Learning a Low-rank Tensor of Pharmacogenomic Multi-relations from Biomedical Networks

Zhuliu Li∗, Wei Zhang†, R Stephanie Huang∗ and Rui Kuang∗

∗University of Minnesota Twin Cities, MN, USA
†University of Central Florida, FL, USA

∗{lixx3617, rshuang, kuang}@umn.edu, †wzhang.cs@ucf.edu

Abstract—Learning pharmacogenomic multi-relations among diseases, genes and chemicals from content-rich biomedical and biological networks can provide important guidance for drug discovery, drug repositioning and disease treatment. Most of the existing methods focus on imputing missing values in the disease-gene, disease-chemical and gene-chemical pairwise relations from the observed relations instead of being designed for learning high-order disease-gene-chemical multi-relations. To achieve the goal, we propose a general tensor-based optimization framework and a scalable Graph-Regularized Tensor Completion from Observed Pairwise Relations (GT-COPR) algorithm to infer the multi-relations among the entities across multiple networks in a low-rank tensor, based on manifold regularization with the graph Laplacian of a Cartesian, tensor or strong product of the networks, and consistencies between the collapsed tensors and the observed bipartite relations. Our theoretical analyses also prove the convergence and efficiency of GT-COPR. In the experiments, the tensor fiber-wise and slice-wise evaluations demonstrate the accuracy of GT-COPR for predicting the disease-gene-chemical associations across the large-scale protein-protein interactions network, chemical structural similarity network and phenotype-based human disease network; and the validation on Genomics of Drug Sensitivity in Cancer cell line dataset shows a potential clinical application of GT-COPR for learning disease-specific chemical-gene interactions. Statistical enrichment analysis demonstrates that GT-COPR is also capable of producing both topologically and biologically relevant disease, gene and chemical components with high significance.

Source code: https://github.com/kuanglab/GT-COPR

Index Terms—multi-relational learning, drug repositioning, disease gene prioritization, product graphs, tensor completion.

I. INTRODUCTION

The relationships among different entities across multiple biomedical and biological networks carry rich information of the associations between heterogeneous objects of diseases, genes (proteins) and chemicals in the networks. Inferring the disease-gene-chemical multi-relations based on the network topologies can guide the development and application of therapies for precision medicine [1]. An example of real known pharmacogenomic multi-relations across three biomedical subnetworks are shown in Fig. 1, e.g. the multi-relation "acute myeloid leukemia"-"NRAS"-"fedratinib" indicates mutation of NRAS (gene) can impact the sensitivity of fedratinib (chemical) in treating acute myeloid leukemia (disease) [1].

Current studies mainly emphasize on imputing unknown pairwise relations among biomedical and biological networks, based on the observed associations and network topological information. For example, [4] and [5] predict novel drug-target associations for drug repositioning using observed associations obtained from publicly available databases, together with the drug structural and target sequence similarity networks; [6] and [7] adopt network propagation based approaches according to the smoothness assumptions made on the protein-protein interactions (PPI) and phenotype similarity networks to identify disease-gene pairs for prioritizing disease causal genes. To infer the disease-gene, disease-chemical and gene-chemical pairwise associations simultaneously, [8] formulates a collective matrix completion problem with graph regularization, and [9] applies network diffusion algorithms on the six-modal network constructed by stacking all the pairwise relational matrices and networks from the three data domains.

To learn the disease-gene-chemical multi-relations directly, we propose a general product-graph-regularized tensor (n-way array) completion framework. Our goal is to predict the unknown n-way relations based on the topological information carried in variant types of product graphs. In the last decade, graph-based tensor completion techniques have received great interest. In [10]–[12], semi-supervised manifold learning technology referred as label propagation [13], [14] is applied for completing the partially observed tensors. Specifically, [10] predicts the link types of the unknown hyperlinks among knowledge graphs in a tensor based on the conjugate gradient descent optimization, whose scalability is later improved in...
[11] through low-rank approximation of the knowledge graphs; [12] introduces the product graph regularization into the tensor completion objective via a Gaussian random fields prior to infer the cross-graph multi-relations. As the graph-regularized matrix factorization [15]–[17] has been extensively explored and achieved substantial success in data mining areas from collaborative filtering to link prediction, [18] generalized the idea to tensor completion via decomposing an incomplete tensor into low-rank matrices with their values jointly smoothing over the manifolds of the tensor product graph. When dealing with the temporal-bipartite-relational tensor completion problem, [19] proposes to collapse the sub-tensor of previous time stamps into a matrix (bipartite graph), followed by applying the well-known Katz measure for pairwise link prediction at a future time stamp. Similar tensor collapsing idea has also been adopted in [20] for nodes classification on the temporal bipartite graphs.

As reviewed above, the existing tensor completion methods all require training with observed multi-relations, which are however very scarce or even unavailable in many situations, especially in high-order multi-relations. The requirement severely limits the applicability of these methods to predicting disease-gene-chemical associations, in which pairwise relations among disease-gene, disease-chemical and gene-chemical widely exist in public available databases such as CTD [21], DrugBank [22] and ChEMBL [23], while the curated triple-wise relations are extremely sparse. Therefore, we consider utilizing the observed pairwise relations, together with the knowledge graphs to solve the multi-relational learning problem. In this work, we formally establish a novel and general tensor-based optimization objective and a scalable iterative method GT-COPR (Graph-Regularized Tensor Completion from Observed Pairwise Relations) to efficiently infer the n-way relations in a compressed tensor. The key novelties of our model are:

- We propose to learn a compressed tensor in CPD-form to guarantee the space and time efficiencies for learning high-order multi-relations.
- We propose to co-regularize tensor elements with the Laplacian of three types of product graphs (Cartesian, tensor and strong product) to introduce the local consistency among the n-way relations in multiple entities.
- We propose to apply tensor collapsing to capture the cross-mode dependencies and the global consistencies with the observed bipartite relations exist in database.

II. PRELIMINARIES

In this section, we summarize the notations and definitions used in the forthcoming sections. Other general knowledge of tensor computation can be found in the survey paper [24].

1) Notations

Vector: \( x \) Hadamard product: \( \circ \)
Tensor: \( X \) Khatri-Rao product: \( \odot \)
Vectorization of tensor: \( \text{vec}(X) \) Vector outer product: \( \odot \)
The \((a,b)\)-th element of matrix \( X \): \( X_{ab} \) Kronecker sum: \( \oplus \)
The \((a,b,c)\)-th element of tensor \( X \): \( X_{abc} \) Kronecker product: \( \otimes \)

2) CANDECOMP/PARAFAC decomposition (CPD)

An \( n \)-way tensor \( T \in \mathbb{R}^{I_n \times I_{n-1} \times \cdots \times I_1} \) of rank \( K \) can be written as

\[
T = \sum_{k=1}^{K} A_k^{(1)} \circ A_k^{(n-1)} \circ \cdots \circ A_k^{(1)} = [A^{(n)}, A^{(n-1)}, \ldots, A^{(1)}],
\]

where \( A_k^{(i)} \) is the \( k \)-th column of factor matrix \( A^{(i)} \in \mathbb{R}^{I_i \times K} \).

Vectorization property: The CPD-form tensor described above can be vectorized as \( \text{vec}(T) = (A^{(1)} \odot A^{(2)} \odot \cdots \odot A^{(n)})1 \), where \( 1 \) is a vector with all-ones.

III. TENSOR-BASED MULTI-RELATIONAL LEARNING

As illustrated in Fig. 2, our learning task is to infer the multi-relations in an \( n \)-way tensor \( T \in \mathbb{R}^{I_n \times I_{n-1} \times \cdots \times I_1} \) across the nodes of \( n \) knowledge graphs \( \{G^{(i)} = (V^{(i)}, E^{(i)}) : |V^{(i)}| = I_i, \forall i = 1, \ldots, n \} \) where \( V^{(i)} \) and \( E^{(i)} \) are the nodes and edges of \( G^{(i)} \) (Fig. 2 (III)), given a set of non-negative matrices \( \{R_{i,j} \in \mathbb{R}_{+}^{I_i \times I_j} : \forall i, j \in [1, n] \) and \( i < j \) (Fig. 2 (I)) with \( R_{i,j} \) holding the observed pairwise relations between the nodes of graphs \( G^{(i)} \) and \( G^{(j)} \) with zeros representing the unknown relations. To solve the multi-relational learning problem, we first propose our optimization formulation, and then present an efficient iterative algorithm GT-COPR to minimize the optimization objective, followed by theoretical analyses of the convergence and efficiency of GT-COPR.

A. Optimization Formulation

Our key ideas of solving the \( n \)-way relational learning problem are 1) \( n \)-way relations inferred in the tensor \( T \) are required to be consistent with each other by the connectivity in the product graph which will be defined in Section III-B; 2) the collapsed tensors are required to be consistent with the corresponding bipartite relational matrices; and 3) the inferred tensor \( T \) is compressed in its CPD form for space and time efficiencies, together in a novel optimization formulation presented below in Proposition 1.

Proposition 1. The tensor \( T \in \mathbb{R}^{I_n \times I_{n-1} \times \cdots \times I_1} \) of inferred \( n \)-way relations can be approximated and compressed in the rank-\( K \) CPD form \( \hat{T} = [A^{(n)}, A^{(n-1)}, \ldots, A^{(1)}] \), which is obtained by solving the following optimization problem:

\[
\begin{align*}
\min_{\{A^{(i)}: i=1,\ldots,n\}} \quad & \mathcal{J} = \sum_{i,j: i<j} \frac{1}{\prod_{l \neq i,j} I_l} \text{collap}(\hat{T}, i, j)^2_F + \lambda \text{vec}(\hat{T})^T L \text{vec}(\hat{T}) + \frac{\beta}{2} \sum_{i=1}^{n} ||A^{(i)}||_F^2 \\
\text{subject to} \quad & A^{(i)} \geq 0, \forall i = 1, \ldots, n,
\end{align*}
\]

where \( \text{collap}(\hat{T}, i, j) \) denotes collapsing tensor \( \hat{T} \) into an \( I_i \times I_j \) matrix by summing over the tensor slices along the corresponding modes (illustrated in Fig. 2 (I)); \( L \) is the Laplacian matrix of the graph \( G = (V, E) \), which can be any
one of the three types of product graphs defined in Section III-B; $\lambda$ and $\beta \in (0, 1)$ are hyperparameters.

The first term of the objective function $J$ in Equation (1) requires averaging over the slices of tensor $\hat{T}$ to be globally consistent with the observed bipartite relational matrices; the second term is called Laplacian regularization or smoothness constraint involving the product graph which will be discussed in Section III-B; the third term is the standard Tikhonov regularization which penalizes overly complex model to avoid over-fitting; and the CPD formulation of $\hat{T}$ guarantees the inferred $n$-way relations to be in a compressed form with low space complexity $O(K \sum I_i)$.

B. Laplacian regularization using product graphs

We denote the adjacency, degree and graph Laplacian matrices of graph $G^{(i)}$ as $W^{(i)}$, $D^{(i)}$ and $L^{(i)} = D^{(i)} - W^{(i)}$ for all $i = 1, \ldots, n$ respectively. The following three types of product graphs [25] are considered in this work.

1) Cartesian product graph $G^c$: two different nodes $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ in $G^c$ are connected if and only if there is a pair of nodes $a_i$ and $b_i$ connected in $G_i^{(i)}$ such that $(a_i, b_i) \in E^{(i)}$, and $a_j = b_j$ for all $j \neq i$. The adjacency and graph Laplacian matrices of $G^c$ are obtained as $W^c = \oplus_{i=1}^n W^{(i)}$ and $L^c = \oplus_{i=1}^n L^{(i)}$ respectively.

2) Tensor product graph $G^t$: two different nodes $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ in $G^t$ are connected if and only if $(a_i, b_i) \in E^{(i)}$ for all $i = 1, \ldots, n$. The adjacency and graph Laplacian matrices of $G^t$ are given by $W^t = \otimes_{i=1}^n W^{(i)}$ and $L^t = \otimes_{i=1}^n D^{(i)} - \otimes_{i=1}^n W^{(i)}$ respectively.

3) Strong product graph $G^s$: two different nodes $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ in $G^s$ are connected if and only if they are connected in either $G^c$ or $G^t$. The adjacency and graph Laplacian matrices of $G^s$ are obtained as $W^s = W^c + W^t$ and $L^s = L^c + L^t$ respectively.

Denoting $G = (V, E)$ as one of the three types of product graphs defined above, it is clear that the total number of nodes $|V| = \prod_{i=1}^n I_i$ in $G$ equals the number of elements in the tensor $\hat{T}$ in Equation (1). Therefore, $G$ can be regarded as a high-order representation of all the knowledge graphs $\{G^{(i)} = (V^{(i)}, E^{(i)}) : i = 1, \ldots, n\}$, whose graph Laplacian $L$ can be used to regularize the whole tensor $\hat{T}$. By introducing the second term $\text{vec}(\hat{T})^T L \text{vec}(\hat{T})$ in Equation (1), the approximated $n$-way relations in $\hat{T}$ are ensured to be smooth over the manifolds of the product graph $G$, such that a pair of tensor elements $\hat{T}_{a_1 \ldots a_n \ldots b_1}$ and $\hat{T}_{b_n \ldots b_1}$ share similar values if the nodes $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ in $G$ are connected. Note that by the definitions, strong and tensor product graphs have more edges compared with Cartesian product graph, and thus might encode richer similarity information among the tensor elements, while Cartesian product graph bridges two tensor elements under stricter condition which might incur less noise in defining...
the similarities. We consider all the three types of product graph manifolds in our model since both Cartesian and tensor product similarities exist in the real biomedical and biological networks as shown in Fig. 1, and Cartesian product graph also has a strong capability of jointly clustering the topologically related objects from multiple data domains as will be shown in the experiments.

C. GT-COPR Algorithm

The objective function $\mathcal{J}$ in Equation (1) is non-convex on variables $\{A^{(i)} : i = 1, \ldots, n\}$ jointly, thus finding its global minimum is difficult. In the following, we propose an efficient iterative algorithm Graph-Regularized Tensor Completion from Observed Pairwise Relations (GT-COPR) as summarized in Algorithm 1 to find the local minimum of $\mathcal{J}$ based on the multiplicative updating rule structurally similar to the method for solving non-negative matrix factorization (NMF) problems [15], [26]. Without loss of generality, we only show the derivations for the strong product graph regularization with the Laplacian matrix given by

$$L = L^c + L^t = \bigoplus_{i=1}^{n} L^{(i)} + \bigotimes_{i=1}^{n} D^{(i)} - \bigotimes_{i=1}^{n} W^{(i)}.$$ 

The derivations can be easily degenerated for the Cartesian ($L^c$) or tensor product ($L^t$) graph regularized GT-COPR with small modifications.

For simplicity, we rewrite the objective function in Equation (1) as $\mathcal{J} = \mathcal{J}_1 + \lambda \mathcal{J}_2 + \beta \mathcal{J}_3$. Let $a_T^{(i)} = 1_T^T A^{(i)}$ be the row summation of matrix $A^{(i)}$, we first define five auxiliary variables in Table I which are required for the derivations.

By expanding the tensor collapsing operator in $\mathcal{J}_1$ as:

$$\text{collapse}(\prod_{i,j} (A^{(n)}, A^{(n-1)}, \ldots, A^{(1)}), i, j) = \prod a_T^{(i)}, a_{T-1}^{(i)}, A^{(i)}, a_T^{(i+1)}, \ldots, a_T^{(n)}),$$

we obtain the partial derivative of the first term in a linear form of $A^{(i)}$ as:

$$\frac{\partial \mathcal{J}_1}{\partial A^{(i)}} = - (\sum_{j \neq i} \Theta_j^{(-i)}) + \sum_{j, k \neq i : j < k} \Theta_jk^{(-i)} + 1_T^T A^{(i)} (\sum_{j, k \neq i : j < k} \Phi_jk^{(-i)}) + A^{(i)} (\sum_{j \neq i} \Phi_j^{(-i)}).$$

Next, we expand the second objective term as follows

$$\mathcal{J}_2 = \frac{1}{2} \text{vec}(\prod_{i} A^{(n)}, A^{(n)}_1, \ldots, A^{(1)}_1) \text{vec}(\prod_{i} A^{(n)}, A^{(n)}_1, \ldots, A^{(1)}_1),$$

$$= \frac{1}{2} 1_T^T (\bigotimes_{i=1}^{n} A^{(i)}_T)^T L (\bigotimes_{i=1}^{n} A^{(i)}) 1$$

$$= \frac{1}{2} 1_T^T (\bigotimes_{i} A^{(i)}_T D^{(i)} A^{(i)}_T) - \bigotimes_{i} A^{(i)}_T W^{(i)} A^{(i)}_T$$

$$+ \sum_{i=1}^{n} (A^{(i)}_T L^{(i)} A^{(i)}_T) \bigotimes_{i \neq i} (A^{(i)}_T A^{(i)}_T 1).$$

Equation (2) holds by the Vectorization property of the CPD form given in Section II. Lemmas 1 and 2 are applied to obtain Equation (3). The partial derivative of $\mathcal{J}_2$ to $A^{(i)}$ is then obtained as

$$\frac{\partial \mathcal{J}_2}{\partial A^{(i)}} = L^{(i)} (\bigotimes_{j \neq i} (A^{(j)}_T D^{(j)} A^{(j)}_T)) + A^{(i)} (\sum_{j \neq i} \Phi_j^{(-i)}) +$$

$$D^{(i)} A^{(i)} (\bigotimes_{j \neq i} (A^{(j)}_T D^{(j)} A^{(j)}_T)) - W^{(i)} A^{(i)} (\bigotimes_{j \neq i} (A^{(j)}_T W^{(j)} A^{(j)}_T)).$$

**Lemma 1.** If matrices $A, B, C$ and $D$ are of such size that one can form the operation $(A \odot B), (C \odot D), (A^T C)$ and $(B^T D)$, then equality $(A \odot B)^T (C \odot D) = (A^T C) \odot (B^T D)$ holds.

**Lemma 2.** If matrices $A, B, C$ and $D$ are of such size that one can form the operation $(AC) \odot (BD)$, then equality $(A \odot B)(C \odot D) = (AC) \odot (BD)$ holds.

Combining $\frac{\partial \mathcal{J}_1}{\partial A^{(i)}}, \frac{\partial \mathcal{J}_2}{\partial A^{(i)}}, \frac{\partial \mathcal{J}_3}{\partial A^{(i)}}$ and $\frac{\partial \mathcal{J}_4}{\partial A^{(i)}} = A^{(i)}$, we obtain the partial derivative of $\mathcal{J}$ to factor matrix $A^{(i)}$ as the following linear form:

$$\frac{\partial \mathcal{J}}{\partial A^{(i)}} = -X^1 - X^2 A^{(i)} X^3 + X^4 A^{(i)} X^5$$

$$+ X^6 A^{(i)} X^7 + A^{(i)} X^8 + \beta A^{(i)},$$

where the matrices

$$X^1 = \sum_{j \neq i} \Theta_j^{(-i)}) + \sum_{j, k \neq i : j < k} \Theta_jk^{(-i)}, X^2 = \lambda W^{(i)}, X^3 = \bigotimes_{j \neq i} (A^{(j)}_T D^{(j)} A^{(j)}_T) + \bigotimes_{j \neq i} (A^{(j)}_T W^{(j)} A^{(j)}_T),$$

$$X^4 = \lambda D^{(i)}, X^5 = \bigotimes_{j \neq i} (A^{(j)}_T D^{(j)} A^{(j)}_T), X^6 = 1_T^T, X^7 = \sum_{j, k \neq i : j < k} \Phi_jk^{(-i)}$$

are all non-negative. According to Equation (4) we present Theorem 1 below to provide the updating rule of the proposed GT-COPR algorithm.

**Theorem 1.** Updating variables $\{A^{(i)} : i = 1, \ldots, n\}$ alternatively according to the rule given below can monotonically decrease the objective function $\mathcal{J}$ until they converge to the fixed-point solution.

$$A^{(i)}_T \leftarrow \frac{A^{(i)}_T (X^1 + X^2 A^{(i)} X^3)_T}{(X^4 A^{(i)})^T X^5 + X^6 A^{(i)} X^7 + A^{(i)} X^8 + \beta A^{(i)}},$$

**D. Theoretical Analysis**

Here, we provide theoretical analyses of GT-COPR with strong product graph regularization (given in Algorithm 1) to analyze the convergence and efficiency. The same analyses are also applicable to Cartesian and tensor product graph regularization with slight modifications.
Algorithm 1 GT-COPR (for strong product graph)

1: **Input:** 1) knowledge graphs \( \{G^{(i)} : i = 1, \ldots, n\} \), 2) observed bipartite relational matrices \( \{R_{i,j} \in \mathbb{R}_{+}^{I_i \times I_j} : \forall i, j \in [1, n] \text{ and } i < j\} \), 3) hyper parameters \( \lambda \) and \( \beta \), and 4) randomly initialized non-negative low-rank matrices \( \{A^{(i)} : i = 1, \ldots, n\} \) with rank \( K \).
2: **Output:** a CPD-form tensor \( T = \prod \{A^{(n)}, A^{(n-1)}, \ldots, A^{(1)}\} \) which stores the inferred \( n \)-way relations.
3: **while** not converge **do**
4: 5: **for** \( i = 1 \) to \( n \) **do**
6: 7: **end while**
8: **return** \( \{A^{(i)} : i = 1, \ldots, n\} \).

1) **Convergence Analysis:** As the objective function \( J \) is clearly bounded from below by zero, the convergence of the updating rule given in Theorem 1 can be proved by showing that \( J \) is non-increasing under updating. We adopt the similar proof procedure as in non-negative matrix factorization (NMF) [26] using the auxiliary function defined in Theorem 2.

**Theorem 2.** A function \( J(h) \) is non-increasing under the update \( h^* = \arg \min_{h} G(h, \hat{h}) \) if \( G(h, \hat{h}) \) is an auxiliary function for \( J(h) \), such that the following conditions are satisfied:

\[
G(h, \hat{h}) \geq J(h), \quad G(h, h) = J(h).
\]

By Theorem 2, the convergence claimed in Theorem 1 can be proved if the rule given in Theorem 1 is an update of one proper auxiliary function of \( J(A^{(i)}) \), which is defined in Theorem 3.

**Theorem 3.** The following function

\[
G(A^{(i)}_{ab}, \tilde{A}_{ab}) = J(\tilde{A}_{ab}) + J'(\tilde{A}_{ab})(A^{(i)}_{ab} - \tilde{A}_{ab}) + \frac{(X^4 \tilde{A}^{(i)} X_5 + X^6 \tilde{A}^{(i)} X_7 + \tilde{A}^{(i)} X^8 + \beta \tilde{A}^{(i)} ab (A^{(i)}_{ab} - \tilde{A}_{ab})^2}{2 \tilde{A}_{ab}}
\]

is an auxiliary function of \( J(A^{(i)}_{ab}) \) and has its global minimum.

Proof: First, it is obvious that \( G(A^{(i)}_{ab}, \tilde{A}_{ab}) = J(\tilde{A}_{ab}) \). To show \( G(A^{(i)}_{ab}, \tilde{A}_{ab}) \geq J(\tilde{A}_{ab}) \) we obtain the second-order Taylor expansion of \( J(A^{(i)}_{ab}) \) at the point \( \tilde{A}_{ab} \) as

\[
J(A^{(i)}_{ab}) = J(\tilde{A}_{ab}) + J'(\tilde{A}_{ab})(A^{(i)}_{ab} - \tilde{A}_{ab}) + \frac{1}{2} J''(\tilde{A}_{ab})(A^{(i)}_{ab} - \tilde{A}_{ab})^2,
\]

with the second-order derivative given below:

\[
J''(\tilde{A}_{ab}) = -X^2_{aa}X^3_{bb} + X^4_{aa}X^5_{bb} + X^6_{aa}X^7_{bb} + X^8_{bb} + \beta.
\]

Thus, the inequality \( G(A^{(i)}_{ab}, \tilde{A}_{ab}) \geq J(\tilde{A}_{ab}) \) holds if

\[
\frac{(X^4 \tilde{A}^{(i)} X_5 + X^6 \tilde{A}^{(i)} X_7 + \tilde{A}^{(i)} X^8 + \beta \tilde{A}^{(i)} ab (A^{(i)}_{ab} - \tilde{A}_{ab})^2}{2 \tilde{A}_{ab}} \geq J''(\tilde{A}^{(i)}_{ab}),
\]

which can be demonstrated by the facts that \( X^2_{aa}X^3_{bb} \geq 0 \), \( X^4_{aa}X^5_{mb} \geq X^4_{aa}X^3_{bb} \tilde{A}^{(i)}_{ab} \), \( X^6_{aa}X^7_{mb} \geq X^6_{aa}X^5_{bb} \tilde{A}^{(i)}_{ab} \), and \( (\tilde{A}^{(i)} X^8_{ab}) = \sum_i \tilde{A}^{(i)} X^8_{lb} \geq X^8_{bb} \tilde{A}^{(i)}_{ab} \). (End of Proof)

As the auxiliary function \( G(A^{(i)}_{ab}, \tilde{A}_{ab}) \) in Equation (5) is a quadratic function on variable \( A^{(i)}_{ab} \), its minimum can be easily obtained in a closed-form as

\[
A^{(i)*}_{ab} = \arg \min_{A^{(i)}_{ab}} G(A^{(i)}_{ab}, \tilde{A}_{ab}) = \frac{(\tilde{A}^{(i)}_{ab} \sum_1 X^1 + X^6 \tilde{A}^{(i)} X^8 + \beta \tilde{A}^{(i)}_{ab})}{(X^4 \tilde{A}^{(i)} X_5 + X^6 \tilde{A}^{(i)} X_7 + \tilde{A}^{(i)} X^8 + \beta \tilde{A}^{(i)}_{ab})},
\]

which leads to the updating rule in Theorem 1.

To analyze the optimality of the fixed point after convergence, we first define \( \{A^{(i)} \in \mathbb{R}^{I_i \times K} : i = 1, \ldots, n\} \) to be the matrices of Lagrange multipliers with the Lagrange function

\[
\mathcal{L} = J - \sum_{i=1}^n \text{tr}(A^{(i)} A^{(i)^T}).
\]

Settting \( \frac{\partial \mathcal{L}}{\partial A^{(i)}_{ab}} \) to be zero, we obtain \( \Lambda^{(i)}_{ab} = \frac{\partial J}{\partial A^{(i)}_{ab}} \). Furthermore, when \( A^{(i)} \) is a fixed point under the updating in Theorem 1 we have

\[
(X^4 \tilde{A}^{(i)} X_5 + X^6 \tilde{A}^{(i)} X_7 + \tilde{A}^{(i)} X^8 + \beta \tilde{A}^{(i)}_{ab} A^{(i)}_{ab}) = 0,
\]

which implies the KKT complementary slackness condition \( \Lambda^{(i)}_{ab} = 0 \) is satisfied.

2) **Complexity Analysis:** Assuming that \( n < I_i \) and \( K < I_i \), for all \( i = 1, \ldots, n \), and with a slight abuse of notation by denoting \( |S| \) also as the number of non-zeros in matrix \( S \), we first summarize the time complexities of computing the five auxiliary variables in Table 1. Based on it, we obtain the time complexity required to compute

\[
X^1 \text{ as } O(\sum_{j,k \in I_j} K|K| |R_{j,k} |), \text{ to compute } X^2 A^{(i)} X^3 + X^4 A^{(i)} X^5 + X^6 A^{(i)} X^7 + A^{(i)} X^8 + \beta A^{(i)}_{ab} A^{(i)}_{ab} = 0,
\]

Thus, the overall time complexity of updating the \( i \)-th factor matrix \( A^{(i)} \) in one iteration of GT-COPR is \( O(\sum_{j,k \in I_j} K|K| |W_{j,k} |) \). The space required to store the pairwise relational matrices is \( O(\sum_{j,k \in I_j} |R_{j,k} |) \), and to store the networks is \( O(\sum_{j,k \in I_j} |W_{j,k} |) \), and to store the factor matrices is \( O(\sum_{j,k \in I_j} K|K|) \). Thus, the overall space complexity is \( O(\sum_{j,k \in I_j} |R_{j,k} | + |W_{j,k} |) \). Though we only show the derivation for the strong product graph regularization, it is not hard to observe that the three types of product graphs share the same theoretical time and
space complexities. Note that the regularization with the tensor product graph does not involve the $\Psi_j^{-1}$ term in Table I, and thus can be empirically faster than using Cartesian or strong product regularization.

IV. EXPERIMENTS

![Fig. 3: Experimental data integrated from multiple sources.](image)

In this section we first describe the datasets that we integrate for learning disease-gene-chemical relations. Then we compare the prediction performance of GT-COPR with other methods through tensor fiber-wise and slice-wise evaluations. To show the potential clinical value of GT-COPR, we further validate the inferred triple-wise relations with the significant cancer-specific pharmacogenomic interactions reported in [1] from the analysis of the Genomics of Drug Sensitivity in Cancer (GDSC) [27] cell line dataset. Finally, we perform statistical analysis on the learned factor matrices to show that GT-COPR also detect topologically and biologically relevant disease, gene and chemical components.

A. Data Integration

We integrate multiple data sources to build the pairwise relations and knowledge graphs to infer the disease-gene-chemical associations. We first downloaded the pairwise relations from the Comparative Toxicogenomics Database (CTD) [21], which provides manually curated associations between chemicals, diseases and genes extracted from the published literature, with chemicals and diseases represented as Medical Subject Headings (MeSH) terms and genes represented as official gene symbols. Next, to obtain the networks we 1) downloaded the Homo sapiens protein-protein interactions (PPI) network from BioGRID 3.5 [3] as the gene-gene network; 2) we downloaded the human common disease network [2] as our disease-disease network where two common diseases are connected if their Human Phenotype Ontology (HPO) based phenotypic profiles share a high similarity; and 3) we construct the chemical-chemical network by first converting those CTD chemicals to PubChem 881-bit structure fingerprints using ChemRICH database [28], then add an edge between every pair of chemicals if the Tanimoto coefficient between their fingerprints is above 0.75 as suggested by [28]. The statistics of the integrated dataset are summarized in Fig. 3, which includes the data sources, numbers of graph nodes, numbers of pairwise relations, and densities of the networks and pairwise relational matrices.

B. Methods for Comparison

As mentioned in Section I, there is no tensor-based method developed for inferring cross-graph multi-relations from the observed pairwise relations. To benchmark the performance of GT-COPR, we compared it with five graph-based non-negative matrix factorization (NMF) methods. 1) wiZAN-Dual [17]: a dual graph regularized NMF with weight and imputation matrices in the objective, which was applied for prediction of drug-target interactions in [5]. 2) GWNMF [16]: a dual graph regularized NMF with the binary weight matrix indicating the observed and unobserved pairwise relations. 3) GWNTMF [16]: an alteration of the GWNMF to non-negative matrix tri-factorization. 4) FASCINATE [8]: a generalization of wiZAN-Dual to the joint factorization of multiple matrices, which was applied to infer the disease-gene, disease-chemical and gene-chemical pairwise relations simultaneously. 5) SNMF: symmetric NMF [29] applied on the matrix constructed by putting all graphs on the diagonal blocks and all pairwise relational matrices on the off-diagonal blocks.

We choose $\alpha$ from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ and fix $\beta = 0.1$ for GT-COPR. The graph hyperparameters of wiZAN-Dual/FASCINATE and GWNMF/GWNTMF are set by searching the grids $\{0.1, 0.5, 0.9, 1\}$ and $\{0.1, 1, 10, 100\}$ respectively, as suggested in their papers. Note that GT-COPR and the baseline methods use different scales of graph hyperparameters since the gradients of their variables are in different scales. To determine the rank $K$ of the factor matrices, we plot the top-1000 sorted singular values of each bipartite relational matrices in Fig. 4. Interestingly, we can observe that the spectral energy of each of the three bipartite relational matrices is dominated by their top-200 singular values. Therefore, we set $K$ to be the "elbow point" 200 for all the methods.

![Fig. 4: Elbow plots of the singular values of the bipartite relational matrices.](image)

C. Evaluation of the Predictive Performance

To evaluate the performance of predicting the disease-gene-chemical triples, we first construct a 3-way binary ground truth tensor $T$ with $T_{ijk}$ denoting the association between the i-th disease, j-th gene and k-th chemical, which is positive if at least two interactions among the triple are observed in the pairwise relational matrices. Then we evaluate the performances for predicting the disease-gene-chemical triples by disease, gene and chemical tensor fibers and slices respectively. Without loss of generality, we use the disease fiber...
and slice as examples to explain our evaluation procedures as follows.

- **Fiber-wise evaluation**: When evaluating the disease fiber $T_{jk}$ we first eliminate all the connections to the $j$-th gene and $k$-th chemical from the disease-gene and disease-chemical relational matrices respectively; next, we construct $T_{jk}$ using the learned low-rank matrices as described in Proposition 1, which is then treated as a score vector to predict $T_{jk}$.

- **Slice-wise evaluation**: When evaluating the disease slice $T_{i}$ we first eliminate all the connections to the $i$-th disease from the disease-gene and disease-chemical relational matrices respectively; next, we construct $T_{i}$ using the learned low-rank matrices as described in Proposition 1, which is then treated as a score matrix to predict $T_{i}$.

We randomly choose 10% of diseases, genes or chemicals as validation (5%) and test (5%) data and treat the rest 90% as training data for both fiber-wise and slice-wise evaluations. The random sampling procedures are repeated 10 times. The non-negative factor matrices of the baseline methods are treated as the CPD components for constructing the tensor $T$ as what GT-COPR reports. The predictive performances on the test data of all the methods are evaluated by average scores of AUROC (area under the receiver operating characteristics), MAP (mean average precision), hits at top 10 (Hits@10) and hits at top 5 (Hits@5), which are summarized in Table II and III. We can observe that all the three types of product graph regularized GT-COPR have very similar performances in both fiber-wise and slice-wise evaluations; by utilizing the topological information of the product graph via jointly regularizing the tensor elements with the product graph manifolds, GT-COPR consistently and significantly outperforms the other matrix factorization methods in all the fiber-wise evaluations and most of the slice-wise evaluations; SNMF and the soft-weighted methods FASCINATE and wiZAN-Dual also perform clearly better than the binary-weighted methods GWMTF and GWNF, which implies that the discrimination between the observed and unobserved pairwise relations are informative in our data, and thus it is more reasonable to consider the unobserved pairwise relations as negative samples than simply treat them as missing entries.

**D. Validation on Cancer Cell Line Data**

In [1], ANOVA analyses of the Genomics of Drug Sensitivity in Cancer (GDSC) cell line data find 182 significant cancer-specific interactions between differential drug sensitivity and cancer functional events (CFEs). The analyses are based on 1250 CFEs including somatic mutations, copy number alterations and DNA hypermethylation; and 265 clinical, clinical developmental and experimental anti-cancer drug compounds. This dataset is considered as an independent source from the integrated dataset described in Section IV-A.

We first map the 12 diseases (cancers), 1250 CEFs and 265 compounds to the ids in our integrated dataset, resulting in 104 interactions among 9 diseases, 670 genes and 100 chemicals. Next, we train each method with its optimal parameter found in the previous section, using all the observed pairwise relations and networks from the integrated database to obtain the low-rank factor matrices, which are then used to construct the $9 \times 670 \times 100$ subtensor in $\hat{T}$. Then we measure the performances of all the methods (GWNF/GWMTF are excluded due to their poor performances) by AUROC scores for predicting the disease specific gene-chemical interactions across all the 9 cancers (correspond to 9 tensor slices). To understand if the predictions made by GT-COPR are biased towards the observed pairwise relations, we also add a baseline using the binary tensor $\hat{T}$ constructed from the CTD pairwise relations mentioned in Section IV-C to make the same predictions. The scatter plots in Fig. 5 show that GT-COPR clearly outperforms the other methods in all the 9 cancer types, which provides the evidence of GT-COPR learning clinically meaningful pharmacogenomic multi-relations. The right-most column in Fig. 5 shows that $\hat{T}$ has a poor prediction performance on all

---

**TABLE II: Fiber-wise evaluation**

<table>
<thead>
<tr>
<th>Methods</th>
<th>AUROC</th>
<th>MAP</th>
<th>Hits@10</th>
<th>Hits@5</th>
<th>AUROC</th>
<th>MAP</th>
<th>Hits@10</th>
<th>Hits@5</th>
<th>AUROC</th>
<th>MAP</th>
<th>Hits@10</th>
<th>Hits@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT-COPR (Cartesian)</td>
<td>0.9125</td>
<td>0.1928</td>
<td>0.3605</td>
<td>0.4027</td>
<td>0.9092</td>
<td>0.2827</td>
<td>0.2252</td>
<td>0.2522</td>
<td>0.9032</td>
<td>0.4066</td>
<td>0.4377</td>
<td>0.5630</td>
</tr>
<tr>
<td>GT-COPR (Tensor)</td>
<td>0.9130</td>
<td>0.1894</td>
<td>0.3468</td>
<td>0.4192</td>
<td>0.8697</td>
<td>0.2829</td>
<td>0.4344</td>
<td>0.5489</td>
<td>0.9750</td>
<td>0.2759</td>
<td>0.3562</td>
<td>0.4463</td>
</tr>
<tr>
<td>GT-COPR (Strong)</td>
<td>0.8660</td>
<td>0.1604</td>
<td>0.2827</td>
<td>0.2864</td>
<td>0.7467</td>
<td>0.1463</td>
<td>0.2102</td>
<td>0.2481</td>
<td>0.9236</td>
<td>0.1097</td>
<td>0.1563</td>
<td>0.1877</td>
</tr>
<tr>
<td>SNMF</td>
<td>0.8978</td>
<td>0.1414</td>
<td>0.2579</td>
<td>0.2522</td>
<td>0.8378</td>
<td>0.2209</td>
<td>0.3310</td>
<td>0.3483</td>
<td>0.9704</td>
<td>0.1489</td>
<td>0.1535</td>
<td>0.1453</td>
</tr>
<tr>
<td>FASCINATE</td>
<td>0.8732</td>
<td>0.0924</td>
<td>0.0604</td>
<td>0.0321</td>
<td>0.3753</td>
<td>0.0886</td>
<td>0.1909</td>
<td>0.1948</td>
<td>0.7076</td>
<td>0.0185</td>
<td>0.0588</td>
<td>0.0519</td>
</tr>
</tbody>
</table>

**TABLE III: Slice-wise evaluation**

<table>
<thead>
<tr>
<th>Methods</th>
<th>AUROC</th>
<th>MAP</th>
<th>Hits@10</th>
<th>Hits@5</th>
<th>AUROC</th>
<th>MAP</th>
<th>Hits@10</th>
<th>Hits@5</th>
<th>AUROC</th>
<th>MAP</th>
<th>Hits@10</th>
<th>Hits@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT-COPR (Cartesian)</td>
<td>0.9945</td>
<td>0.0687</td>
<td>0.2523</td>
<td>0.6905</td>
<td>0.9853</td>
<td>0.0388</td>
<td>0.0935</td>
<td>0.0903</td>
<td>0.9708</td>
<td>0.0337</td>
<td>0.2123</td>
<td>0.1890</td>
</tr>
<tr>
<td>GT-COPR (Tensor)</td>
<td>0.9945</td>
<td>0.0687</td>
<td>0.2523</td>
<td>0.6905</td>
<td>0.9853</td>
<td>0.0388</td>
<td>0.0935</td>
<td>0.0903</td>
<td>0.9708</td>
<td>0.0337</td>
<td>0.2123</td>
<td>0.1890</td>
</tr>
<tr>
<td>GT-COPR (Strong)</td>
<td>0.9919</td>
<td>0.0802</td>
<td>0.3475</td>
<td>0.6905</td>
<td>0.9840</td>
<td>0.0303</td>
<td>0.0694</td>
<td>0.0710</td>
<td>0.9874</td>
<td>0.0392</td>
<td>0.2123</td>
<td>0.1890</td>
</tr>
<tr>
<td>SNMF</td>
<td>0.9032</td>
<td>0.0159</td>
<td>0.0759</td>
<td>0.0993</td>
<td>0.8568</td>
<td>0.0152</td>
<td>0.1403</td>
<td>0.1387</td>
<td>0.9181</td>
<td>0.0321</td>
<td>0.1156</td>
<td>0.1240</td>
</tr>
<tr>
<td>FASCINATE</td>
<td>0.9861</td>
<td>0.0223</td>
<td>0.0588</td>
<td>0.0182</td>
<td>0.8998</td>
<td>0.0159</td>
<td>0.0710</td>
<td>0.0613</td>
<td>0.9687</td>
<td>0.0155</td>
<td>0.0532</td>
<td>0.0474</td>
</tr>
<tr>
<td>wiZAN-Dual</td>
<td>0.9642</td>
<td>0.0369</td>
<td>0.0339</td>
<td>0.0616</td>
<td>0.9198</td>
<td>0.0111</td>
<td>0.0993</td>
<td>0.1000</td>
<td>0.9435</td>
<td>0.0096</td>
<td>0.1146</td>
<td>0.0994</td>
</tr>
<tr>
<td>GWMTF</td>
<td>0.7624</td>
<td>0.0003</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.8645</td>
<td>0.0005</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>GWNF</td>
<td>0.7583</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.7589</td>
<td>0.0004</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.8550</td>
<td>0.0003</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
the cancer types except the Lung Adenocarcinoma (LUAD), which means the CTD database covers only few interactions reported by [1], and further implies the knowledge graphs carry plenty of information for the multi-relational inference. One can also observe that the performances of using the three types of product graph regularization are very close in most of the cancer types. One remarkable difference is that when predicting the gene-chemical interactions in LGG (low grade glioma), the AUROC score is lifted by ~15% via using tensor product graph regularization.

E. Statistical Component Analysis

To evaluate the capability of CT-COPR to find topologically and biologically relevant disease, gene and chemical components, we 1) convert every column (component) in the rank-200 factor matrices $A^{(1)}$ (disease), $A^{(2)}$ (gene) and $A^{(3)}$ (chemical) learned by CT-COPR with the Cartesian product graph regularization to z-score vector, 2) select diseases, genes and chemicals with z-scores $> 2.33$ (p-value $< 0.01$) matched by components, 3) perform right-tailed Fisher’s exact test to find the most significant (p-value $< 10^{-3}$) disease components via comparing with the 13 disease categories reported by [2], and 4) apply gene and chemical enrichment analyses with Gene Ontology (GO) Consortium [30] and IMPaLA [31] respectively to find the GO terms and pathways related to the selected genes and chemicals in each of the disease components found in step 3).

There are 101 biological related disease, gene and chemical components found by the procedures described above. Due to the page limit, we only show the two components corresponding to cardiovascular system and cellular proliferation disease categories respectively. Fig. 6 shows the top-5 enriched GO terms and chemical pathways, and three sub-networks containing subsets of the matched diseases, genes and chemicals related to the cardiovascular system disease category. Interestingly, the significantly enriched GO terms and chemical pathways are all closely related to the cardiovascular system. For example, the heart failure is known to be a syndrome characterized by up regulation of the "sympathetic nervous" system [32]; and "organic cation transporters" (OCT1-3 and OCTN1/2) facilitate cardiac uptake of endogenous compounds and numerous drugs [33]. Moreover, each of the three sub-networks is fully connected, with densities of the networks and cross-network bipartite relational matrices significantly higher than the background densities of the integrated dataset given in Fig. 3. We also show the subsets of the detected diseases in cellular proliferation component, together with two lists of top enriched GO terms and chemical pathways in Table IV. The first column shows that all the diseases are neoplasms (caused by an abnormal proliferation of tissues). The last two columns show that the enriched GO terms and chemical pathways...
are all cancer related. For example, the GO terms "lipid metabolic process" and "cellular lipid metabolic process" are believed to be cancer-development related by a recent study [34]; and the "cytochrome P450" enzymes are known to be important targets in cancer, due to their role in "xenobiotic metabolism" [35]. Overall, the results demonstrate that the components produced by GT-COPR have both topologically and biologically interpretations.

F. Implementation and Running time

We implemented GT-COPR using MATLAB (R2018a) with the Tensor Toolbox v 2.6 [36] on a server with Intel(R) Xeon(R) CPU E5-2450 (32 cores 2.10GHz, 2 CPUs) and 196GB of RAM. Fig. 7 shows that GT-COPR is able to learn the disease-gene-chemical multi-relations using the integrated dataset in less than one CPU hour. Empirically, the implementation of the tensor product graph regularized GT-COPR scales almost linearly with the tensor rank and is faster than the Cartesian and strong product graph regularized versions.

TABLE IV: Detected components of cellular proliferation (p = 4.7 × 10^{-13})

<table>
<thead>
<tr>
<th>Disease names</th>
<th>Enriched GO terms</th>
<th>Enriched chemical pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>breast neoplasms</td>
<td>small molecule metabolic process (p = 3.1 × 10^{-21})</td>
<td>pathways in cancer (KEGG) (p = 9.1 × 10^{-10})</td>
</tr>
<tr>
<td>hepatocellular carcinoma</td>
<td>lipid metabolic process (p = 1.6 × 10^{-18})</td>
<td>xenobiotics (Reactome) (p = 2.4 × 10^{-8})</td>
</tr>
<tr>
<td>adenocarcinoma</td>
<td>cellular lipid metabolic process (p = 1.3 × 10^{-18})</td>
<td>cytochrome P450 (Reactome) (p = 2.0 × 10^{-7})</td>
</tr>
<tr>
<td>squamous cell carcinoma</td>
<td>metabolic process (p = 5.9 × 10^{-18})</td>
<td>biological oxidation (Reactome) (p = 3.8 × 10^{-7})</td>
</tr>
<tr>
<td>colonic neoplasms</td>
<td>carboxylic acid metabolic process (p = 1.9 × 10^{-17})</td>
<td>prostate cancer (KEGG) (p = 4.7 × 10^{-7})</td>
</tr>
<tr>
<td>liver neoplasms</td>
<td>oxoacid metabolic process (p = 2.8 × 10^{-17})</td>
<td>chemical carcinogenesis (KEGG) (p = 1.6 × 10^{-6})</td>
</tr>
<tr>
<td>lung neoplasms</td>
<td>organic acid metabolic process (p = 5.3 × 10^{-17})</td>
<td>breast cancer pathway (Wikipathways) (p = 6.3 × 10^{-6})</td>
</tr>
<tr>
<td>lymphoma</td>
<td>organic substance metabolic process (p = 1.67 × 10^{-15})</td>
<td>hepatocellular carcinoma (KEGG) (p = 3.4 × 10^{-5})</td>
</tr>
<tr>
<td>cholangiocarcinoma</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6: Sub-networks of the detected components in the cardiovascular system disease category. The densities of each sub-network and bipartite sub-network are given together with the background densities in parentheses.

Fig. 7: Comparison of the running time of CT-COPR with three types of product graph regularization.

V. CONCLUSION

In this study, we introduced a novel and general tensor-based algorithm GT-COPR for learning pharmacogenomic
multi-relations across multiple networks, utilizing the observed pairwise relations and the topological information of three different types of product graphs, without relying on known initial multi-relations. We proved the convergence and efficiency of GT-COPR by theoretical analyses. We observed initial multi-relations. We proved the convergence and efficiency of GT-COPR by theoretical analyses. We observed 

**REFERENCES**


