

Kennesaw State University
DigitalCommons@Kennesaw State University

Faculty Publications

12-28-2017

Stable solution to $l_{2,1}$ -based robust inductive matrix completion and its application in linking long noncoding RNAs to human diseases

Ashis Kumer Biswas

University of Colorado - Denver

Dongchul Kim

University of Rio Grande Valley

Mingon Kang

Kennesaw State University, mkang9@kennesaw.edu

Chris Ding

University of Texas at Arlington

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/facpubs>

 Part of the [Computer Sciences Commons](#)

Recommended Citation

Biswas, Ashis Kumer; Kim, Dongchul; Kang, Mingon; and Ding, Chris, "Stable solution to $l_{2,1}$ -based robust inductive matrix completion and its application in linking long noncoding RNAs to human diseases" (2017). *Faculty Publications*. 4391.
<https://digitalcommons.kennesaw.edu/facpubs/4391>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Faculty Publications by an authorized administrator of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

RESEARCH

Open Access



Stable solution to $l_{2,1}$ -based robust inductive matrix completion and its application in linking long noncoding RNAs to human diseases

Ashis Kumer Biswas¹, Dongchul Kim³, Mingon Kang⁴, Chris Ding² and Jean X. Gao^{2*}

From IEEE BIBM International Conference on Bioinformatics & Biomedicine (BIBM) 2016
Shenzhen, China. 15-18 December 2016

Abstract

Backgrounds: A large number of long intergenic non-coding RNAs (lincRNAs) are linked to a broad spectrum of human diseases. The disease association with many other lincRNAs still remain as puzzle. Validation of such links between the two entities through biological experiments are expensive. However, a plethora lincRNA-data are available now, thanks to the High Throughput Sequencing (HTS) platforms, Genome Wide Association Studies (GWAS), etc, which opens the opportunity for cutting-edge machine learning and data mining approaches to extract meaningful relationships among lincRNAs and diseases. However, there are only a few *in silico* lincRNA-disease association inference tools available to date, and none of them utilizes side information of both the entities simultaneously in a single framework.

Methods: The recently developed Inductive Matrix Completion (IMC) technique provides a recommendation platform among two entities considering respective side information about them. However, the formulation of IMC is incapable of handling noise and outliers that may be present in the datasets, while data sparsity consideration is another issue with the standard IMC method. Thus, a robust version of IMC is needed that can solve the two issues. As a remedy, in this paper, we propose Stable Robust Inductive Matrix Completion (SRIMC) that utilizes the $l_{2,1}$ norm based regularization to optimize the objective function with a unique 2-step stable solution approach.

Results: We applied SRIMC to the available association data between human lincRNAs and OMIM disease phenotypes as well as a diverse set of side information about the lincRNAs and the diseases. The method performs better than the state-of-the-art methods in terms of *precision@k* and *recall@k* at the top-*k* disease prioritization to the subject lincRNAs. We also demonstrate that SRIMC is equally effective for querying about novel lincRNAs, as well as predicting rank of a newly known disease for a set of well-characterized lincRNAs.

Conclusions: With the experimental results and computational evaluation, we show that SRIMC is robust in handling datasets with noise and outliers as well as dealing with novel lincRNAs and disease phenotypes.

Keywords: Matrix completion, Inductive learning, Long noncoding RNA, Human disease phenotypes, Association inference

*Correspondence: gao@uta.edu

²Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, Texas 76019, USA

Full list of author information is available at the end of the article

Background

LincRNA-disease association inference problem

It is a surprising fact that, only 2% of the entire human genome codes for proteins [1]. In recent years, it has become evident that the non-protein coding portion of the genome, especially the long intergenic non-coding RNAs (lincRNAs) having length more than 200 bases each with no overlaps with any annotated protein-coding regions, are of critical functional importance. These lincRNAs demonstrate diverse molecular mechanisms and implicate various human diseases [2]. With the advent of the high-throughput genomic technologies, a large number of lincRNAs have been cataloged [3]. However, fully annotating the functions of the lincRNAs and their involvements in human disease implications still remain a challenge for the researchers. Developing machine learning algorithm to rank disease implications by a given lincRNA based on prior knowledge would be beneficial to the community for tackling the challenge.

Limitations of existing algorithms

There are several long non-coding RNA (lncRNA)-disease association inference tools developed in the past few years. But, there is a small number of tools that actually solved the lincRNA-disease inference problem. Due to the complexities in the relationship and the available datasets, only a small number of experimentally validated associations have been reported in the lncRNADisease database [4]. Therefore, using multiple complementary data sources in the algorithm is important to predict potential lincRNA and disease associations. For example, LRLSLDA [5], K-RWRH [6], and TslncRNA-disease [7] belong to a family of network based association identification methods. Each of the algorithms use biological networks, such as lincRNA similarity network and disease similarity network to develop their prediction model. Then by using the model they infer lincRNA-disease connections by either using random walk procedure on a derived biological network or by computing a similarity measure between nodes with known disease implications. The association inference problem can also be tackled through the use of matrix completion based algorithms; Non-negative Matrix Factorization (NMF) belongs to this family of solution strategy. But, it suffers the cold start problem, due to the inability to address the inference predictions of the diseases for novel lincRNAs and vice versa. Furthermore, these methods were presented on a very small set of associations and developed without considering the scalability (e.g., around 200 lincRNAs compared to more than 8000 lincRNAs available to date from the research by [3] remain overlooked). However, the methods utilizing the lincRNA-expression profiles to build similarity networks dealing with a small number of disease

classes. So, they fall short in generalizing their prediction to identify novel diseases-lincRNA connections. Owing to the fact that, a plethora of side information about the lincRNAs and the disease phenotypes are available, and the data is growing extensively every single day. Inductive Matrix Completion (IMC) based algorithms utilize side information about both the lincRNAs and diseases along with the known association evidences to predict missing associations [8]. But, the standard IMC uses the least square error function which is known to be unstable in presence of noise and outliers [9]. A stable and robust version of the IMC is thus needed in this problem.

Outline of our proposed approach

We propose a novel stable robust formulation of IMC using $\ell_{2,1}$ norm based error function, as well as $\ell_{2,1}$ based regularizer. We call the proposed method “robust” because it can handle noise better than the standard IMC. Also, we call the method “stable” because of the fact that it utilizes a 2-step stable strategies to solve the problem.

Summary of contributions

- We first describe a Robust IMC approach that introduces $l_{2,1}$ norms in both its penalty function and the regularization. We then propose Stable Robust IMC (SRIMC) that can handle outliers and noise in the dataset and also joint sparsity. The solution strategy breaks the problem into two separate and independent problems, where each of the sub-problem has stable solution and easy to compute. Hence, in terms of computational complexity and reliability SRIMC should be a better option.
- We provide an application of our RIMC and SRIMC methods to solve the lincRNA-disease association inference problem. We show that RIMC and SRIMC can perform induction to decipher associations between a novel disease and a novel lincRNA, based on the side information about them we have, that are not provided during learning phase. This is unlike the traditional matrix factorization methods and network-based inference methods discussed earlier which are transductive in nature.
- We demonstrate that the integration of diverse features of the lincRNAs and the diseases available through publicly available data-servers can overcome worse predictive performance issue faced by the inference tools which occurs due to the extreme sparsity inherent to the lincRNA-disease association dataset.
- We present a comparison of our proposed RIMC, SRIMC with standard IMC as well as the state-of-the-art lincRNA-disease association methods.

The rest of the paper is organized as follows. In “Methods” section we propose the robust IMC formulation using $\ell_{2,1}$ norm, underline the advantages of the proposed algorithm compared with the standard IMC as well as standard NMF approaches. Here we also show the correctness of the proposed algorithm. In “Results” section we present the experimental setup and the dataset. In “Discussions” section, we present the results of the experiments, and comparative study on the performance of the proposed algorithm with the base-line methods. Finally, in “Conclusions” section we conclude the paper. A preliminary version of this work has been reported in [10].

Methods

Stable robust inductive Matrix completion (SRIMC) strategy
 In this section we review standard Inductive Matrix Completion method; then we present our robust IMC (RIMC) formulation. And finally, we present the Stable Robust IMC (SRIMC) algorithm. Later, we provide a computational algorithm for our proposed method along with the correctness of the algorithm.

Review on standard IMC

The Inductive Matrix Completion approach [8] includes side information of both the row and column entities. The formulation solves the issue “cold-start” problem in a transductive setup (e.g., standard NMF, etc.). Therefore we can predict association between new entities that are not included in the data matrix available during the training time. Let’s consider an association matrix, $A \in \mathbb{R}^{M \times N}$ denoting the links between M row entities and N column entities. We also have side information of both the row and column entities and the information is encapsulated in two matrices $X \in \mathbb{R}^{M \times m}$ and $Y \in \mathbb{R}^{N \times n}$ containing m features of the M row entities and n features of the N column entities respectively. Equation 1 defines the the objective function of the standard IMC.

$$\min_{W,H} \varphi = \frac{1}{2} \left\| A - XWH^T Y^T \right\|_F^2 + \frac{\lambda_1}{2} \left\| W \right\|_F^2 + \frac{\lambda_2}{2} \left\| H \right\|_F^2$$

such that, $W \geq 0, H \geq 0,$ (1)

where λ_1, λ_2 are the regularization parameters that weighing between the accrued loss on the observed entries and the trace norm regularization constraints. Here, an entry A_{ij} is modeled as $\mathbf{x}_i^T Z \mathbf{y}_j$, where $Z \in \mathbb{R}^{m \times n}$ is a low-rank matrix to be recovered by solving Eq. 1. It is solved in a way that Z becomes the multiplication of two factor matrices W and H , that is, WH^T , where $W \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{n \times r}$. Equation 1 can be easily solved using Algorithm 1.

After Algorithm 1 returns, we get the two factor matrices W and H . These two matrices can be used to compute missing association scores between the row and the column entities. It can also provide prediction score for an

Algorithm 1: Solve Standard Inductive Matrix Completion

Data: $A \in \mathbb{R}^{M \times N}, X \in \mathbb{R}^{M \times m}, Y \in \mathbb{R}^{N \times n}$

Result: $W \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{n \times r}$

Initialize W and H with random number maintaining the non-negativity constraints $W_{ik} \geq 0, H_{jk} \geq 0;$

while convergence criteria not met **do**

$$H_{jk} \leftarrow H_{jk} \frac{(Y^T A^T X W)_{jk}}{(Y^T Y H W^T X^T X W + \lambda_2 H)_{jk}} \quad (2)$$

$$W_{ik} \leftarrow W_{ik} \frac{(X^T A Y H)_{ik}}{(X^T X W H^T Y^T Y H + \lambda_1 W)_{ik}} \quad (3)$$

end

return W, H

association between a known row entity with an new column entity, or a known column entity with a new row entity, or both new row and column entities.

Robust IMC (RIMC) formulation

One limitation of the standard IMC is that it is prone to outliers in the given dataset. Given $A \in \mathbb{R}^{M \times N}, X \in \mathbb{R}^{M \times m}, Y \in \mathbb{R}^{N \times n}$, the loss function of the standard IMC is:

$$\left\| A - XWH^T Y^T \right\|_F^2 = \sum_{i=1}^M \left\| A_{i,:} - X_{i,:} W H^T Y^T \right\|_2^2 \quad (4)$$

Here, a squared residual error gets accumulated in each iteration in the optimization step, meaning only a few outliers may result in large error. Another shortcoming of the the standard IMC is that it can not handle joint sparsity across feature data matrices X and Y . Therefore, a solution to each of the limitations is needed. The initial hypothesis of RIMC was presented by [10]. The robust IMC, instead of using the ℓ_2 norm based loss function involves $\ell_{2,1}$ norm in defining the loss function which is:

$$\left\| A - XWH^T Y^T \right\|_{2,1} = \sum_{i=1}^M \sqrt{\sum_{j=1}^N (A - XWH^T Y^T)_{ij}^2} \quad (5)$$

Due to the fact that the errors are not squared in each step, the approach has great advantage to handle outliers than that of standard IMC based approaches. The generalized objective function of the RIMC can be stated as:

$$\min_{W,H} \varphi = \left\| A - XWH^T Y^T \right\|_{2,1} + \lambda_1 R(W) + \lambda_2 R(H)$$

such that, $W \geq 0, H \geq 0$ (6)

Here, we have several options as the regularization function $R(\cdot)$; such as: $R_1(B) = \|B\|_F^2, R_2(B) = \sum_{i=1}^M \|B_{i,:}\|_1,$

$R_3(B) = \sum_{i=1}^M \|B_{i,:}\|_2^0$ and $R_4(B) = \sum_{i=1}^M \|B_{i,:}\|_2$. Here, $R_1(\cdot)$ is the ridge regularization and is adapted in the standard IMC formulation, $R_2(\cdot)$ is the LASSO regularization which is a non-convex function and difficult to optimize. $R_3(\cdot)$ involves the ℓ_0 norm and is the most desirable [11], and $R_4(\cdot)$ employs the $\ell_{2,1}$ norm. $R_4(\cdot)$ was chosen because the function is convex and we can easily optimize the objective function involving this kind of regularizer [12].

Thus given the data matrices A, X, Y , in this paper we optimize the following robust IMC formulation:

$$\min_{W, H} \varphi = \|A - XWH^T Y^T\|_{2,1} + \lambda_1 \|W\|_{2,1} + \lambda_2 \|H\|_{2,1}$$

such that, $W \geq 0, H \geq 0$ (7)

Algorithm for RIMC (version 1)

In order to solve Eq. 7, Algorithm 2 can be adapted [10].

Algorithm 2: Solve Robust Inductive Matrix Completion (Version 1)

Data: $A \in \mathbb{R}^{M \times N}, X \in \mathbb{R}^{M \times m}, Y \in \mathbb{R}^{N \times n}$

Result: $W \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{n \times r}$

Initialize W and H with random number maintaining the non-negativity constraints $W_{ik} \geq 0, H_{jk} \geq 0$;
 Initialize $D \in \mathbb{R}^{M \times M}, P \in \mathbb{R}^{m \times m}, Q \in \mathbb{R}^{n \times n}$ as identity matrices.;

while convergence criteria not met **do**

$$H_{jk} = H_{jk} \frac{(e_n e_M^T A e_N e_M^T D e_M e_n^T Y^T e_N e_M^T X W)_{ik}}{(e_n e_M^T X W H^T Y^T e_N e_M^T D e_M e_n^T Y^T e_N e_M^T X W + \lambda_2 Q H)_{jk}}$$

$$D_{ii} = 1 / \sqrt{\sum_{j=1}^N (A - XWH^T Y^T)_{ij}^2}$$

$$Q_{ii} = 1 / \sqrt{\sum_{j=1}^r H_{ij}^2}$$

$$W_{ik} = W_{ik} \frac{(e_m e_M^T A e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H)_{ik}}{(e_m e_M^T X W H^T Y^T e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H + \lambda_1 P W)_{ik}}$$

$$P_{ii} = 1 / \sqrt{\sum_{j=1}^r W_{ij}^2}$$

end

return W, H

Correctness of the RIMC Algorithm (version 1)

Theorem 1 At convergence, the converged solution W^* of the updating rule in Algorithm 2 satisfies the KKT condition.

Proof The KKT condition for W with constraints $W_{ik} \geq 0$, with $i = 1 \dots m, k = 1 \dots r$ is:

$$\frac{\partial \varphi(W)}{\partial W_{ik}} W_{ik} = 0, \forall i, k \tag{8}$$

Now, the partial derivative is

$$\begin{aligned} \frac{\partial \varphi(W)}{\partial W_{ik}} &= \sum_{\alpha=1}^M \frac{1}{\sqrt{\sum_{\beta=1}^N (A - XWH^T Y^T)_{\alpha\beta}^2}} \cdot \sum_{\beta'=1}^N (A - XWH^T Y^T)_{\alpha\beta'} \cdot \\ &\frac{\partial}{\partial W_{ik}} (A - XWH^T Y^T)_{\alpha\beta'} + \lambda_1 \sum_{\alpha=1}^m \frac{1}{\sum_{\beta=1}^r W_{\alpha\beta}^2} \cdot W_{\alpha\beta} \cdot \frac{\partial W_{\alpha\beta}}{\partial W_{ik}} \\ &= \sum_{\alpha=1}^M D_{\alpha\alpha} \sum_{\beta'=1}^N (A - XWH^T Y^T)_{\alpha\beta'} (-X^T e_M e_N^T Y H)_{ik} \\ &+ \lambda_1 P_{ii} W_{ik} \end{aligned} \tag{9}$$

where $e_s = (1, \dots, 1)^T \in \mathbb{R}^s$ is a vector with all 1s. Also, $D, P \in \mathbb{R}^{m \times m}$ are the two diagonal matrices with the diagonal elements given by:

$$D_{ii} = 1 / \sqrt{\sum_{j=1}^N (A - XWH^T Y^T)_{ij}^2} \tag{10}$$

$$P_{ii} = 1 / \sqrt{\sum_{j=1}^r W_{ij}^2} \tag{11}$$

Now, let us continue from Eq. 9:

$$\begin{aligned} \frac{\partial \varphi(W)}{\partial W_{ik}} &= - (e_m e_M^T A e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H)_{ik} \\ &+ (e_m e_M^T X W H^T Y^T e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H)_{ik} \\ &+ \lambda_1 P_{ii} W_{ik} \end{aligned} \tag{12}$$

Thus, the KKT condition for W is:

$$\begin{aligned} &\left[- (e_m e_M^T A e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H)_{ik} \right. \\ &+ (e_m e_M^T X W H^T Y^T e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H)_{ik} \\ &\left. + \lambda_1 P_{ii} W_{ik} \right] W_{ik} = 0, \forall i, k \end{aligned} \tag{13}$$

But, once W converges (according to Algorithm 2), the converged solution (W^*) satisfies

$$W_{ik}^* = W_{ik}^* \frac{(e_m e_M^T A e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H)_{ik}}{(e_m e_M^T X W^* H^T Y^T e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H + \lambda_1 P W^*)_{ik}}$$

which can be written as

$$\begin{aligned} & \left[- \left(e_m e_M^T A e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H \right)_{ik} \right. \\ & \left. + \left(e_m e_M^T X W^* H^T Y^T e_N e_M^T D e_M e_m^T X^T e_M e_N^T Y H \right)_{ik} \right. \\ & \left. + \lambda_1 P_{ii} W_{ik}^* \right] W_{ik}^* = 0 \end{aligned}$$

This is identical to Eq. 43. Thus, the converged solution W^* satisfies the KKT condition. \square

Theorem 2 *At convergence, the converged solution H^* of the updating rule in Algorithm 2 satisfies the KKT condition.*

Proof The KKT condition for H with constraints $H_{jk} \geq 0$, with $j = 1 \dots n, k = 1 \dots r$ is:

$$\frac{\partial \varphi(H)}{\partial H_{jk}} H_{jk} = 0, \forall j, k \quad (14)$$

Now, the partial derivative is

$$\begin{aligned} \frac{\partial \varphi(H)}{\partial H_{jk}} &= \sum_{\alpha=1}^M \frac{1}{\sqrt{\sum_{\beta=1}^N (A - XWH^T Y^T)_{\alpha\beta}^2}} \cdot \sum_{\beta'=1}^N (A - XWH^T Y^T)_{\alpha\beta'} \cdot \\ & \frac{\partial}{\partial H_{jk}} (A - XWH^T Y^T)_{\alpha\beta'} + \lambda_2 \sum_{\alpha=1}^n \frac{1}{\sum_{\beta=1}^r H_{\alpha\beta}^2} \cdot H_{\alpha\beta} \cdot \frac{\partial H_{\alpha\beta}}{\partial H_{jk}} \\ &= \sum_{\alpha=1}^M D_{\alpha\alpha} \sum_{\beta'=1}^N (A - XWH^T Y^T)_{\alpha\beta'} (-Y^T e_N e_M^T XW)_{jk} \\ &+ \lambda_2 Q_{jj} H_{jk}, \end{aligned} \quad (15)$$

where D is already defined in Eq. 10, and $Q \in \mathbb{R}^{n \times n}$ is a diagonal matrix with the diagonal elements given by:

$$Q_{jj} = 1 / \sqrt{\sum_{i=1}^r H_{ji}^2} \quad (16)$$

Now, let us continue from Eq. 15:

$$\begin{aligned} \frac{\partial \varphi(H)}{\partial H_{jk}} &= - \left(e_n e_M^T A e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW \right)_{jk} \\ & \left(e_n e_M^T XWH^T Y^T e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW \right)_{jk} \\ &+ \lambda_2 Q_{jj} H_{jk} \end{aligned} \quad (17)$$

Thus, the KKT condition for H is:

$$\begin{aligned} & \left[- \left(e_n e_M^T A e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW \right)_{jk} \right. \\ & \left. + \left(e_n e_M^T XWH^T Y^T e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW \right)_{jk} \right. \\ & \left. + \lambda_2 Q_{jj} H_{jk} \right] H_{jk} = 0, \forall j, k \end{aligned} \quad (18)$$

But, once H converges (according to Algorithm 2), the converged solution (H^*) satisfies

$$H_{jk}^* = H_{jk}^* \frac{\left(e_n e_M^T A e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW \right)_{ik}}{\left(e_n e_M^T XWH^* Y^T e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW + \lambda_2 QH^* \right)_{jk}}$$

which can be written as

$$\begin{aligned} & \left[- \left(e_n e_M^T A e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW \right)_{jk} \right. \\ & \left. + \left(e_n e_M^T XWH^* Y^T e_N e_M^T D e_M e_n^T Y^T e_N e_M^T XW \right)_{jk} \right. \\ & \left. + \lambda_2 Q_{jj} H_{jk}^* \right] H_{jk}^* = 0 \end{aligned} \quad (19)$$

This is identical to Eq. 47. Thus, the converged solution H^* satisfies the KKT condition. \square

Algorithm for RIMC (version 2)

We can also solve the robust IMC optimization problem (Eq. 7) without the use of the e vectors. It is demonstrated in Algorithm 3.

Algorithm 3: Solve Robust Inductive Matrix Completion (Version 2)

Data: $A \in \mathbb{R}^{M \times N}, X \in \mathbb{R}^{M \times m}, Y \in \mathbb{R}^{N \times n}$

Result: $W \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{n \times r}$

Initialize W and H with random number maintaining the non-negativity constraints $W_{ik} \geq 0, H_{jk} \geq 0$;
Initialize $D \in \mathbb{R}^{M \times M}, P \in \mathbb{R}^{m \times m}, Q \in \mathbb{R}^{n \times n}$ as identity matrices.;

while convergence criteria not met **do**

$$H_{\gamma\psi} = H_{\gamma\psi} \frac{(Y^T A^T D X W)_{\gamma\psi}}{(Y^T Y H W^T X^T D X W + \lambda_2 Q H)_{\gamma\psi}}$$

$$D_{ii} = 1 / \sqrt{\sum_{j=1}^N (A - XWH^T Y^T)_{ij}^2}$$

$$Q_{ii} = 1 / \sqrt{\sum_{j=1}^r H_{ij}^2}$$

$$W_{\alpha\beta} = W_{\alpha\beta} \frac{(X^T D A Y H)_{\alpha\beta}}{(X^T D X W H^T Y^T Y H + \lambda_1 P W)_{\alpha\beta}}$$

$$P_{ii} = 1 / \sqrt{\sum_{j=1}^r W_{ij}^2}$$

end
return W, H

Convergence of the RIMC Algorithm (version 2)

Here, we present the proof of the convergence of Algorithm 3.

Theorem 3 Algorithm 3 will monotonically decrease the objective function of the problem (Eq. 7) in each iteration and converge to the global optimum of the problem.

However, it can be rephrased using the following two statements:

- (A) Updating H using the H update equation in Algorithm 3 while fixing W, the objective function of the problem (Eq. 7) monotonically decreases.
- (B) Updating W using the W update equation in Algorithm 3 while fixing H, the objective function of the problem (Eq. 7) monotonically decreases.

Proof We prove Theorem 3 (A, B) separately in the following two sections. □

Proof of Theorem 3(A): Updating of H

Proof We now focus on proving Theorem 3(A). The proof requires the following two lemmas: (Lemma 4 and 5). □

Lemma 4 Let, $H^{(t)}$ be the H at the t^{th} iteration, and $H^{(t+1)}$ is obtained from the next iteration. Then, under the H update rule in Algorithm 3, the following inequality holds.

$$\begin{aligned} & \text{tr} \left(\left(A - XWH^{(t+1)T} Y^T \right)^T D \left(A - XWH^{(t+1)T} Y^T \right) \right) \\ & \quad + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^{(t+1)T} Q H^{(t+1)} \right) \\ & \leq \text{tr} \left(\left(A - XWH^{(t)T} Y^T \right)^T D \left(A - XWH^{(t)T} Y^T \right) \right) \\ & \quad + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^{(t)T} Q H^{(t)} \right), \end{aligned} \tag{20}$$

where, $D_{ii} = 1 / \sqrt{\sum_{j=1}^N (A - XWH^{(t)T} Y^T)_{ij}^2}$ and $Q_{ii} = 1 / \sqrt{\sum_{j=1}^r H_{ij}^{(t)T}}$

The proof of Lemma 4 is given in section Proof of Lemma 4.

Lemma 5 Under the H update rule in Algorithm 3, the following inequality holds:

$$\begin{aligned} & \left\| A - XWH^{(t+1)T} Y^T \right\|_{2,1} + \lambda_1 \|W\|_{2,1} + \lambda_2 \|H^{(t+1)}\|_{2,1} \\ & - \left\| A - XWH^{(t)T} Y^T \right\|_{2,1} - \lambda_1 \|W\|_{2,1} - \lambda_2 \|H^{(t)}\|_{2,1} \leq \end{aligned}$$

$$\begin{aligned} & \frac{1}{2} \left\{ \text{tr} \left(\left(A - XWH^{(t+1)T} Y^T \right)^T D \left(A - XWH^{(t+1)T} Y^T \right) \right) \right. \\ & \quad + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^{(t+1)T} Q H^{(t+1)} \right) \\ & \quad - \text{tr} \left(\left(A - XWH^{(t)T} Y^T \right)^T D \left(A - XWH^{(t)T} Y^T \right) \right) \\ & \quad \left. - \lambda_1 \text{tr} \left(W^T P W \right) - \lambda_2 \text{tr} \left(H^{(t)T} Q H^{(t)} \right) \right\}, \end{aligned} \tag{21}$$

where D, P, Q matrices are defined earlier.

The proof of Lemma 5 is given in section Proof of Lemma 5.

Now, if we take a look at the right hand side of the inequality in Eq. 21, the value is negative or zero according to Lemma 4. This completes the proof that the objective function of Eq. 7 decreases monotonically.

Proof of Theorem 3(B): updating of W

Proof We now focus on proving Theorem 3(B). The proof requires the following two lemmas: (Lemma 6 and 7). □

Lemma 6 Let, $W^{(t)}$ be the W at the t^{th} iteration, and $W^{(t+1)}$ is obtained from the next iteration. Then, under the W update rule in Algorithm 3, the following inequality holds.

$$\begin{aligned} & \text{tr} \left(\left(A - XW^{(t+1)} H^T Y^T \right)^T D \left(A - XW^{(t+1)} H^T Y^T \right) \right) \\ & \quad + \lambda_1 \text{tr} \left(W^{(t+1)T} P W^{(t+1)} \right) + \lambda_2 \text{tr} \left(H^T Q H \right) \\ & \leq \text{tr} \left(\left(A - XW^{(t)} H^T Y^T \right)^T D \left(A - XW^{(t)} H^T Y^T \right) \right) \\ & \quad + \lambda_1 \text{tr} \left(W^{(t)T} P W^{(t)} \right) + \lambda_2 \text{tr} \left(H^T Q H \right), \end{aligned} \tag{22}$$

where, D, P, Q are defined earlier.

Proof of Lemma 6 is provided in section Proof of Lemma 6.

Lemma 7 Under the W update rule in Algorithm 3, the following inequality holds:

$$\begin{aligned} & \left\| A - XW^{(t+1)} H^T Y^T \right\|_{2,1} + \lambda_1 \|W^{(t+1)}\|_{2,1} + \lambda_2 \|H\|_{2,1} \\ & - \left\| A - XW^{(t)} H^T Y^T \right\|_{2,1} - \lambda_1 \|W^{(t)}\|_{2,1} - \lambda_2 \|H\|_{2,1} \leq \\ & \frac{1}{2} \left\{ \text{tr} \left(\left(A - XW^{(t+1)} H^T Y^T \right)^T D \left(A - XW^{(t+1)} H^T Y^T \right) \right) \right. \\ & \quad + \lambda_1 \text{tr} \left(W^{(t+1)T} P W^{(t+1)} \right) + \lambda_2 \text{tr} \left(H^T Q H \right) \\ & \quad - \text{tr} \left(\left(A - XW^{(t)} H^T Y^T \right)^T D \left(A - XW^{(t)} H^T Y^T \right) \right) \\ & \quad \left. - \lambda_1 \text{tr} \left(W^{(t)T} P W^{(t)} \right) - \lambda_2 \text{tr} \left(H^T Q H \right) \right\}, \end{aligned} \tag{23}$$

where D, P, Q matrices are defined earlier.

Proof of Lemma 7 is provided in section Proof of Lemma 7.

Now, if we take a look at the right hand side of the inequality in Eq. 23, the value is negative or zero according to Lemma 6. This completes the proof that the objective function of Eq. 7 decreases monotonically.

Proof of Lemma 4

Proof We can re-write Eq. 20 as follows:

$$J(H^{(t+1)}) \leq J(H^{(t)}), \tag{24}$$

where

$$J(H) = tr(A - XWH^T Y^T)^T D(A - XWH^T Y^T) + \lambda_1 tr(W^T P W) + \lambda_2 tr(H^T Q H) \tag{25}$$

And, according to the statement of Lemma 4, under the H update rule Algorithm 3, $J(H)$ monotonically decreases. In order to prove the statement, we follow the approaches utilizing auxiliary functions [13, 14]. \square

Definition 1 $G(H, H')$ is an auxiliary function for the function $J(H)$ if $G(H, H') \geq J(H)$ for all H' and $G(H, H) = J(H)$.

Now, we define:

$$H^{(t+1)} = \underset{H}{\operatorname{argmin}} G(H, H^{(t)})$$

So, we have

$$J(H^{(t+1)}) = G(H^{(t+1)}, H^{(t+1)}) \leq G(H^{(t+1)}, H^{(t)}) \leq G(H^{(t)}, H^{(t)}) = J(H^{(t)})$$

This proves that $J(H^{(t)})$ is monotonically decreasing.

Now the important steps in the remainder of the proof are: (a) determine a proper auxiliary function, and (b) find the global minima of the auxiliary function.

Lemma 8 *The function*

$$G(H, H') = tr(A^T D A) - 2tr(Y H W^T X^T D A) + \lambda_1 tr(W^T P W) + \lambda_2 tr(H^T Q H) + \sum_{i=1}^n \sum_{j=1}^r \frac{(Y^T Y H' W^T X^T D X W)_{ij} H_{ij}^2}{H'_{ij}} \tag{26}$$

is an auxiliary function for J .

Proof Now $J(H)$ of Eq. 25 can be re-written as:

$$J(H) = tr(A^T D A) - 2tr(Y H W^T X^T D A) + \lambda_1 tr(W^T P W) + \lambda_2 tr(H^T Q H) + tr(H^T Y^T Y H W^T X^T D X W) \tag{27}$$

Now we will be applying the following inequality of matrices according to the investigations by [14, 15]:

$$tr(H^T \Lambda H B) \leq \sum_i \sum_j (\Lambda H' B)_{ji} \frac{H_{ij}^2}{H'_{ij}}, \tag{28}$$

where, Λ, B, H are non-negative matrices, and Λ, B are symmetric matrices. And obviously the equality holds in Eq. 28 when $H = H'$.

In Eq. 28, if we do the substitutions: $\Lambda = Y^T Y, B = W^T X^T D X W, H = H, H' = H'$, we see that the fifth term of Eq. 27 is smaller than the fifth term of Eq. 26. However, the equality holds when $H = H'$. Thus $G(H, H')$ in Eq. 26 is an auxiliary function of $J(H)$. \square

Now, we need to find the global minimum of Eq. 26. Let $f(H) = G(H, H')$. The gradient of $f(H)$ is

$$\frac{\partial f(H)}{\partial H_{ij}} = -2(Y^T A^T D X W)_{ij} + 2\lambda_2(QH)_{ij} + 2 \frac{(Y^T Y H' W^T X^T D X W)_{ij} H_{ij}}{H'_{ij}} \tag{29}$$

However, the second order derivative (i.e., the Hessian matrix) would be

$$\frac{\partial^2 f(H)}{\partial H_{ij} \partial H_{kl}} = 2(Q)_{jl} \delta_{ik} \delta_{k\beta} + \left(2 \frac{(Y^T D Y H' W^T X^T X W)_{ij}}{H'_{ij}} \right) \delta_{jl} \delta_{ik} \tag{30}$$

The Hessian matrix (Eq. 30 is semi-positive definite implying that $f(H) = G(H, H')$ is a convex function. Thus, there exists a unique global minimum for $f(H)$. The global minimum can be obtained by setting the gradient of $f(H)$ to zero and solve for H . Thus from Eq. 29 we get

$$H_{ij} = H'_{ij} \frac{(Y^T A^T D X W)_{ij}}{(Y^T Y H' W^T X^T D X W + \lambda_2 Q H)_{ij}} \tag{31}$$

By replacing $H^{(t+1)} = H$ and $H^{(t)} = H'$, we would obtain the H update rule in Algorithm 3. Therefore, under this rule, the objective function $J(H)$ of Eq. 25 decreases monotonically, and hence completes the proof.

Proof of Lemma 5

Proof We know that,

$$\begin{aligned} & \text{tr} \left(A - XWH^{(t)T} Y^T \right)^T D \left(A - XWH^{(t)T} Y^T \right) \\ & + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^{(t)T} Q H^{(t)} \right) \\ & = \sum_{i=1}^M \sum_{j=1}^N \left(A - XWH^{(t)T} Y^T \right)_{ij} D_{ii} \\ & + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \sum_{k=1}^n \sum_{l=1}^r H_{kl}^{(t)2} Q_{kk} \\ & = \sum_{i=1}^M \left\| A_i - \left(XWH^{(t)T} Y^T \right)_i \right\|^2 D_{ii} \\ & + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \sum_{k=1}^n \left\| H_k^{(t)} \right\|^2 Q_{kk} \end{aligned}$$

Similarly, we can see that

$$\begin{aligned} & \text{tr} \left(A - XWH^{(t+1)T} Y^T \right)^T D \left(A - XWH^{(t+1)T} Y^T \right) \\ & + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^{(t+1)T} Q H^{(t+1)} \right) \\ & = \sum_{i=1}^M \left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\|^2 D_{ii} \\ & + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \sum_{k=1}^n \left\| H_k^{(t+1)} \right\|^2 Q_{kk} \end{aligned}$$

Then, the right-hand side (*r.h.s*) of Eq. 21 becomes

$$\begin{aligned} r.h.s & = \frac{1}{2} \sum_{i=1}^M \left(\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\|^2 \right. \\ & - \left\| A_i - \left(XWH^{(t)T} Y^T \right)_i \right\|^2 \right) D_{ii} + \lambda_2 \sum_{k=1}^n \left(\left\| H_k^{(t+1)} \right\|^2 \right. \\ & - \left. \left\| H_k^{(t)} \right\|^2 \right) Q_{kk} \\ & = \frac{1}{2} \sum_{i=1}^M \left(\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\|^2 D_{ii} - \frac{1}{D_{ii}} \right) \\ & + \lambda_2 \sum_{k=1}^n \left(\left\| H_k^{(t+1)} \right\|^2 Q_{kk} - \frac{1}{Q_{kk}} \right) \end{aligned}$$

And, the left-hand side (*l.h.s*) of Eq. 21 becomes

$$\begin{aligned} l.h.s & = \sum_{i=1}^M \left(\sqrt{\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\|^2} \right. \\ & - \left. \sqrt{\left\| A_i - \left(XWH^{(t)T} Y^T \right)_i \right\|^2} \right) \\ & + \lambda_2 \sum_{k=1}^n \left(\sqrt{\left\| H_k^{(t+1)} \right\|^2} - \sqrt{\left\| H_k^{(t)} \right\|^2} \right) \\ & = \sum_{i=1}^M \left(\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\| \right. \\ & - \left. \left\| A_i - \left(XWH^{(t)T} Y^T \right)_i \right\| \right) \\ & + \lambda_2 \sum_{k=1}^n \left(\sqrt{\left\| H_k^{(t+1)} \right\|^2} - \sqrt{\left\| H_k^{(t)} \right\|^2} \right) \\ & = \sum_{i=1}^M \left(\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\| - \frac{1}{D_{ii}} \right) \\ & + \lambda_2 \left(\sum_{k=1}^n \left\| H_k^{(t+1)} \right\| - \frac{1}{Q_{kk}} \right) \end{aligned}$$

Now, we compute the difference between the *l.h.s* and *r.h.s*,

$$\begin{aligned} l.h.s - r.h.s & = \sum_{i=1}^M \left(\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\| \right. \\ & - \left. \left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\|^2 \frac{D_{ii}}{2} - \frac{1}{2D_{ii}} \right) \\ & + \lambda_2 \sum_{k=1}^n \left(\left\| H_k^{(t+1)} \right\| - \left\| H_k^{(t+1)} \right\|^2 \frac{Q_{kk}}{2} - \frac{1}{2Q_{kk}} \right) \\ & = \sum_{i=1}^M \frac{D_{ii}}{2} \left(\frac{\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\|}{D_{ii}} \right. \\ & - \left. \left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\|^2 - \frac{1}{D_{ii}^2} \right) \\ & + \lambda_2 \sum_{k=1}^n \frac{Q_{kk}}{2} \left(\frac{\left\| H_k^{(t+1)} \right\|}{Q_{kk}} - \left\| H_k^{(t+1)} \right\|^2 - \frac{1}{Q_{kk}^2} \right) \\ & = \sum_{i=1}^M \frac{(-D_{ii})}{2} \left(\left\| A_i - \left(XWH^{(t+1)T} Y^T \right)_i \right\| - \frac{1}{D_{ii}} \right)^2 \\ & + \lambda_2 \sum_{k=1}^n \frac{(-Q_{kk})}{2} \left(\left\| H_k^{(t+1)} \right\| - \frac{1}{Q_{kk}} \right)^2 \\ & \leq 0 \end{aligned}$$

The above inequality holds because, D, Q are non-negative matrices, and the sum of non-positive numbers is always non-positive. This completes the proof. \square

Proof of Lemma 6

Proof We can re-write Eq. 22 as follows:

$$J(W^{(t+1)}) \leq J(W^{(t)}), \tag{32}$$

where

$$J(W) = \text{tr} \left(A - XWH^T Y^T \right)^T D \left(A - XWH^T Y^T \right) + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^T Q H \right) \tag{33}$$

And, according to the statement of Lemma 6, under the W update rule in Algorithm 3, $J(W)$ monotonically decreases. In order to prove the statement, we follow the approaches utilizing auxiliary functions [13, 14]. \square

Definition 2 $G(W, W')$ is an auxiliary function for the function $J(W)$ if $G(W, W') \geq J(W)$ for all W' and $G(W, W) = J(W)$.

Now, we define:

$$W^{(t+1)} = \underset{W}{\text{argmin}} \quad G \left(W, W^{(t)} \right)$$

So, we have

$$J \left(W^{(t+1)} \right) = G \left(W^{(t+1)}, W^{(t+1)} \right) \leq G \left(W^{(t+1)}, W^{(t)} \right) \leq G \left(W^{(t)}, W^{(t)} \right) = J \left(W^{(t)} \right)$$

This proves that $J(W^{(t)})$ is monotonically decreasing. Now the important steps in the remainder of the proof are: (a) determine a proper auxiliary function, and (b) find the global minima of the auxiliary function.

Lemma 9 The function

$$G(W, W') = \text{tr} \left(A^T D A \right) - 2 \text{tr} \left(Y H W^T X^T D A \right) + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^T Q H \right) + \sum_{i=1}^m \sum_{j=1}^r \frac{\left(X^T D X W' H^T Y^T Y H \right)_{ij} W_{ij}^2}{W'_{ij}} \tag{34}$$

is an auxiliary function for J .

Proof Now $J(W)$ of Eq. 41 can be re-written as:

$$J(W) = \text{tr} \left(A^T D A \right) - 2 \text{tr} \left(Y H W^T X^T D A \right) + \lambda_1 \text{tr} \left(W^T P W \right) + \lambda_2 \text{tr} \left(H^T Q H \right) + \text{tr} \left(W^T X^T D X W H^T Y^T Y H \right) \tag{35}$$

Now we will be applying the following inequality of matrices according to the investigations by [14, 15]:

$$\text{tr} \left(W^T \Lambda W B \right) \leq \sum_i \sum_j \left(\Lambda W' B \right)_{ji} \frac{W_{ij}^2}{W'_{ij}}, \tag{36}$$

where, Λ, B, W are non-negative matrices, and Λ, B are symmetric matrices. And obviously the equality holds in Eq. 36 when $W = W'$.

In Eq. 36, if we do the substitutions: $\Lambda = X^T D X, B = H^T Y^T Y H, W = W, W' = W'$, we see that the fifth term of Eq. 35 is smaller than the fifth term of Eq. 34. However, the equality holds when $W = W'$. Thus $G(W, W')$ in Eq. 34 is an auxiliary function of $J(W)$. \square

Now, we need to find the global minimum of Eq. 34. Let $f(W) = G(W, W')$. The gradient of $f(W)$ is

$$\frac{\partial f(W)}{\partial W_{ij}} = -2 \left(X^T D A Y H \right)_{ij} + 2 \lambda_1 \left(P W \right)_{ij} + 2 \frac{\left(X^T D X W' H^T Y^T Y H \right)_{ij} W_{ij}}{W'_{ij}} \tag{37}$$

However, the second order derivative (i.e., the Hessian matrix) would be

$$\frac{\partial^2 f(W)}{\partial W_{ij} \partial W_{kl}} = 2(P)_{ij} \delta_{jl} \delta_{ik} + \left(2 \frac{\left(X^T D X W' H^T Y^T Y H \right)_{ij}}{W'_{ij}} \right) \delta_{jl} \delta_{ik} \tag{38}$$

The Hessian matrix (Eq. 38) is semi-positive definite implying that $f(W) = G(W, W')$ is a convex function. Thus, there exists a unique global minimum for $f(W)$. The global minimum can be obtained by setting the gradient of $f(W)$ to zero and solve for W . Thus from Eq. 37 we get

$$W_{ij} = W'_{ij} \frac{\left(X^T D A Y H \right)_{ij}}{\left(X^T D X W' H^T Y^T Y H + \lambda_1 P W \right)_{ij}} \tag{39}$$

By replacing $W^{(t+1)} = W$ and $W^{(t)} = W'$, we would obtain the W update rule in Algorithm 3. Therefore, under this rule, the objective function $J(W)$ of Eq. 41 decreases monotonically, and hence completes the proof.

Proof of Lemma 7

Proof We know that,

$$\begin{aligned} & \text{tr} \left(A - XW^{(t)}H^T Y^T \right)^T D \left(A - XW^{(t)}H^T Y^T \right) \\ & + \lambda_1 \text{tr} \left(W^{(t)T} P W^{(t)} \right) + \lambda_2 \text{tr} \left(H^T Q H \right) \\ & = \sum_{i=1}^M \sum_{j=1}^N \left(A - XW^{(t)}H^T Y^T \right)_{ij} D_{ii} \\ & + \lambda_1 \sum_{k=1}^m \sum_{l=1}^r W_{kl}^{(t)2} P_{kk} + \lambda_2 \text{tr} \left(H^T Q H \right) \\ & = \sum_{i=1}^M \left\| A_i - \left(XW^{(t)}H^T Y^T \right)_i \right\|^2 D_{ii} \\ & + \lambda_1 \sum_{k=1}^m \left\| W_k^{(t)} \right\|^2 P_{kk} + \lambda_2 \text{tr} \left(H^T Q H \right) \end{aligned}$$

Similarly, we can see that

$$\begin{aligned} & \text{tr} \left(A - XW^{(t+1)}H^T Y^T \right)^T D \left(A - XW^{(t+1)}H^T Y^T \right) \\ & + \lambda_1 \text{tr} \left(W^{(t+1)T} P W^{(t+1)} \right) + \lambda_2 \text{tr} \left(H^T Q H \right) \\ & = \sum_{i=1}^M \left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\|^2 D_{ii} \\ & + \lambda_1 \sum_{k=1}^m \left\| W_k^{(t+1)} \right\|^2 P_{kk} + \lambda_2 \text{tr} \left(H^T Q H \right) \end{aligned}$$

Then, the right-hand side (*r.h.s*) of Eq. 23 becomes

$$\begin{aligned} r.h.s &= \frac{1}{2} \sum_{i=1}^M \left(\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\|^2 \right. \\ & - \left. \left\| A_i - \left(XW^{(t)}H^T Y^T \right)_i \right\|^2 \right) D_{ii} + \lambda_1 \sum_{k=1}^m \left(\left\| W_k^{(t+1)} \right\|^2 \right. \\ & - \left. \left\| W_k^{(t)} \right\|^2 \right) P_{kk} \\ & = \frac{1}{2} \sum_{i=1}^M \left(\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\|^2 D_{ii} - \frac{1}{D_{ii}} \right) \\ & + \lambda_1 \sum_{k=1}^m \left(\left\| W_k^{(t+1)} \right\|^2 P_{kk} - \frac{1}{P_{kk}} \right) \end{aligned}$$

And, the left-hand side (*l.h.s*) of Eq. 23 becomes

$$\begin{aligned} l.h.s &= \sum_{i=1}^M \left(\sqrt{\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\|^2} \right. \\ & - \left. \sqrt{\left\| A_i - \left(XW^{(t)}H^T Y^T \right)_i \right\|^2} \right) \\ & + \lambda_1 \sum_{k=1}^m \left(\sqrt{\left\| W_k^{(t+1)} \right\|^2} - \sqrt{\left\| W_k^{(t)} \right\|^2} \right) \\ & = \sum_{i=1}^M \left(\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\| \right. \\ & - \left. \left\| A_i - \left(XW^{(t)}H^T Y^T \right)_i \right\| \right) \\ & + \lambda_1 \sum_{k=1}^m \left(\sqrt{\left\| W_k^{(t+1)} \right\|^2} - \sqrt{\left\| W_k^{(t)} \right\|^2} \right) \\ & = \sum_{i=1}^M \left(\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\| - \frac{1}{D_{ii}} \right) \\ & + \lambda_1 \left(\sum_{k=1}^m \left\| W_k^{(t+1)} \right\| - \frac{1}{P_{kk}} \right) \end{aligned}$$

Now, we compute the difference between the *l.h.s* and *r.h.s*,

$$\begin{aligned} l.h.s - r.h.s &= \sum_{i=1}^M \left(\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\| \right. \\ & - \left. \left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\|^2 \frac{D_{ii}}{2} - \frac{1}{2D_{ii}} \right) \\ & + \lambda_1 \sum_{k=1}^m \left(\left\| W_k^{(t+1)} \right\| - \left\| W_k^{(t+1)} \right\|^2 \frac{P_{kk}}{2} - \frac{1}{2P_{kk}} \right) \\ & = \sum_{i=1}^M \frac{D_{ii}}{2} \left(\frac{\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\|}{D_{ii}} \right. \\ & - \left. \left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\|^2 - \frac{1}{D_{ii}^2} \right) \\ & + \lambda_1 \sum_{k=1}^m \frac{P_{kk}}{2} \left(\frac{\left\| W_k^{(t+1)} \right\|}{P_{kk}} - \left\| W_k^{(t+1)} \right\|^2 - \frac{1}{P_{kk}^2} \right) \\ & = \sum_{i=1}^M \frac{(-D_{ii})}{2} \left(\left\| A_i - \left(XW^{(t+1)}H^T Y^T \right)_i \right\| - \frac{1}{D_{ii}} \right)^2 \\ & + \lambda_1 \sum_{k=1}^m \frac{(-P_{kk})}{2} \left(\left\| W_k^{(t+1)} \right\| - \frac{1}{P_{kk}} \right)^2 \\ & \leq 0 \end{aligned}$$

The above inequality holds because, *D, P* are non-negative matrices, and the sum of non-positive numbers is always non-positive. This completes the proof. \square

Correctness of the RIMC Algorithm (version 2)

In this section we are going to prove that the converged solution presented in Algorithm 3 is the correct optimal solution. In fact, we will show that the converged solution satisfies the Karush-Kuhn-Tucker (KKT) condition of the constrained optimization theory. At first, we have Theorem 10 to prove the correctness of the algorithm with respect to W . Theorem 11 will prove the correctness of the algorithm with respect to H .

Theorem 10 *At convergence, the converged solution W^* of the updating rule in Algorithm 3 satisfies the KKT condition.*

Proof The KKT condition for W with constraints $W_{\alpha\beta} \geq 0$, with $\alpha = 1, \dots, m; \beta = 1, \dots, r$ is:

$$\frac{\partial J(W)}{\partial W_{\alpha\beta}} W_{\alpha\beta} = 0, \forall \alpha, \beta \tag{40}$$

Similar to Eq. 25, the $J(W)$ can be written as:

$$J(W) = \text{tr} \left(A - XWH^T Y^T \right)^T D \left(A - XWH^T Y^T \right) + \lambda_1 \text{tr} \left(W^T PW \right) + \lambda_2 \text{tr} \left(H^T QH \right) \tag{41}$$

Now, the partial derivative of $J(W)$ can be expressed as:

$$\frac{\partial J(W)}{\partial W_{\alpha\beta}} = -2 \left(X^T DAYH \right)_{\alpha\beta} + 2\lambda_1 (PW)_{\alpha\beta} + 2 \left(X^T DXWH^T Y^T YH \right)_{\alpha\beta} \tag{42}$$

Thus, the KKT condition for W is:

$$\left[- \left(X^T DAYH \right)_{\alpha\beta} + \lambda_1 (PW)_{\alpha\beta} + \left(X^T DXWH^T Y^T YH \right)_{\alpha\beta} \right] W_{\alpha\beta} = 0, \forall \alpha, \beta \tag{43}$$

But, once W converges (according to Algorithm 3), the converged solution W^* satisfies the following:

$$W_{\alpha\beta}^* \leftarrow W_{\alpha\beta}^* \frac{\left(X^T DAYH \right)_{\alpha\beta}}{\left(X^T DXW^* H^T Y^T YH + \lambda_1 PW^* \right)_{\alpha\beta}}$$

which can be written as

$$\left[- \left(X^T DAYH \right)_{\alpha\beta} + \lambda_1 (PW^*)_{\alpha\beta} + \left(X^T DXW^* H^T Y^T YH \right)_{\alpha\beta} \right] W_{\alpha\beta}^* = 0, \forall \alpha, \beta \tag{44}$$

This is identical to Eq. 43. Thus, the converged solution W^* satisfies the KKT condition. \square

Theorem 11 *At convergence, the converged solution H^* of the updating rule in Algorithm 3 satisfies the KKT condition.*

Proof The KKT condition for H with constraints $H_{\gamma\psi} \geq 0$, with $\gamma = 1, \dots, n, \psi = 1, \dots, r$ is:

$$\frac{\partial J(H)}{\partial H_{\gamma\psi}} H_{\gamma\psi} = 0, \forall \gamma, \psi \tag{45}$$

Now, the partial derivative of $J(H)$ from Eq. 25 is

$$\frac{\partial J(H)}{\partial H_{\gamma\psi}} = -2 \left(Y^T A^T DXW \right)_{\gamma\psi} + 2\lambda_2 (QH)_{\gamma\psi} + 2 \left(Y^T YHW^T X^T DXW \right)_{\gamma\psi} \tag{46}$$

Thus, the KKT condition for H is:

$$\left[- \left(Y^T A^T DXW \right)_{\gamma\psi} + \lambda_2 (QH)_{\gamma\psi} + \left(Y^T YHW^T X^T DXW \right)_{\gamma\psi} \right] H_{\gamma\psi} = 0, \forall \gamma, \psi \tag{47}$$

But, once H converges (according to Algorithm 3), the converged solution, H^* satisfies the following:

$$H_{\gamma\psi}^* \leftarrow H_{\gamma\psi}^* \frac{\left(Y^T A^T DXW \right)_{\gamma\psi}}{\left(Y^T YH^* W^T X^T DXW + \lambda_2 QH^* \right)_{\gamma\psi}}$$

which can be written as

$$\left[- \left(Y^T A^T DXW \right)_{\gamma\psi} + \lambda_2 (QH^*)_{\gamma\psi} + \left(Y^T YH^* W^T X^T DXW \right)_{\gamma\psi} \right] H_{\gamma\psi}^* = 0, \forall \gamma, \psi$$

This is identical to Eq. 47. Thus, the converged solution H^* satisfies the KKT condition. \square

Stable robust IMC (SRIMC) formulation

Instead of solving the RIMC objective function (Eq. 7) directly, here we propose a two-step solution strategy to the RIMC formulation, and we call this new algorithm SRIMC.

Step 1: solving matrix Z from a matrix equation

In this step, we consider the following matrix equation

$$XZY^T = A, \tag{48}$$

where Z is an $m \times n$ matrix of unknowns, X is the $M \times m$ feature matrix of the row entities, Y is the $N \times n$ is the feature matrix of the column entities. And, A is the $M \times N$ binary association matrix between the row and column entities.

Now, in Eq. 48, if we left multiply by X^T and right multiply by Y , we get the following equation

$$X^T XZY^T Y = X^T AY \tag{49}$$

If X has full column rank and Y has a full row rank, then both $X^T X$ and $Y^T Y$ are invertible. Therefore, we can solve for Z .

$$\begin{aligned} ZY^T Y &= (X^T X)^{-1} X^T A Y \\ \Rightarrow \hat{Z} &= (X^T X)^{-1} X^T A Y (Y^T Y)^{-1} \end{aligned} \quad (50)$$

Step 2: robust NMF on matrix Z

$$\begin{aligned} \min_{W,H} \quad & \varphi = \|Z - WH^T\|_{2,1} + \lambda_1 \|W\|_{2,1} + \lambda_2 \|H\|_{2,1} \\ \text{such that,} \quad & W \geq 0, H \geq 0 \end{aligned} \quad (51)$$

This is a modified non-negative matrix factorization (NMF) problem; only difference is the usage of the $\ell_{2,1}$ norms instead of ℓ_2 norms in the loss function and the regularizers.

Algorithm for SRIMC

We can also solve the Stable Robust IMC optimization problem by solving the two problems mentioned above. It is demonstrated in Algorithm 4.

Algorithm 4: Solve Stable Robust Inductive Matrix Completion (SRIMC)

Data: $A \in \mathbb{R}^{M \times N}$, $X \in \mathbb{R}^{M \times m}$, $Y \in \mathbb{R}^{N \times n}$

Result: $W \in \mathbb{R}^{m \times r}$ and $H \in \mathbb{R}^{n \times r}$

Initialize W and H with random number maintaining the non-negativity constraints $W_{ik} \geq 0, H_{jk} \geq 0$;

Initialize $D \in \mathbb{R}^{M \times M}$, $P \in \mathbb{R}^{m \times m}$, $Q \in \mathbb{R}^{n \times n}$ as identity matrices.;

Solve Z from the equation $XZY^T = A$;

while convergence criteria not met **do**

$$W_{ik} = W_{ik} \frac{(ZDH)_{ik}}{(WH^T DH + \lambda_1 PW)_{ik}},$$

$$H_{jk} = H_{jk} \frac{(DZ^T W)_{jk}}{(DHW^T W + \lambda_2 QH)_{jk}},$$

$$D_{ii} = \frac{1}{\sum_{j=1}^n (Z - WH^T)_{ij}^2},$$

$$P_{ii} = \frac{1}{\sum_{j=1}^r W_{ij}^2},$$

$$Q_{ii} = \frac{1}{\sum_{j=1}^r H_{ij}^2}$$

end

return W, H

Results

Disease-LincRNA association datasets

We prepared a sparse association matrix by extracting the lincRNA-disease association dataset from the LincRNA-Disease [4] with sparsity index 0.22%. LincRNA expression dataset was obtained from the co-expression based association study [7]. Finally, we cataloged 8194 lincRNAs and 2148 human disease phenotypes and the resulting association matrix contains 46,934 associations among these two entities. We followed a standard naming of the disease phenotypes by OMIM identification numbers. We extracted top-5 OMIM phenotypes matching the human disease names using OMIM API [16].

LincRNA feature datasets

The features of LincRNAs consist of four groups of information: (i) expression profiles, (ii) transcription factor binding sites (TFBS), (iii) functional annotations and (iv) single nucleotide polymorphism (SNP) information. The RNA-seq expression profiles of the 8194 lincRNAs on 22 human tissues were collected from the Human BodyMap Project 2.0 [3]. The expression scores were measured in FPKM (Fragments Per Kilobase of exons per Million Fragments mapped) unit. Then, TFBS information about the lincRNAs in our study with 120 transcription factors were obtained from ChIP-base dataset [17]. Linc2GO is a public data repository containing functional annotations of lincRNAs [18]. There are three different types of functions cataloged in the Linc2GO dataset: gene ontology biological process (GO BP), gene ontology molecular function (GO MF) and KEGG pathways. The 8194 lincRNAs with the functional annotation together make a sparse matrix with sparsity index 0.11%. We performed singular value decomposition on the matrix to compute and use the leading 100 singular vectors in our study as part of the features of the lincRNAs. We extracted links among 368,494 SNPs and the lincRNAs from our study from the LincRNASNP dataset [19]. Again, the SNP-lincRNA association matrix turned out to a sparse matrix with the sparsity index 0.0077%. Therefore, we performed singular value decomposition on the matrix to compute and use the leading 100 singular vectors. Finally, we performed a filtering on all the four groups of features of the lincRNAs in our study. We found that 6540 out of the initial 8194 lincRNAs have data from all the four groups of featureset. Therefore, our final lincRNA feature matrix (X in our study) has 6540 rows (lincRNAs) and 342 columns (features).

Disease feature datasets

The disease feature dataset consists of two groups of information: (i) term frequency inverse document frequency (TF-IDF) scores and (ii) phenotype similarity scores. The TFIDF scores were prepared by mining the OMIM text corpus on the 2661 OMIM phenotypes, resulting a 20491

term scores of each of the 2148 phenotypes from our study. We took leading 100 singular vectors as part of the disease feature. The phenotype-phenotype similarity scores were retrieved from a study conducted by [20]. The similarity profiles after encapsulated in a square matrix of dimension 2148 by 2148, had to go through a singular value decomposition module to extract leading 100 singular vectors that constitute the part of the feature matrix of the diseases in our study. Finally, our disease feature matrix contains 200 features of the 2148 diseases.

Baseline algorithms

We conducted a comparative study of our proposed algorithms with five baseline methods: (i) NMF [13], (ii) LRLSLDA [5], (iii) TsLincRNA-Disease [7], (iv) K-RWRH [6] and (v) standard IMC [21]. The NMF based approach finds the two factors W and H by just working on the lincRNA-disease association matrix A . The LRLSLDA ranks the lincRNAs with a disease by the use of a classifier trained on two similarity feature matrices. The method was developed with eight parameters to train before getting good prediction results. The TsLincRNA-Disease utilizes a series of statistical significance tests on a co-expression network obtained from tissue-specific and non-tissue-specific lincRNA expression information. Apart from the expression data, this method lacks the integration of other types of information available about the lincRNAs and the disease. The K-RWRH is a stochastic algorithm developed on top of the random walk on a three heterogeneous networks. The method is very complex and it is harder to obtain a steady state distribution for the dataset our study.

Evaluation metrics

We define two metrics for evaluating our proposed algorithm and the baseline algorithms. The metrics are popular in evaluating any recommender style systems as in [22].

precision@k: The ratio of the number of recovered disease phenotypes to recommended k phenotypes for a target lincRNA. We take average of the ratios for every lincRNAs of our study. The metric is defined as follows:

$$precision@k = \frac{1}{N_l} \sum_{l=1}^{N_l} \frac{|P_l(k) \cap D_l|}{k}, \tag{52}$$

where, $P_l(k)$ is the top- k ranked diseases for an lincRNA l , D_l is the set of diseases related to the lincRNA l deleted during the training phase. And, N_l is the total number lincRNAs in the test set.

recall@k: The ratio of recovered disease phenotypes to the set of hidden phenotypes in the test dataset. Again, we take average of the ratios for every lincRNAs in the study. The metric is defined as follows:

$$recall@k = \frac{1}{N_l} \sum_{l=1}^{N_l} \frac{|P_l(k) \cap D_l|}{|D_l|}, \tag{53}$$

We repeated the experiments for various values of k , from 5 to 100. We conducted 10-fold cross-validation in each of the experiments listed in the following sections.

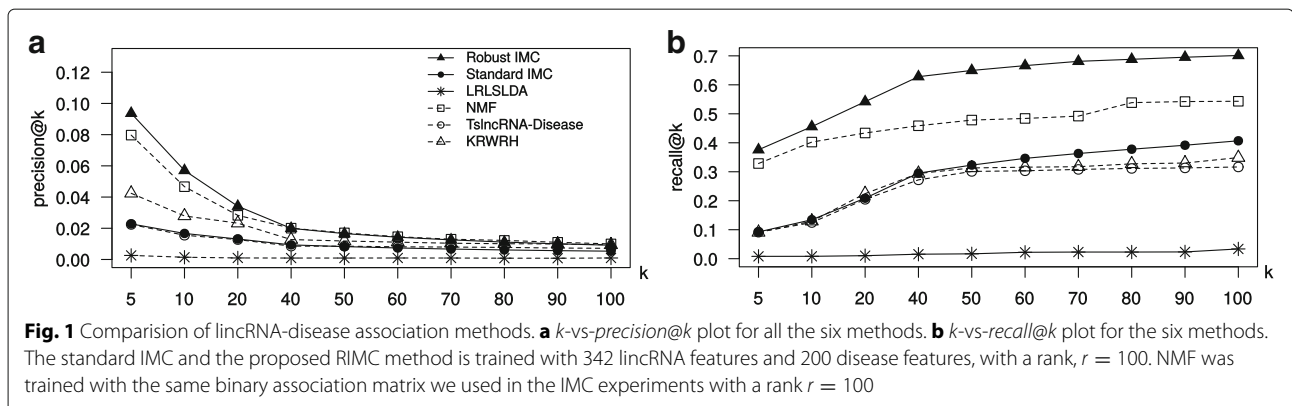
Discussions

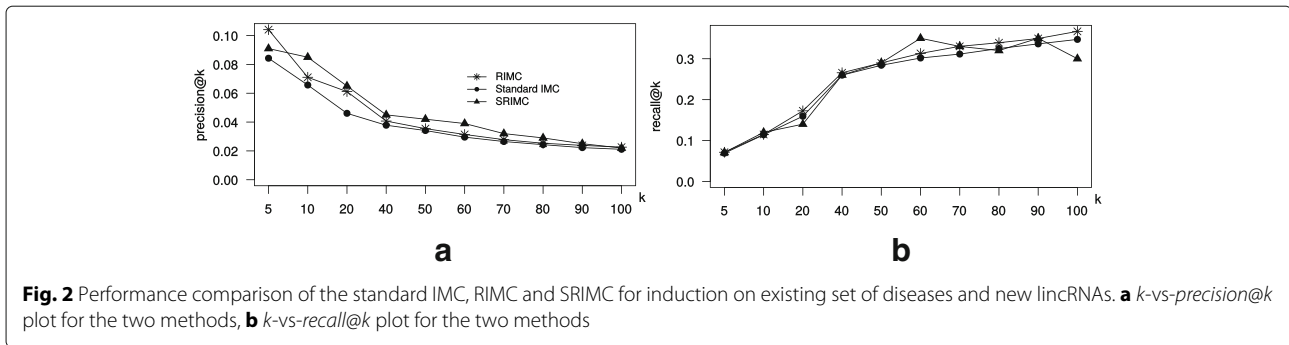
True LincRNA-disease association retrieval

Figure 1 shows the performance of RIMC along with other base-line algorithms to predict true lincRNA-disease associations. A 10-fold cross-validation was conducted on the 2418 OMIM phenotypes. We find that our RIMC method leads in identifying true associations than all the baseline algorithms for all k values. The NMF based algorithm is better than the three other baseline algorithms. LRLSLDA’s association retrieval was the worse due to the fact that it relies only on known association matrix and the expression profiles of the lincRNAs that seems to be not sufficient to build one predictive model.

Induction on new associations

Here we conducted a thorough comparative study on the three algorithms including two of ours (RIMC and





SRIMC) to predict associations between novel lincRNAs and/or diseases. We assume that all the features of the novel lincRNAs and/or diseases that we bring into our prediction framework can be computed or available. Note that, none of the baseline algorithms except the standard inductive matrix completion based approach (standard IMC) are missing in all the experiments from this sections due to the fact that none are capable of doing induction on novel associations.

Induction experiments on new LincRNAs

From the dataset in our study we selected a list of 10% lincRNAs and deleted all the entries of these randomly selected lincRNAs from the three training matrices A , X and Y . The deleted entries will serve as test set during evaluation. Then, RIMC, SRIMC and the standard IMC were trained with modified training matrices. Once, training is done on the reduced dataset, each of the obtained three modules were evaluated with the test set that were extracted at the beginning of this step. We repeat the entire training and test steps 10 times and reported the average performance score of all the three methods. Figure 2 illustrates the performance comparison of the three methods for predicting association between a new lincRNA with an existing set of diseases. RIMC and SRIMC show better $precision@k$ than the standard IMC based approach for predicting upto the top-50 disease associations with the new lincRNAs. For higher values

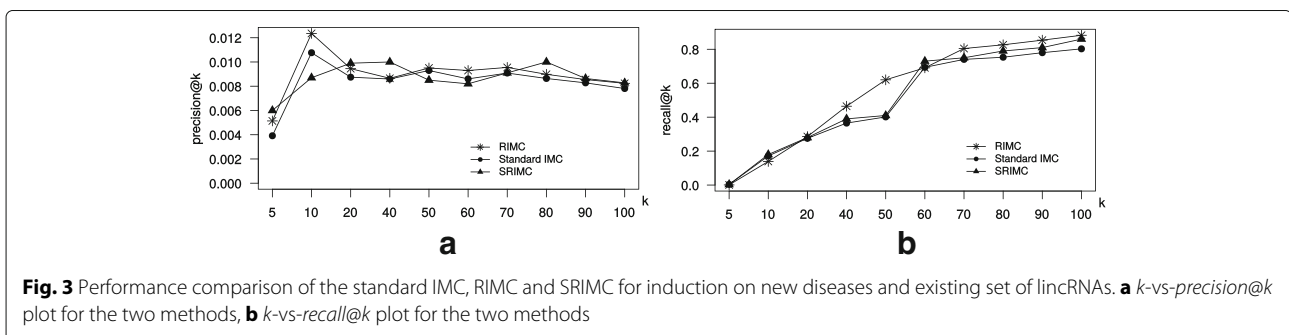
of k in the top- k predictions, both RIMC and the standard IMC show similar performance. But in terms of numerical precision, RIMC exceeds the performance of standard IMC. However, in terms of $recall@k$, we can see that SRIMC and RIMC perform superior than that of the standard IMC method.

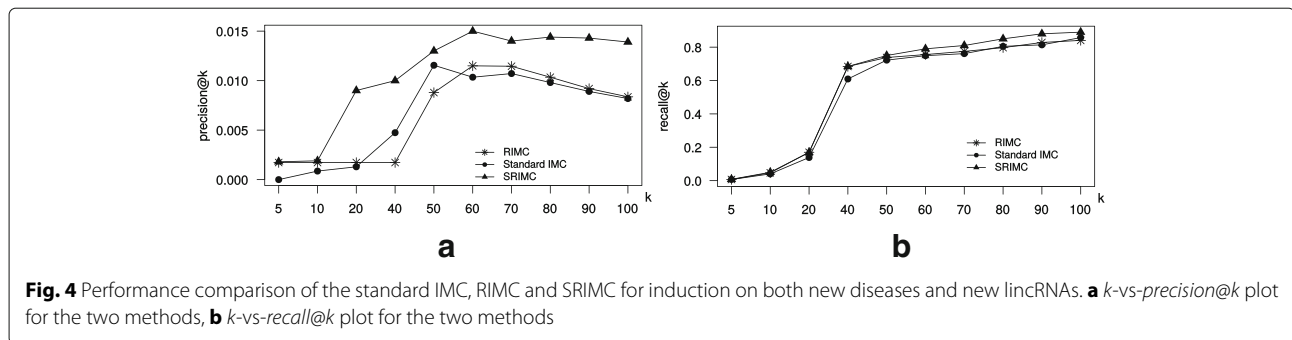
Induction experiments on new diseases

Similar to the approach mentioned in the previous section, we randomly selected 10% of the total disease phenotypes from the dataset of the study, and deleted all the entries related to the diseases. The deleted entries is going to be our test set. The reduced dataset is going to serve as training dataset. The RIMC, SRIMC and the standard IMC were trained on the reduced training dataset and evaluated against the test set. The entire training and evaluation were repeated 10 times and the average performance scores were reported. Figure 3 illustrates the performance comparison of the three methods to predict associations among known list of lincRNAs with a novel disease. Here, both RIMC and SRIMC demonstrates better induction performance in terms of the $precision@k$ and $recall@k$ values.

Induction experiments on both new LincRNAs and new diseases

Finally, in this batch of induction experiment, we randomly picked 5% of the subject disease entries, and 5%





of the subject lincRNA entries and deleted the respective connections between the two entities from the three data matrices A , X and Y . The deleted connections and feature set are treated as the test-set, while the reduced data matrices are used to train the three algorithms. We repeat the above steps 10 times and compute the average performance scores. Figure 4 illustrates the performance comparison of our proposed RIMC, SRIMC and the only baseline algorithm applicable here which is the standard IMC to predict association between a new lincRNA and a new disease based on the model trained on data about a limited set of lincRNAs and disease phenotypes not including these two lincRNA and disease phenotypes. The $precision@k$ plot of for the RIMC and SRIMC show better performance than the standard IMC based approach for predicting for both lower and higher values of k in the top- k association ranking with the novel diseases. However, from the $recall@k$ curve of the both algorithms, we can see that both RIMC and standard IMC performs similar in the top- k association prediction problem. But, SRIMC performs superior than both of the algorithms.

Conclusions

In this article, we propose theoretical foundations of robust inductive matrix completion method using $\ell_{2,1}$ norm. We provided three algorithms to solve our robust induction matrix completion objective function. The first two algorithms are equivalent, but the third one what we call Stable Robust Inductive Matrix Completion (SRIMC) breaks the problem into two sub-problems. But it turns out to be a simple, stable and better solution strategy. We applied the proposed methods in identifying missing links between putative lincRNAs and human disease phenotypes. All the three variants of robust inductive matrix completion are well suited for noisy type of datasets. Besides the standard IMC formulation, our proposed method also outperformed other four lincRNA-disease association solutions. The proposed methods are applicable to predict associations among between well-studied

lincRNAs with novel disease, or novel lincRNAs with well-studied diseases, or a set of novel lincRNAs with novel diseases.

Abbreviations

FPKM: Fragments per kilobase of exons per million fragments; GWAS: Genome wide association studies; HTS: High throughput sequencing; IMC: Inductive matrix completion; lincRNAs: Long intergenic non-coding RNAs; lncRNAs: Long non-coding RNAs; NMF: Non-negative matrix factorization; RIMC: Robust inductive matrix completion; SRIMC: Stable robust inductive matrix completion; SNP: Single nucleotide polymorphism; TFBS: Transcriptor factor binding sites; TF-IDF: Term frequency inverse document frequency

Acknowledgements

We express deepest gratitude to Dr. Kytai Nguyen in Bioengineering Department at University of Texas at Arlington for the comments and feedback on our lincRNAs discovery results.

Funding

Not applicable.

Availability of data and materials

The ChIP-base dataset is available at <https://omictools.com/chipbase-tool>. The Linc2GO dataset is available at: <https://omictools.com/linc2go-tool>. The SNP-lincRNA data can be found at: <http://bioinfo.life.hust.edu.cn/lncRNASNP/>.

About this supplement

This article has been published as part of *BMC Medical Genomics* Volume 10 Supplement 5, 2017: Selected articles from the IEEE BIBM International Conference on Bioinformatics & Biomedicine (BIBM) 2016: medical genomics. The full contents of the supplement are available online at <https://bmcmmedgenomics.biomedcentral.com/articles/supplements/volume-10-supplement-5>.

Authors' contributions

AKB conceived the package and wrote the manuscript. DK and MK contributed to data analysis and programming for the experiment. CD contributed to the two-step algorithm to solve the proposed RIMC algorithm. JG provided overall supervision. All authors reviewed, edited and approved the final manuscript.

Ethics approval and consent to participate

All the datasets used in this study are from publicly available data repository. No patients samples were used or collected in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Computer Science and Engineering, University of Colorado Denver, Denver, Colorado 80204, USA. ²Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, Texas 76019, USA. ³Department of Computer Science, University of Rio Grande Valley, Edinburg, Texas 78541, USA. ⁴Department of Computer Science, Kennesaw State University, Marietta, Georgia 30060, USA.

Published: 28 December 2017

References

- Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. *Nat Rev Genet.* 2010;11(8):559–71.
- Esteller M. Non-coding rnas in human disease. *Nat Rev Genet.* 2011;12(12):861–74.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding rnas reveals global properties and specific subclasses. *Genes Dev.* 2011;25(18):1915–27.
- Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* 2013;41(D1):983–6. doi:10.1093/nar/gks1099.
- Chen X, Yan GY. Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics.* 2013;29(20):2617–24.
- Ganegoda GU, Li M, Wang W, Feng Q. Heterogeneous network model to infer human disease-long intergenic non-coding rna associations. *NanoBioscience IEEE Trans.* 2015;14(2):175–83.
- Liu MX, Chen X, Chen G, Cui QH, Yan GY. A computational framework to infer human disease-associated long noncoding rnas. *PLoS ONE.* 2014;9(1):84408.
- Jain P, Dhillon IS. Provable inductive matrix completion. arXiv preprint arXiv:1306.0626. 2013. <https://arxiv.org/abs/1306.0626>.
- Liu W, Zheng N, You Q. Nonnegative matrix factorization and its applications in pattern recognition. *Chin Sci Bull.* 2006;51(1):7–18.
- Biswas AK, Kim DC, Kang M, Gao JX. Robust inductive matrix completion strategy to explore associations between lincrnas and human disease phenotypes. In: *Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference On. Shenzhen: IEEE; 2016.* p. 334–9.
- Luo D, Ding C, Huang H. Towards structural sparsity: an explicit l2/l0 approach. In: *2010 IEEE International Conference on Data Mining. Sydney; 2010.* p. 344–53. doi:10.1109/ICDM.2010.155.
- Nie F, Huang H, Cai X, Ding CH. Efficient and robust feature selection via joint $\ell_2,1$ -norms minimization. In: *Advances in Neural Information Processing Systems. Vancouver; 2010.* p. 1813–21.
- Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature.* 1999;401(6755):788–91.
- Kong D, Ding C, Huang H. Robust nonnegative matrix factorization using l2,1-norm. In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management. Glasgow: ACM; 2011.* p. 673–82.
- Ding CH, Li T, Jordan MI. Convex and semi-nonnegative matrix factorizations. *IEEE Trans Pattern Anal Mach Intell.* 2010;32(1):45–55.
- Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 2015;43(D1):789–98.
- Yang JH, Li JH, Jiang S, Zhou H, Qu LH. CHIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res.* 2013;41(D1):177–87.
- Liu K, Yan Z, Li Y, Sun Z. Linc2GO: a human lincRNA function annotation resource based on ceRNA hypothesis. *Bioinformatics.* 2013;29(17):2221–2.
- Gong J, Liu W, Zhang J, Miao X, Guo AY. lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res.* 2015;43(D1):181–6.
- Caniza H, Romero AE, Paccanaro A. A network medicine approach to quantify distance between hereditary disease modules on the interactome. *Sci Rep.* 2015;5:17658. doi:10.1038/srep17658.
- Natarajan N, Dhillon IS. Inductive matrix completion for predicting gene–disease associations. *Bioinformatics.* 2014;30(12):60–8.
- Lian D, Zhao C, Xie X, Sun G, Chen E, Rui Y. Geomf: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York City: ACM; 2014.* p. 831–40.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

