, and the specification of $\boldsymbol{\varphi}_{it}$ and $\boldsymbol{\mu}$ can be analyzed in the following cases:

(i) If $\boldsymbol{\mu} > \mathbf{K}_{it}$, then $\boldsymbol{\omega}_{it}^h > 0$: because of the complementary slackness $\boldsymbol{\varphi}_{it}\boldsymbol{\omega}_{it}^h = 0, \boldsymbol{\varphi}_{it} = 0$ and $\boldsymbol{\omega}_{it}^h = \frac{\boldsymbol{\mu}-\mathbf{K}_{it}}{2\boldsymbol{\beta}}$

(ii) If $\boldsymbol{\mu} = \mathbf{K}_{it}$, because of $\boldsymbol{\varphi}_{it}\boldsymbol{\omega}_{it}^h = 0$ and $\boldsymbol{\omega}_{it}^h = \frac{\boldsymbol{\mu}-\mathbf{K}_{it}}{2\boldsymbol{\beta}}$, then $\boldsymbol{\varphi}_{it} = 0$ and $\boldsymbol{\omega}_{it}^h = 0$

(iii) If $\boldsymbol{\mu} < \mathbf{K}_{it}$, since $\boldsymbol{\omega}_{it}^h \geq 0$, it requires $\boldsymbol{\varphi}_{it} > 0$; because $\boldsymbol{\varphi}_{it}\boldsymbol{\omega}_{it}^h = 0$, then $\boldsymbol{\omega}_{it}^h = 0$

From the above analysis, we can set $\boldsymbol{\omega}_{it}^h$ as:

$$\boldsymbol{\omega}_{it}^h = \begin{cases} \frac{\boldsymbol{\mu}-\mathbf{K}_{it}}{2\boldsymbol{\beta}} & \text{if} \boldsymbol{\mu} > \mathbf{K}_{it} \text{ and} t \leq \max_i t_i \\ 0 & \text{if} \boldsymbol{\mu} \leq \mathbf{K}_{it} \text{ and} t > \max_i t_i \end{cases}, \tag{A.23}$$

where $\mathbf{K}_{it} = \left\|\mathbf{X}_{it} - \mathbf{G}_i\mathbf{U}_{it}^T\right\|_F^2$ is the reconstruction loss for i-th type of objects in t-th heterogeneous feature matrix. Let $\mathbf{v}_K$ store the entries of vector $vec(\mathbf{K})$ in ascending order with entries corresponding to $\mathbf{X}_{it}$, $t > \max_i t_i$ removed. Accordingly, $\mathbf{v}^h$ stores the corresponding entries of $vec(\boldsymbol{\omega}^h)$ with entries corresponding to $\mathbf{X}_{it}$, $t > \max_i t_i$ removed. For a not too big predefined $\boldsymbol{\beta}$, there exists $q \in \{1, 2, \ldots, |\mathcal{X}|\}$, $|\mathcal{X}|$ is the number of feature matrices for all the types, with $\mathbf{v}_K(q) < \boldsymbol{\mu}$ and $\mathbf{v}_K(q+1) \geq \mu$, satisfying $\sum \mathbf{v}_K = \sum_{\mathbf{v}_K(q)<\boldsymbol{\mu}} \frac{\boldsymbol{\mu}-\mathbf{v}_K(q)}{2\boldsymbol{\beta}} = 1$. Then $\mathbf{v}^h(q')$ has the following explicit solution:

$$\mathbf{v}^h(q') = \begin{cases} \frac{\boldsymbol{\mu}-\mathbf{v}_K(q')}{2\boldsymbol{\beta}} & \text{if} q' \leq q \\ 0 & \text{if} q' > q \end{cases}, \tag{A.24}$$

From $\sum_{q'=1}^{|\mathcal{X}|} \mathbf{v}^h(q') = \sum_{q'=1}^{q} \frac{\boldsymbol{\mu}-\mathbf{v}_K(q')}{2\boldsymbol{\beta}} = 1$, we can get the value for $\boldsymbol{\mu}$ as:

$$\boldsymbol{\mu} = \frac{2\boldsymbol{\beta} + \sum_{q'=1}^{q} \mathbf{v}^h(q')}{q} \tag{A.25}$$

We compute $\mathbf{v}_K$ via finding an appropriate q that satisfies $\boldsymbol{\mu} - \mathbf{v}_K(q) > 0$ and $\boldsymbol{\mu} - \mathbf{v}_K(q+1) \leq 0$. We decrease q from $|\mathcal{X}|$ to 1 step by step, and list the procedure in Algorithm 1. The seek of $\mathbf{v}^r$ in Eq. (A.17) can be similarly attained.

Algorithm 1: A method to seek q and compute $\mathbf{v}^h$

Input: $\mathbf{v}_K$, β
Output: q, $\mathbf{v}^h$
Initialize $q = |\mathcal{X}|$, $\mu = 0$;
while $q > 0$ **do**
  Update μ using Eq. (A.25);
  if $\boldsymbol{\mu} - \mathbf{v}_K(q) > 0$ **then**
    break;
  else
    $q \leftarrow q - 1$;
  end
end
$\mathbf{v}^h(q') \leftarrow \frac{\boldsymbol{\mu}-\mathbf{v}_K(q')}{2\boldsymbol{\beta}}$, for $q' = 1, \ldots, q$;
$\mathbf{v}^h(q') \leftarrow 0$, for $q' = q+1, \ldots, |\mathcal{X}|$;
return q and $\mathbf{v}^h$

From the Eq. (A.24), we can see that if $\mathbf{v}_K(q')$ is smaller than $\mathbf{v}_K(q'')$ ($q'' \in \{1, 2, \ldots, |\mathcal{X}|\}$) and $\boldsymbol{\mu} - \mathbf{v}_K(q') > 0$, then a larger weight is assigned to the feature matrix corresponding to $\mathbf{v}_K(q')$ than to $\mathbf{v}_K(q'')$. It can be also observed that if $q' > q$, $\mathbf{v}^h(q') = 0$, which means that the attribute matrices are *automatically removed* from the optimization process. This is because these attribute data matrices with larger reconstruction loss. Similarly, we observe that if $\mathbf{v}_L(p')$ is smaller than $\mathbf{v}_L(p'')$ ($p'' \in \{1, 2, \ldots, |\mathcal{R}|\}$) and $\boldsymbol{\eta} - \mathbf{v}_L(p') > 0$ in Eq. (A.17), then a larger weight is assigned to the relational data matrix corresponding to $\mathbf{v}_L(p')$ than to $\mathbf{v}_L(p'')$. We can further observe that $\mathbf{v}^a(p') = 0$ if $p' > p$, which means that the relational data matrices are *automatically removed* also during the optimization process. This is because these relational matrices with larger reconstruction loss, possibly due to the noisy entries of respective association matrices. Therefore, AHNF has the capability to automatically remove the noisy relational data matrices and noisy attribute data matrices.

References

[1] V. Gligorijević, N. Pržulj, Methods for biological data integration: perspectives and challenges, J. R. Soc. Interface 12 (112) (2015) 20150571.
[2] Y. Li, F.-X. Wu, A. Ngom, A review on machine learning principles for multi-view biological data integration, Brief. Bioinform. 19 (2) (2016) 325–340.
[3] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, M.M. Hoffman, Machine learning for integrating data in biology and medicine: principles, practice, and opportunities, Inf. Fusion 50 (2019) 71–91.
[4] P. Pavlidis, J. Weston, J. Cai, W.S. Noble, Learning gene functional classifications from multiple data types, J. Comput. Biol. 9 (2) (2002) 401–411.
[5] C. Wu, S. Ma, A selective review of robust variable selection with applications in bioinformatics, Brief. Bioinform. 16 (5) (2014) 873–883.
[6] J. Zeng, G. Yu, J. Wang, M. Guo, X. Zhang, DMIL-III: isoform-isoform interaction prediction using deep multi-instance learning method, in: IEEE International Conference on Bioinformatics and Biomedicine, 2019, pp. 171–176.
[7] N. Cesa-Bianchi, M. Re, G. Valentini, Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference, Mach. Learn. 88 (1–2) (2012) 209–241.
[8] X. Chen, G.-Y. Yan, Novel human LncRNA–disease association inference based on LncRNA expression profiles, Bioinformatics 29 (20) (2013) 2617–2624.
[9] G. Yu, H. Rangwala, C. Domeniconi, G. Zhang, Z. Yu, Protein function prediction using multi-label ensemble classification, IEEE/ACM Trans. Comput. Biol.Bioinform. 10 (4) (2013) 1045–1057.
[10] L. Zhang, G. Yu, D. Xia, J. Wang, Protein–protein interactions prediction based on ensemble deep neural networks, Neurocomputing 324 (2019) 10–19.
[11] M. Žitnik, B. Zupan, Data fusion by matrix factorization, IEEE Trans. Pattern Anal. Mach.Intell. 37 (1) (2015) 41–53.
[12] G. Aceto, D. Ciuonzo, A. Montieri, A. Pescapÿÿ, Mimetic: Mobile encrypted traffic classification using multimodal deep learning, Comput. Netw. 165 (2019) 106944.
[13] J. Wang, Z. Yang, C. Domeniconi, X. Zhang, G. Yu, Cooperative driver pathway discovery via fusion of multi-relational data of genes, mirnas and pathways, Brief. Bioinform. 99 (1) (2020) 1–14.
[14] J. Zhao, X. Xie, X. Xu, S. Sun, Multi-view learning overview: recent progress and new challenges, Inf. Fusion 38 (2017) 43–54.
[15] Q. Tan, G. Yu, C. Domeniconi, J. Wang, Z. Zhang, Incomplete multi-view weak-label learning, in: International Joint Conference on Artificial Intelligence, 2018, pp. 2703–2709.
[16] Y. Xing, G. Yu, C. Domeniconi, J. Wang, Z. Zhang, M. Guo, Multi-view multi-instance multi-label learning based on collaborative matrix factorization, in: AAAI Conference on Artificial Intelligence, 33, 2019, pp. 5508–5515.
[17] Z. Yan, J. Liu, L.T. Yang, W. Pedrycz, Data fusion in heterogeneous networks, Inf. Fusion 53 (1) (2020) 1–3.
[18] M. Gönen, E. Alpaydın, Multiple kernel learning algorithms, Journal of Mach. Learn. Res. 12 (7) (2011) 2211–2268.
[19] G. Yu, H. Rangwala, C. Domeniconi, G. Zhang, Z. Zhang, Protein function prediction by integrating multiple kernels, in: International Joint Conference on Artificial Intelligence, 2013, pp. 1869–1875.
[20] W. Lan, M. Li, K. Zhao, J. Liu, F.-X. Wu, Y. Pan, J. Wang, Ldap: a web server for LncRNA-disease association prediction, Bioinformatics 33 (3) (2016) 458–460.
[21] G. Yu, G. Fu, C. Lu, Y. Ren, J. Wang, Brwlda: bi-random walks for predicting LncRNA-disease associations, Oncotarget 8 (36) (2017) 60429–60446.
[22] M. Zhou, X. Wang, J. Li, D. Hao, Z. Wang, H. Shi, L. Han, H. Zhou, J. Sun, Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous LncRNA and disease network, Mol. BioSyst. 11 (3) (2015) 760–769.
[23] H. Wang, H. Huang, C. Ding, Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization, in: Proceedings of ACM International Conference on Information and Knowledge Management, 2011, pp. 279–284.
[24] G. Fu, J. Wang, C. Domeniconi, G. Yu, Matrix factorization-based data fusion for the prediction of LncRNA-disease associations, Bioinformatics 34 (9) (2018) 1529–1537.
[25] G. Yu, Y. Wang, J. Wang, G. Fu, M. Guo, C. Domeniconi, Weighted matrix factorization based data fusion for predicting LncRNA-disease associations, in: IEEE International Conference on Bioinformatics and Biomedicine, 2018, pp. 572–577.
[26] Y. Wang, G. Yu, C. Domeniconi, J. Wang, X. Zhang, M. Guo, Selective matrix factorization for multi-relational data fusion, in: International Conference on Database Systems for Advanced Applications, 2019, pp. 313–329.

[27] C. Lu, M. Yang, F. Luo, F.-X. Wu, M. Li, Y. Pan, Y. Li, J. Wang, Prediction of LncRNA-disease associations based on inductive matrix completion, Bioinformatics 34 (19) (2018) 3357–3364.
[28] A. Biswas, M. Kang, D. Kim, J. Gao, Robust inductive matrix completion strategy to explore associations between lincrnas and human disease phenotypes, IEEE/ACM Trans. Comput. Biol.Bioinform. 99 (1) (2019) 1–12.
[29] J.R. Prensner, A.M. Chinnaiyan, The emergence of LncRNAs in cancer biology, Cancer Discov. 1 (5) (2011) 391–407.
[30] X. Chen, C.C. Yan, X. Zhang, Z.-H. You, Long non-coding RNAs and complex diseases: from experimental results to computational models, Brief. Bioinform. 18 (4) (2017) 558–576.
[31] C. Ding, T. Li, W. Peng, H. Park, Orthogonal nonnegative matrix t-factorizations for clustering, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006, pp. 126–135.
[32] Q. Tan, G. Yu, C. Domeniconi, J. Wang, Z. Zhang, Multi-view weak-label learning based on matrix completion, in: SIAM International Conference on Data Mining, 2018, pp. 450–458.
[33] G. Yu, G. Fu, J. Wang, H. Zhu, Predicting protein function via semantic integration of multiple networks, IEEE/ACM Trans. Comput. Biol.Bioinform. 13 (2) (2016) 220–232.
[34] S. Zhou, E. Zhu, X. Liu, T. Zheng, Q. Liu, J. Xia, J. Yin, Subspace segmentation-based robust multiple kernel clustering, Inf. Fusion 53 (1) (2020) 145–154.
[35] S. Boyd, et al., Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach.Learn. 3 (1) (2011) 1–122.
[36] T.R. Mercer, M.E. Dinger, J.S. Mattick, Long non-coding RNAs: insights into functions, Nat. Rev. Genet. 10 (3) (2009) 155.
[37] L.M. Schriml, et al., Disease ontology: a backbone for disease semantic integration, Nucleic Acids Res. 40 (D1) (2011) D940–D946.
[38] Y. Sun, Y. Zhou, Y. Bai, Q. Wang, J. Bao, Y. Luo, Y. Guo, L. Guo, A long non-coding RNA hottip expression is associated with disease progression and predicts outcome in small cell lung cancer patients, Mol. Cancer 16 (1) (2017) 162.
[39] Y. Sun, B. Hu, Q. Wang, M. Ye, Q. Qiu, Y. Zhou, F. Zeng, X. Zhang, Y. Guo, L. Guo, Long non-coding RNA hottip promotes BCL-2 expression and induces chemoresistance in small cell lung cancer by sponging mir-216a, Cell Death Dis. 9 (2) (2018) 85.
[40] Z. Bao, Z. Yang, Z. Huang, Y. Zhou, Q. Cui, D. Dong, Lncrnadisease 2.0: an updated database of long non-coding RNA-associated diseases, Nucleic Acids Res. 47 (D1) (2018) D1034–D1037.
[41] Y. Gao, P. Wang, Y. Wang, X. Ma, H. Zhi, D. Zhou, X. Li, Y. Fang, W. Shen, Y. Xu, et al., Lnc2cancer v2. 0: updated database of experimentally supported long non-coding rnas in human cancers, Nucleic Acids Res. 47 (D1) (2018) D1028–D1033.
[42] Y. Wang, J. Zhou, Z. Wang, P. Wang, S. Li, Upregulation of sox2 activated LncRNA pvt1 expression promotes breast cancer cell growth and invasion, Biochem. Biophys. Res.Commun. 493 (1) (2017) 429–436.
[43] J. Li, C. Gao, C. Liu, C. Zhou, X. Ma, H. Li, J. Li, X. Wang, L. Qi, Y. Yao, et al., Four LncRNAs associated with breast cancer prognosis identified by coexpression network analysis, J. Cell. Physiol. 234 (8) (2019) 14019–14030.
[44] C. Li, G. Liang, W. Yao, J. Sui, X. Shen, Y. Zhang, S. Ma, Y. Ye, Z. Zhang, W. Zhang, et al., Differential expression profiles of long non-coding rnas reveal potential biomarkers for identification of human gastric cancer, Oncol.Reports 35 (3) (2016) 1529–1540.
[45] E.T. Elsayed, P.E. Salem, A.M. Darwish, H.M. Fayed, Plasma long non-coding RNA hotair as a potential biomarker for gastric cancer, Int. J. Biol. Mark. 33 (4) (2018) 528–533.
[46] Z. Chang, J. Cui, Y. Song, Long noncoding RNA PVT1 promotes EMT via mediating microrna-186 targeting of twist1 in prostate cancer, Gene 654 (2018) 36–42.
[47] H. Chen, Y. He, Y. Hou, D. Chen, S. He, Y. Cao, X. Wu, Long non-coding rna ccat1 promotes the migration and invasion of prostate cancer pc-3 cells, Eur. Rev. Med. Pharmacol. Sci. 22 (10) (2018) 2991–2996.
[48] Y. Zhou, T. Shan, W. Ding, Z. Hua, Y. Shen, Z. Lu, B. Chen, T. Dai, Study on mechanism about long noncoding RNA malat1 affecting pancreatic cancer by regulating hippo-yap signaling, J. Cell. Physiol. 233 (8) (2018) 5805–5814.
[49] L. Ma, F. Wang, C. Du, Z. Zhang, H. Guo, X. Xie, H. Gao, Y. Zhuang, M. Kornmann, H. Gao, et al., Long non-coding rna meg3 functions as a tumour suppressor and has prognostic predictive value in human pancreatic cancer, Oncol. Reports 39 (3) (2018) 1132–1140.
[50] W. Wang, W. Lou, B. Ding, B. Yang, H. Lu, Q. Kong, W. Fan, A novel MRNA-MIRNA-LncRNA competing endogenous rna triple sub-network associated with prognosis of pancreatic cancer, Aging 11 (9) (2019) 2610.
[51] J. Davis, M. Goadrich, The relationship between precision-recall and RoC curves, in: International Conference on Machine Learning, 2006, pp. 233–240.
[52] T. Fawcett, An introduction to RoC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
[53] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: Advances in Neural Information Processing Systems, 2001, pp. 556–562.
[54] C.H. Ding, T. Li, M.I. Jordan, Convex and semi-nonnegative matrix factorizations, IEEE Trans. Pattern Anal. Mach.Intell. 32 (1) (2008) 45–55.
[55] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.