Graph Regularized Probabilistic Matrix Factorization for Drug-Drug Interactions Prediction

Stuti Jain, Emilie Chouzenoux, Kriti Kumar, Angshul Majumdar

Graph Regularized Probabilistic Matrix Factorization for Drug-Drug Interactions Prediction

Stuti Jain, Emilie Chouzenoux, *Senior Member, IEEE,* Kriti Kumar, *Member, IEEE* and Angshul Majumdar, *Senior Member, IEEE*

***Abstract*—Co-administration of two or more drugs simultaneously can result in adverse drug reactions. Identifying drug-drug interactions (DDIs) is necessary, especially for drug development and for repurposing old drugs. DDI prediction can be viewed as a matrix completion task, for which matrix factorization (MF) appears as a suitable solution. This paper presents a novel Graph Regularized Probabilistic Matrix Factorization (GRPMF) method, which incorporates expert knowledge through a novel graph-based regularization strategy within an MF framework. An efficient and sounded optimization algorithm is proposed to solve the resulting non-convex problem in an alternating fashion. The performance of the proposed method is evaluated through the DrugBank dataset, and comparisons are provided against state-of-the-art techniques. The results demonstrate the superior performance of GRPMF when compared to its counterparts.**

***Index Terms*—Matrix factorization, Probabilistic matrix factorization, Graph regularization, Drug-drug interaction prediction**

I. Introduction

Drug-Drug Interaction (DDI) refers to the effects of a given drug when taken together with another drug, at the same time. Co-administration of two or more drugs simultaneously can affect the pharmokinetics and/or pharmacodynamics of one or more drugs, which can cause unexpected and even adverse drug reactions [1]. These effects can cause severe injuries to the patients and even be responsible for deaths. Thus, it is necessary to know the DDI for the drugs used in the market, for clinical safety. Knowledge of DDI is also vital for developing new drugs, as well as for repurposing old drugs from clinical and public health perspectives. Preclinical identification of DDIs is an ill-posed problem as clinical testing (e.g., in vitro, in vivo, and in populo) is usually conducted on a small group of drugs. Such a process is time and cost intensive. Thus, DDI prediction employing computational approaches that are implementable on a large scale has become a popular research topic in recent years [2].

Computational approaches for DDI prediction can be broadly divided into two categories: (i) Similarity-based approaches, that are based on the similarity of drug information (e.g., chemical structure [3], targets [4], side-effects [5], and (ii) knowledge-based approaches, that employ text mining from scientific literature [6], electronic medical record database [7] and the FDA Adverse Event Reporting System [8] to predict DDI. Note that the latter group of approaches do not perform learning per se, as it is mostly a tool to mine clinical findings. Although different machine learning and artificial intelligence models have been used to address the problem [9], DDI prediction remains a challenging problem to address.

In this work, the DDI prediction task is formulated as a matrix completion problem, involving a symmetric matrix with rows/columns corresponding to drugs. We adopt a matrix factorization paradigm, when the sought matrix is defined as the product of two latent factors satisfying some prior knowledge. Our contribution lies in the construction of an original prior well suited to DDI prediction. It incorporates expert knowledge on the DDIs within a graph-based regularization term similar to [10], with the aim to favor expected (dis)similarities between drug pairs. This formulation results in the so-called Graph Regularized Probabilistic Matrix Factorization (GRPMF) method, for which we also propose a sounded optimization algorithm relying on modern proximal methods. We evaluate the performance of our method on the DrugBank dataset, and present comparisons against state-of-the-art techniques.

The paper is organized as follows. Section II discusses related works on DDI prediction. Section III provides a brief overview of various MF-based methods. Section IV presents our main contribution, that is the proposed GRPMF formulation and the optimization algorithm to resolve it. Section V presents our experimental results and comparisons with the state-of-the-art methods, and finally, Section VI concludes the work.

II. Related Works

Clinical trials are time and money consuming. Many DDIs thus remain unknown, mostly because they are not tested during trials. Machine learning-based methods have been widely investigated to predict the unobserved DDIs. A heterogeneous network-assisted inference (HNAI) method is proposed in [11], gathering five prediction models (naive Bayes, decision tree, k-nearest neighbor, logistic regression, and support vector machine) to perform DDIs prediction from drug phenotypic, therapeutic, structural, and genomic similarities. The work in [12] presents a neural network (NN) based method that proposes a heuristic selection of several drug similarity scores integrated with a nonlinear similarity fusion strategy to obtain abstract features for DDI prediction. Another work [13] proposes a semi-supervised learning method for DDI prediction,

S. Jain and E. Chouzenoux are with CVN, Inria Saclay, Univ. Paris Saclay, 91190 Gif-sur-Yvette, France. e-mail: (emilie.chouzenoux@centralesupelec.fr, stuti.jain@inria.fr).

A. Majumdar and K. Kumar is with Dept. of ECE, IIIT - Delhi, India, 110020. e-mail: (angshul@iiitd.ac.in, kritik@iiitd.ac.in).

This work received support from the Associate Team COMPASS between Inria and IIIT Delhi. E.C. and S.J. acknowledge support from the European Research Council Starting Grant MAJORIS ERC-2019-STG-850925.

that uses drug chemical, biological, and phenotype data to calculate the feature similarity of drugs using a regularized least square score minimization.

Some works address the DDI prediction problem as an edge detection (i.e., link prediction) problem where the edges to infer represent connections between drugs. The work in [14] presents both unsupervised and supervised techniques for link prediction using binary classifiers such as tree, k-nearest neighbors, support vector machine, random forest, and gradient boosting machine based on topological and semantic similarity features to estimate the drug interactions. Another work [15] proposes two methods based on NNs and factor propagation over graph nodes, namely, adjacency matrix factorization (AMF) and adjacency matrix factorization with propagation (AMFP) for link prediction for discovering DDIs.

The superior performance of deep learning (DL) techniques across different domains has triggered the interest in such techniques to estimate drug interactions. The work in [16] presents a biomedical resource LSTM (BR-LSTM) that combines biomedical resources with lexical information and entity position information together to extract DDI from the biomedical literature. Note that this model is not DDI prediction per se, but only an automatic tool for mining of information from clinical literature. The work in [17] proposes a convolutional mixture density recurrent NN model that integrates convolutional neural networks, recurrent NNs, and mixture density networks for DDI prediction. An autoencoder-based semi-supervised learning algorithm for feature extraction from FDA adverse event reports to identify potential high priority DDIs for medication alerts is presented in [18]. Another work [19] employs autoencoders and a deep feed-forward network trained on structural similarity profiles (SSP), Gene Ontology (GO) term similarity profiles (GSP), and target gene similarity profiles (TSP) of known drug pairs to predict the effects of DDIs. Due to the black-box nature of the DL models, some work has been done on seeking for explainable DL-based DDI techniques. A comprehensive review of the explainable AI-based techniques to promote the trust of AI models for the critical task of DDI prediction is presented in [20].

Recently, attention-based deep learning models have been explored for DDI prediction task that can predict novel drug-drug interactions and drug-drug interaction-associated events [21], [22], [23]. The work in [21] presented a Siamese self-attention multi-modal neural network for DDI prediction that utilizes multiple drug similarity measures derived from drug characteristics, including drug targets, pathways, and gene expression profiles. Here, model explainability is offered via an Attention mechanism for identifying salient input features. Another work in [22] proposed two attention-mechanism-based encoder–decoder models that incorporate multisource information derived from drugs. They considered three encoding methods to encode the drugs that consisted of (i) integrating the self-attention mechanism, (ii) cross-attention mechanism, and (iii) graph attention network to construct a multisource feature fusion network. Different encoders are constructed based on the different features to learn the interrelationship information between drug pairs, and that information is used to predict DDI. The work in [23] presented a Deep Attention Neural Network based DDI prediction framework, abbreviated as DANN-DDI that uses a graph embedding method to learn drug representations from multiple drug feature networks and adopts an attention neural network to learn representations of drug-drug pairs that are finally used to predict DDI.

Another direction of work formulates the DDI prediction problem as a Matrix Completion task. Here, given the partially observed DDI matrix, the task is to compute the unobserved interactions between the drugs. Some of the popular generic (not tailored for DDI) MF techniques are (i) singular value decomposition (SVD) [24], (ii) non-negative matrix factorization (NMF) [25] and, (iii) probabilistic matrix factorization (PMF) [26]. We will present the two later approaches in detail in our next section, as MC/MF constitutes the core of our contribution.

In addition to the conventional binary DDI prediction, the work in [27] presents an NMF-based approach utilizing drug features for comprehensive DDI prediction. Here, the comprehensive DDI matrix is a signed binary matrix with $+1$ for enhancive drugs, -1 for degressive drugs, and 0 for no drug interactions, respectively, which is rather useful to predict the (positive/negative) behaviors of the interacting drugs. The work in [28] presents an attribute supervised learning model probabilistic dependent matrix tri-factorization (PDMTF) approach for adverse DDI prediction. They utilized two drug attributes, molecular structure, side effects, and their correlation to compute the adverse interactions among drugs. The work in [29] introduces a manifold regularized MF (MRMF) technique to predict DDIs using drug similarities based on drug features like substructures, targets, enzymes, transporters, pathways, indications, side effects, and off side effects. Different drug features and similarity measures are considered to calculate different manifold regularization terms to construct the MRMF models. The inclusion of drug feature-based manifold regularization in the matrix factorization formulation is reported to provide more accurate and robust predictions.

The publicly available large structured biomedical databases has enabled the use of knowledge graph (KG) based approaches for different applications in the biomedical domain. KGs are used to synthesize large biomedical graphs that map similar drug-related entities in the drug database. The work in [30] uses KGs embeddings, namely, RDF2Vec, TransE, TransD, and machine learning algorithms for DDI prediction. A KG NN method (KGNN) that captures the drug and its potential neighborhoods by mining their associated relations in KG for DDI prediction is proposed in [31]. This method utilizes the drugs' topological structures in KG for potential DDI prediction. Another work [32] utilizes KGs combined with DL techniques for estimating DDIs. This work considers the DDI matrix and KG in the form of learned embeddings (like ComplEx, TransE, RDF2Vec, etc.) as input to the Convolutional Neural Networks (CNN) and Long-Short Term Memory (LSTM) model to predict DDIs.

In this work, we focus on the MC/MF based framework, as it presents the advantage of being non supervised and highly interpretable. Our contribution is to incorporate expert knowledge within this family of approach, so as to take advantage

of the aforementioned progressed in database availability.

III. Background

This section presents an overview of MF techniques for MC. We choose here to remain in a generic setting where the matrix to complete is real-valued and rectangular. Note that, for the DDI task, the sought matrix is square symmetric and, in most cases, binary valued, which might lead to simplified formulations.

A. Matrix Completion Problem

Let us consider the problem of a full matrix $R \in \mathbb{R}^{N\times M}$ to recover from partially known matrix $Y \in \mathbb{R}^{N\times M}$. Let

$$\mathcal{D} = \{i \in \{1,\dots,N\}, j \in \{1,\dots,M\} \text{ s.t. } (i,j) \text{ is observed}\}.$$

Non observed entries are typically set to zero. The masking of the indexes outside the set $\mathcal{D}$ is modeled through a Hadamard product $\odot$ with a matrix $B \in \{0,1\}^{N\times M}$, such that $B_{ij} = 1$ if $(i,j) \in \mathcal{D}$, and $B_{ij} = 0$ otherwise. The partially known matrix Y can be expressed as:

$$Y = B \odot R. \tag{1}$$

The task of matrix completion amounts to recovering the entries of R that do not belong to the set of observed indexes $\mathcal{D}$.

B. Matrix Factorization (MF)

MF [33] consists of recovering missing entries in matrix R by minimizing the simple least-squares function $\underset{R}{\text{minimize}}\ \|Y - B\odot R\|_F^2$ under some specific structural prior constraints on R. Under MF prior, R is recast as a product of two matrices $U \in \mathbb{R}^{N\times Z}$ and $V \in \mathbb{R}^{Z\times M}$, where $Z \geq 1$ defines a latent space dimension, typically low compared to (N, M). The matrices U and V are inferred by solving:

$$\underset{U,V}{\text{minimize}}\ \|Y - B \odot (UV)\|_F^2. \tag{2}$$

Subsequently, the complete matrix R is simply recovered by $R = UV$. Problem (2) is however highly under-determined and extra priors are typically introduced to obtain meaningful solutions. The most widely used being probably the positivity of the entries of the latent factors (U, V), yielding the NMF (nonnegative MF) formulation [34].

Let us discuss related formulations for MC. First, another formulation strategy to impose low rank is to resort to nuclear norm minimization [35]. Regularization strategies, based on graph modeling, have been considered in [36] for the MF formulation and in [37] for the nuclear norm formulation. MF models with more than two factors lead to the so-called deep MF approach, investigated for instance in [38] in the context of MC. Graph regularized version for deep MC has been proposed in the recent work [39].

C. Probabilistic Matrix Factorization (PMF)

PMF introduces probabilistic models on the latent factors U and V in the MF formulation [40]. More precisely:

- Each observed entry $(Y_{ij})_{(i,j)\in\mathcal{D}}$ is assumed to follow a Gaussian distribution, with mean $[UV]_{i,j}$ and variance σ^2 (positive scalar assumed to be known).
- Each entry $(U_{iz})_{1\leq i\leq N, 1\leq z\leq Z}$ is assumed to follow a Gaussian distribution with zero mean and variance σ_U^2 (positive scalar assumed to be known).
- Each entry $(V_{zj})_{1\leq z\leq Z, 1\leq j\leq M}$ is assumed to follow a Gaussian distribution with zero mean and variance σ_V^2 (positive scalar assumed to be known).

The maximum a posteriori (MAP) estimator of (U, V) given Y, associated with the above model can be obtained by solving:

$$\underset{U,V}{\text{minimize}} \frac{1}{2\sigma^2}\|Y - B\odot(UV)\|_F^2 + \frac{1}{2\sigma_U^2}\|U\|_F^2 + \frac{1}{2\sigma_V^2}\|V\|_F^2. \tag{3}$$

The minimization with respect to U (resp. V) in this formulation amounts to invert a linear system, which can be performed using a conjugate gradient solver. The PMF formulation can be enhanced by incorporating correlated Gaussian distributions, which results in the PMFG formulation described hereafter.

D. Probabilistic Matrix Factorization with Graph regularization (PMFG)

In PMFG, the prior distributions of U and V now include a graph regularization strategy [10], which amounts to inferring correlations along the rows (resp. columns) of U (resp. V) jointly with the factors U and V. These correlations are modeled through two precision matrices, $\Gamma_U \in S_N^{++}$ and $\Gamma_V \in S_M^{++}$, where $S_N^{++} \in \mathbb{R}^{N\times N}$ and $S_M^{++} \in \mathbb{R}^{M\times M}$ denote symmetric positive definite matrices while S_N, S_M denote symmetric matrices. The prior is the following:

- The columns $u_z \in \mathbb{R}^N$, $z \in \{1,\dots,Z\}$, of U are independent realizations of a multivariate Gaussian distribution with zero mean and covariance $C_U = \Gamma_U^{-1}$;
- The lines $v_z \in \mathbb{R}^M$, $z \in \{1,\dots,Z\}$ are independent realizations of a multivariate Gaussian distribution with zero mean and covariance $C_V = \Gamma_V^{-1}$.

The precision matrices Γ_U and Γ_V are related to Gaussian graphical models associated to the two underlying Gaussian distributions [cite the book Elements-Statistical-Learning-Inference-Prediction], which justifies the name for "graph" regularization. Specifically, matrix Γ_U (resp. Γ_V) can be understood as the adjacency matrix of an undirected graph where each edge identifies with two entries of u_z (resp. v_z) being correlated, given all the others.

The MAP estimate can now be obtained by solving the following:

$$\begin{aligned}\underset{U,V,\Gamma_U,\Gamma_V}{\text{minimize}}\ & \frac{1}{2\sigma^2}\|Y - B\odot(UV)\|_F^2 + \frac{1}{2}\operatorname{tr}(U^\top\Gamma_U U) \\ & + \frac{1}{2}\mathcal{L}(\Gamma_U) + \frac{1}{2}\operatorname{tr}(V\Gamma_V V^\top) + \frac{1}{2}\mathcal{L}(\Gamma_V),\end{aligned} \tag{4}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operation and

$$(\forall \Gamma_U \in S_N) \quad \mathcal{L}(\Gamma_U) = \begin{cases} -\ln\det(\Gamma_U) & \text{if } \Gamma_U \in S_N^{++} \\ +\infty & \text{otherwise} \end{cases} \tag{5}$$

with $\det(\cdot)$ the determinant operation.

PMFG approach provides promising results in [10]. However, it does not incorporate any physical-oriented knowledge in the sought factors (U, V). Actually, in many applications, such as DDI, expert knowledge is available, that dictates more or less likely correlations among the variables. The aim of this present work is to propose a novel formulation to account for such prior knowledge, within the PMFG paradigm.

IV. Proposed Graph Regularized Probabilistic Matrix Factorization (GRPMF)

Let us specify our targeted application. We focus on solving the matrix completion problem arising when predicting interactions between the different drugs, which is the so-called DDI problem. Let $Y \in \mathbb{R}^{N\times N}$ a partially known drug interaction matrix (with unobserved entries set to 0) for N different drugs. The aim is to recover the full drug interaction matrix $R \in \mathbb{R}^{N\times N}$[1]. The proposed method includes expert knowledge within the formulation of PMFG, with the aim of estimating DDIs. Note that unlike general matrix completion problems discussed in previous section, in DDI the matrix to infer is square and symmetric. The sought interaction matrix R can thus be factored as:

$$R = UU^\top. \tag{6}$$

Under this new setting, the graphical model is reduced to a single graph with adjacency matrix Γ_U, and the previously presented PMFG formulation can simply be modified as:

$$\underset{U,\,\Gamma_U}{\text{minimize}} \frac{1}{2\sigma^2}\|Y - B\odot(UU^\top)\|_F^2 + \frac{1}{2}\,\mathrm{tr}(U^\top\Gamma_U U) + \frac{1}{2}\mathcal{L}(\Gamma_U). \tag{7}$$

A. Integrating Expert Knowledge

As already discussed, in DDI application, one might have prior knowledge about the position of the graph edges (i.e. non zero elements in the precision matrix Γ_U), thanks to some expert analysis of the database. This prior knowledge is available through an extra symmetric matrix with positive real entries, namely $A_U \in [0, +\infty[^{N\times N}$. For instance, for the DDI prediction task, A_U could result from precomputing the similarity between the N drugs of the dataset in terms of the SIMCOMP (SIMilar COMPound) scores [41]. Such matrix can then be used as a structural prior on the sought matrix Γ_U, so as to remove spurious edges with no physical meaning, and to promote expected ones in the restored graph. Otherwise stated, if for some (i, j) with $i \neq j$, $[A_U]_{ij}$ is large, then $[\Gamma_U]_{ij} = [\Gamma_U]_{ji}$ should be encouraged to be high as well. In contrast, an entry $[A_U]_{ij}$ close or equal to zero should promote the removal of the edge between nodes i and j in the sought graph.

In order to build a suitable regularization function associated to this new prior, let us introduce the following sets:

$$\mathcal{E}_U = \{(i,j) \in \{1,\dots,N\}^2, i \neq j \text{ and } [A_U]_{ij} > \tau\} \tag{8}$$

$$\overline{\mathcal{E}}_U = \{(i,j) \in \{1,\dots,N\}^2, i \neq j \text{ and } [A_U]_{ij} \leq \tau\} \tag{9}$$

with $\tau \geq 0$ a given detection threshold. In order to avoid a tedious non-convex coupling term in the resulting minimization problem, we furthermore introduce the proxy variable $X \in \mathbb{R}^{N\times N}$ that we penalize so as to be close to the sought product $UU^\top$. We thus propose to solve:

$$\begin{aligned} \underset{X,\,U,\,\Gamma_U}{\text{minimize}} \; & \frac{1}{2\sigma^2}\|Y - B\odot X\|_F^2 + \frac{1}{2}\,\mathrm{tr}(U^\top\Gamma_U U) \\ & + \lambda_U \sum_{(i,j)\in\overline{\mathcal{E}}_U} |[\Gamma_U]_{ij}| - \lambda_U \sum_{(i,j)\in\mathcal{E}_U} \ln(|[\Gamma_U]_{ij}| + \delta) \\ & + \frac{\lambda_R}{2}\|X - UU^\top\|_F^2 + \frac{1}{2}\mathcal{L}(\Gamma_U) + \frac{\lambda_U}{2}\|\Gamma_U\|_F^2. \end{aligned} \tag{10}$$

Hereabove, parameter $\lambda_R > 0$ controls the fulfillment of the equality constraint $X = UU^\top$ while parameter $\lambda_U > 0$ controls the regularization imposed on the precision matrix Γ_U. Our proposed regularization term is made of two parts. We introduce an ℓ_1 term to promote sparsity on the regions where edges should not appear (i.e. $\overline{\mathcal{E}}_U$), and a log-barrier term, smoothed by $\delta > 0$, in regions where edges (i.e., non-zero entries) should be promoted (i.e. $\mathcal{E}_U$). Finally, a quadratic term, that can be viewed as an elastic-net penalty, is added in order to avoid too large values in the entries of the sought covariance matrix.

B. Optimization Algorithm

Problem (10) is highly non-convex, as it is commonly the case in MF formulations. We propose to use an alternating optimization strategy to solve it. Let $F(X, U, \Gamma_U)$ denote the loss function presented in (10). Starting from a given initialization (X^0, U^0, Γ_U^0), for every iteration $k \in \mathbb{N}$, the parameters of the algorithm are updated as:

$$\begin{cases} X^{k+1} = \underset{X\in\mathbb{R}^{N\times N}}{\mathrm{argmin}}\, F(X, U^k, \Gamma_U^k), \\ U^{k+1} = \underset{U\in\mathbb{R}^{Z\times Z}}{\mathrm{argmin}}\, F(X^{k+1}, U, \Gamma_U^k), \\ \Gamma_U^{k+1} = \underset{\Gamma_U\in S_N}{\mathrm{argmin}} F(X^{k+1}, U^{k+1}, \Gamma_U). \end{cases} \tag{11}$$

Such procedure ensures the monotonical decrease of function F. We explicit hereafter each update. Let us remark that, for variables U and Γ_U, the subproblems remain non-convex. The proposed subroutines only amounts to finding a stationary point for each, which might not be a global minimum. No numerical instabilities were observed in our experiments though.

1) Update of X: The matrix X is updated as:

$$X^{k+1} = \underset{X\in\mathbb{R}^{N\times N}}{\mathrm{argmin}} \frac{1}{2\sigma^2}\|Y - B\odot X\|_F^2 + \frac{\lambda_R}{2}\|X - U^k(U^k)^\top\|_F^2. \tag{12}$$

[1] In our experimental part, due to specificity of the retained dataset, R is a binary valued matrix. However, our framework holds for any type of real-valued symmetric DDI matrix R.

This is a strictly convex quadratic problem, whose solution satisfies the following optimality condition:

$$\frac{1}{\sigma^2}Y + \lambda_R U^k (U^k)^\top = \frac{1}{\sigma^2} B \odot X^{k+1} + \lambda_R X^{k+1}. \quad (13)$$

The above equation is a linear system that can be solved efficiently by conjugate gradient [42].

2) Update of U: The matrix U is updated as:

$$U^{k+1} = \underset{U \in \mathbb{R}^{Z \times Z}}{\operatorname{argmin}} \frac{1}{2} \operatorname{tr}(U^\top \Gamma_U^k U) + \frac{\lambda_R}{2} \|X^{k+1} - UU^\top\|_F^2. \quad (14)$$

The above optimization problem is non-convex and differentiable. An efficient nonlinear conjugate gradient method is proposed in [43] to address it.

3) Update of Γ_U: The matrix Γ_U is updated using:

$$\Gamma_U^{k+1} = \underset{\Gamma_U \in S_N}{\operatorname{argmin}} \operatorname{tr}((U^{k+1})^\top \Gamma_U U^{k+1}) + \mathcal{L}(\Gamma_U) + \lambda_U \|\Gamma_U\|_F^2 + 2\lambda_U \sum_{(i,j) \in \overline{\mathcal{E}}_U} |[\Gamma_U]_{ij}| - 2\lambda_U \sum_{(i,j) \in \mathcal{E}_U} \ln(|[\Gamma_U]_{ij}| + \delta) \quad (15)$$

or, equivalently,

$$\Gamma_U^{k+1} = \underset{\Gamma_U \in S_N}{\operatorname{argmin}} f(\Gamma_U) + g(\Gamma_U) \quad (16)$$

with

$$(\forall \Gamma_U \in S_N) \quad f(\Gamma_U) = \operatorname{tr}((U^{k+1})^\top \Gamma_U U^{k+1}) + \mathcal{L}(\Gamma_U) \quad (17)$$

and

$$(\forall \Gamma_U \in S_N) \quad g(\Gamma_U) = 2\lambda_U \sum_{(i,j) \in \overline{\mathcal{E}}_U} |[\Gamma_U]_{ij}| - 2\lambda_U \sum_{(i,j) \in \mathcal{E}_U} \ln(|[\Gamma_U]_{ij}| + \delta) + \lambda_U \|\Gamma_U\|_F^2. \quad (18)$$

The minimization of $f+g$ does not have a close form solution and hereagain, an inner solver is required. Function f is convex, differentiable on its domain S_N^{++}, while g is non-convex, non-differentiable. Luckily, the latter is separable over each of the entries of Γ_U, that is:

$$(\forall \Gamma_U \in S_N) \quad g(\Gamma_U) = \sum_{1 \leq i,j \leq N} g_{ij}([\Gamma_U]_{ij}) \quad (19)$$

with, for every $(i,j) \in \{1, \dots, N\}^2$,

$$(\forall \omega \in \mathbb{R}) \quad g_{ij}(\omega) = \begin{cases} 2\lambda_U |\omega| + \lambda_U \omega^2 & \text{if } (i,j) \in \overline{\mathcal{E}}_U, \\ -2\lambda_U \ln(|\omega| + \delta) + \lambda_U \omega^2 & \text{if } (i,j) \in \mathcal{E}_U. \end{cases} \quad (20)$$

We thus opt for running $L \geq 1$ iterations of a proximal gradient algorithm [44], initialized using the previous value $\Gamma_U^k \in S_N^{++}$ (by construction). This reads as follows:

$$\begin{aligned} &\Gamma_U^{(0)} = \Gamma_U^k \\ &\text{For } \ell = 1, 2, \dots, L \\ &\quad \widetilde{\Gamma}_U^{(\ell)} = \Gamma_U^{(\ell)} - \theta^{(\ell)} \nabla f(\Gamma_U^{(\ell)}) \\ &\quad \Gamma_U^{(\ell+1)} = \operatorname{prox}_{\theta^{(\ell)} g}(\widetilde{\Gamma}_U^{(\ell)}) \\ &\Gamma_U^{k+1} = \Gamma_U^{(L)}. \end{aligned} \quad (21)$$

Hereabove, $(\theta^{(\ell)})_{1 \leq \ell \leq L}$ is a sequence of positive stepsizes obtained through a suitable backtracking strategy, so that all iterates remain in the (open) domain of f. Moreover, prox denotes the proximity operator, for which a definition in the non convex case can be found in [44].

Hereafter, we provide the expression for the gradient of f and the proximity operator of g. First,

$$(\forall \Gamma_U \in S_N^{++}) \quad \nabla f(\Gamma_U) = U^{k+1}(U^{k+1})^\top - \Gamma_U^{-1}, \quad (22)$$

while it is not defined for non definite positive matrices of S_N. Second, due to the separability of function g, we have, for any $\theta > 0$ [45],

$$(\forall \Gamma_U \in S_N) \quad \operatorname{prox}_{\theta g}(\Gamma_U) = \left(\operatorname{prox}_{\theta g_{ij}}([\Gamma_U]_{ij})\right)_{1 \leq i,j \leq N}. \quad (23)$$

The expression for the proximity operator of each term g_{ij}, defined in (20), depends if $(i,j) \in \mathcal{E}_U$ or not.

For $(i,j) \in \overline{\mathcal{E}}_U$, g_{ij} is a convex, proper, lower semicontinuous function on $\mathbb{R}$. Its proximity operator is thus uniquely defined, and it reads:

$$\begin{aligned} (\forall \omega \in \mathbb{R}) \operatorname{prox}_{\theta g_{ij}}(\omega) &= \underset{\xi \in \mathbb{R}}{\operatorname{argmin}} \left(\frac{1}{2}(\xi - \omega)^2 + \theta(2\lambda_U |\xi| + \lambda_U \xi^2) \right) \\ &= \operatorname{prox}_{(1+2\lambda_U\theta)^{-1}|\cdot|} \left((1 + 2\lambda_U \theta)^{-1} \omega \right), \\ &= \mathcal{S}_{(1+2\lambda_U\theta)^{-1}} \left((1 + 2\lambda_U \theta)^{-1} \omega \right), \end{aligned} \quad (24)$$

with $\mathcal{S}_\tau$ the soft thresholding operator with parameter $\tau > 0$:

$$(\forall u \in \mathbb{R}) \quad \mathcal{S}_\tau(u) = \operatorname{sign}(u) \max(0, |u| - \tau). \quad (25)$$

For $(i,j) \in \mathcal{E}_U$, g_{ij} is non-convex. Its proximity operator reads:

$$(\forall \omega \in \mathbb{R}) \quad \operatorname{prox}_{\theta g_{ij}}(\omega) = \underset{\xi \in \mathbb{R}}{\operatorname{argmin}} \left(\frac{1}{2}(\xi - \omega)^2 + \theta(-2\lambda_U \ln(|\xi| + \delta) + \lambda_U \xi^2) \right), \quad (26)$$

$$= \operatorname{prox}_{\frac{2\theta\lambda_U}{1+2\theta\lambda_U}\zeta} \left(\frac{1}{1 + 2\theta\lambda_U} \omega \right), \quad (27)$$

with $\zeta : \omega \mapsto -\ln(|\omega| + \delta)$. Function ζ is non-convex. It is lower semi-continuous, and lower bounded by polynomial $\omega \mapsto -(\omega^2 + \delta)$, thus its proximity operator exists though it is not uniquely defined (i.e., it is set-valued). Let $\tau > 0$, and $\overline{\omega} \in \mathbb{R}$. Set $\varphi : \omega \mapsto \tau\zeta(\omega) + \frac{1}{2}(\omega - \overline{\omega})^2$. The operator $\operatorname{prox}_{\tau\zeta}$ evaluated at $\overline{\omega}$ is defined as the set of (global) minimizers of φ. Hereafter, we study φ, for positive or negative input values, so as to deduce the proximity set of $\tau\zeta$.

First, for every $\omega > 0$, the first and second order derivatives of φ read:

$$\varphi'(\omega) = -\frac{\tau}{\omega + \delta} + \omega - \overline{\omega}, \quad (28)$$

and

$$\varphi''(\omega) = \frac{\tau}{(\omega + \delta)^2} + 1. \quad (29)$$

Function φ' is strictly increasing on $]0,+\infty[$. Moreover, $\lim_{\omega\to 0^+}\varphi'(\omega) = -\frac{\tau}{\delta} - \overline{\omega}$. Thus, φ' cancels on $]0,+\infty[$ if $\overline{\omega} > -\frac{\tau}{\delta}$. Canceling (28) is equivalent to search the roots for the second order polynomial:

$$-\tau + (\omega - \overline{\omega})(\omega + \delta) = \omega^2 - (\overline{\omega} - \delta)\omega - \overline{\omega}\delta - \tau. \quad (30)$$

The discriminant of such polynomial is

$$\Delta^+ = (\overline{\omega} - \delta)^2 + 4(\overline{\omega}\delta + \tau) \quad (31)$$
$$= (\overline{\omega} + \delta)^2 + 4\tau, \quad (32)$$

which is strictly positive under the condition $\overline{\omega} > -\frac{\tau}{\delta}$. Two roots exist and read:

$$\omega_{1,2}^+ = \frac{(\overline{\omega} - \delta) \pm \sqrt{\Delta^+}}{2}. \quad (33)$$

There is a sole positive root, equals to:

$$\omega_2^+ = \frac{(\overline{\omega} - \delta) + \sqrt{(\overline{\omega} + \delta)^2 + 4\tau}}{2}. \quad (34)$$

Second, when $\omega \in]-\infty, 0[$, we have

$$(\forall \omega < 0) \quad \varphi'(\omega) = \frac{\tau}{-\omega + \delta} + \omega - \overline{\omega}, \quad (35)$$

and

$$(\forall \omega < 0) \quad \varphi''(\omega) = \frac{\tau}{(\omega - \delta)^2} + 1. \quad (36)$$

Hereagain, φ' is strictly increasing. Moreover, $\lim_{\omega\to 0^-}\varphi'(\omega) = \frac{\tau}{\delta} - \overline{\omega}$. Therefore, φ' cancels on $]-\infty, 0[$ if $\overline{\omega} < \frac{\tau}{\delta}$. Canceling (35) is equivalent to search the roots for the second order polynomial:

$$-\tau + (\omega - \overline{\omega})(\omega - \delta) = \omega^2 - (\overline{\omega} + \delta)\omega + \overline{\omega}\delta - \tau. \quad (37)$$

The discriminant of such polynomial is

$$\Delta^- = (\overline{\omega} + \delta)^2 - 4(\overline{\omega}\delta - \tau) = (\overline{\omega} - \delta)^2 + 4\tau, \quad (38)$$

which is strictly positive under the condition $\overline{\omega} < \frac{\tau}{\delta}$. Two roots exist and are given by:

$$\omega_{1,2}^- = \frac{(\overline{\omega} + \delta) \pm \sqrt{\Delta^-}}{2}. \quad (39)$$

There is a sole negative root, equals to

$$\omega_1^- = \frac{(\overline{\omega} + \delta) - \sqrt{(\overline{\omega} + \delta)^2 - 4(\overline{\omega}\delta - \tau)}}{2}. \quad (40)$$

Finally, studying the variations of φ on all $\mathbb{R}$ allows us to distinguish three cases:

- Case 1: $\overline{\omega} \in]-\frac{\tau}{\delta}, \frac{\tau}{\delta}[$. Then,

$$\text{prox}_{\tau\zeta}(\overline{\omega}) = \text{argmin}_{p\in\{\omega_1^-, \omega_2^+\}}\varphi(p). \quad (41)$$

- Case 2: $\overline{\omega} \leq -\frac{\tau}{\delta}$. Then,

$$\text{prox}_{\tau\zeta}(\overline{\omega}) = \omega_1^-. \quad (42)$$

- Case 3: $\overline{\omega} \geq \frac{\tau}{\delta}$. Then,

$$\text{prox}_{\tau\zeta}(\overline{\omega}) = \omega_2^+. \quad (43)$$

4) Summarized algorithm: The proposed Graph Regularized Probabilistic Matrix Factorization (GRPMF) method for DDI prediction task is summarized in Algorithm 1. Given the partially observed drug interaction matrix Y, we recall that we aim to recover the full drug interaction matrix R by exploiting the drug similarity information captured in A_U. Matrices X and Γ_U are initialized using identity matrices scaled by a positive value s^0. The first Z left-singular vectors of Y obtained using singular value decomposition (SVD) are considered for initialization of U. The three unknowns, namely X, U and Γ_U, are updated using the updates defined in the previous subsection, in an iterative manner for K iterations. The final matrix R is recovered using $R = UU^\top$.

Algorithm 1 GRPMF for DDI Prediction

Input: Y, A_U
Parameters: $Z, \sigma, s^0, \lambda_R, \lambda_U, K, L$.
Initialization: U^0 (using $\texttt{svd}(Y)$), $X^0 = \Gamma_U^0 = s^0 \times \mathbb{I}_N$.
Compute $\mathcal{E}_U$ and $\overline{\mathcal{E}_U}$ using (8) and (9).
for $k = 1, 2, \ldots, K$ iterations
  Update X^{k+1} using (13);
  Update U^{k+1} using (14);
  Update ${\Gamma_U}^{k+1}$ using (21).
end
Return: $R = U^K (U^K)^\top$.

V. Experimental Results

This section presents our experimental results illustrating the validity of the proposed method. We first introduce the dataset considered for DDI prediction along with the necessary data pre-processing steps carried out. Subsequently, comparison of the proposed GRPMF method against the benchmark algorithms, and an ablation study for GRPMF, are presented.

A. Dataset

We rely on the DDI data from Stanford University [46] that is approved by the U.S. Food and Drug Administration. It contains $48,514$ interactions from $1,514$ drugs extracted from drug labels and scientific publications [47], [48]. Our experiments rely on the publicly available drug dataset from DrugBank that gathers $14,315$ drugs along with their KEGG ID or compound ID. In particular, this latter dataset is used to build the expert knowledge for our regularization term. To do so, we rely on the SIMCOMP score that amounts to comparing the chemical structures of the drugs [41]. SIMCOMP computes the similarity of two chemical compounds by counting the number of matched atoms in those atom alignments, using the KEGG ID reference system. We computed the SIMCOMP score using the SIMCOMP search tool with a cutoff of 0.01 [49]. Due to the unavailability of KEGG ID, and thus SIMCOMP scores, for few drugs, and to the non overlap between [46] and [47] drug lists, the final data reduced to $N = 927$ drugs. For this subset of drugs, we made sure that they each have at least 10 known interactions with other drugs.

From our dataset curating, we end up with a binary matrix $Y \in \{0,1\}^{N\times N}$, where $y_{i,j} = 1$ if drugs i and j are known to interact. An entry $y_{i,j} = 0$ means that no interaction has been reported between drugs i and j so far. Our SIMCOMP

TABLE I: Quality metrics for all compared methods.

Method	AUPR	AUC	Precision	Recall	F1	Accuracy
Graph DDI (GNB)	0.0395	0.5	0.9349	0.9559	0.9406	0.9559
Graph DDI (LogR)	0.0385	0.5008	0.9176	0.9579	0.9373	0.9579
Graph DDI (RF)	0.0385	0.5004	0.9176	0.9579	0.9373	0.9579
KGNN (Sum)	0.1830	0.8255	0.9482	0.8867	0.9117	0.8867
KGNN (Concat)	0.1894	0.8954	0.9502	0.8995	0.9200	0.8995
KGNN (Neighbor)	0.1123	0.758	0.9344	0.8086	0.8613	0.8086
Conv-LSTM	0.0402	0.5006	0.9226	0.8932	0.9073	0.8932
GRMF (p = 2, Z = 50, λ_l = 0.05, $\lambda_d = \lambda_t$ = 0.3)	0.3632	0.8849	0.9558	0.8862	0.9134	0.8862
PMFG (p = 2, Z = 20, θ = 0.3, λ_U = 0, λ_R = 1, $\sigma = s^0 = \delta$ = 0.01)	0.3290	0.8667	0.9495	0.9604	0.9449	0.9604
GRPMF (p = 2, Z = 20, θ = 0.6, λ_U = 0.5, λ_R = 1, $\sigma = s^0 = \delta$ = 0.01)	**0.4636**	**0.9307**	**0.9612**	**0.9575**	**0.9590**	**0.9575**

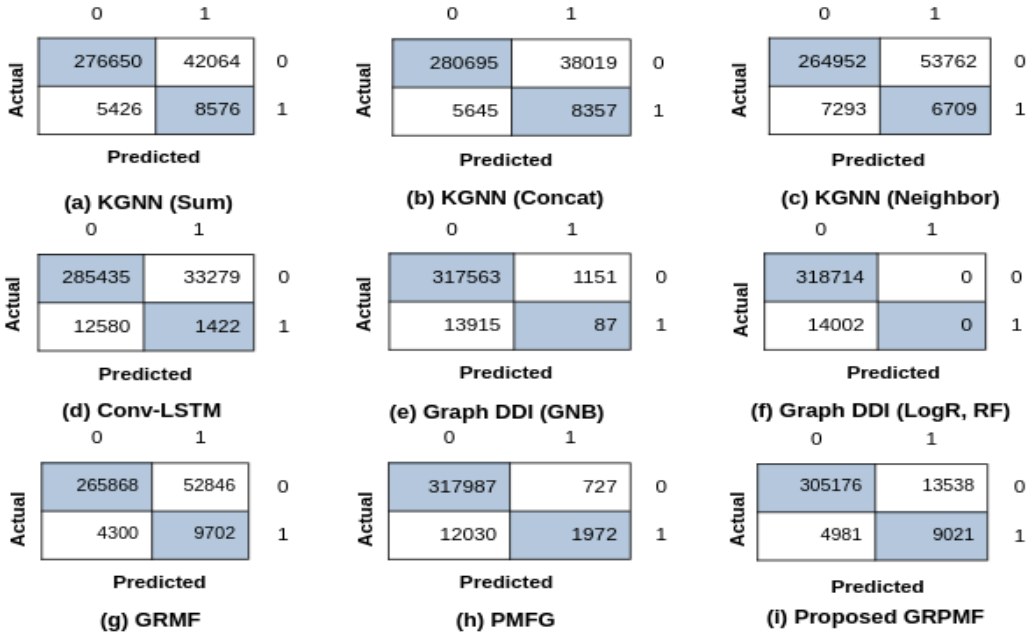


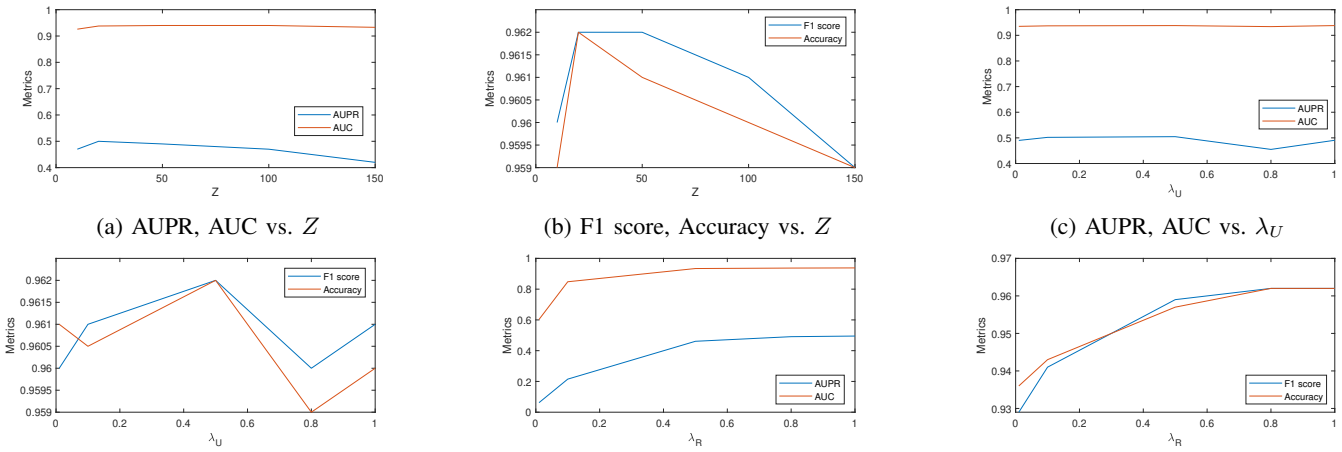
Fig. 1: Confusion matrices for all compared methods.

calculation yields a drug similarity data represented as a symmetric matrix $A_U \in \mathbb{R}^{N\times N}$. The computation for matrix A_U is only required once. On our computational device, it required about 12 hours for 927 drugs on this dataset. Note that the initial SIMCOMP matrix has been sparsified so as to retain only the p-nearest neighbours of each drug with the aim to preserve the local geometry of the original data. Such operation promotes that drugs that are close to one another in the original (chemical structure) space should also be close to one another in the learnt (drug interaction) manifold (i.e., local invariance assumption) [50]. We then compute sets $\mathcal{E}_U$, and $\overline{\mathcal{E}_U}$ using (8) and (9) with $\tau = 0$.

We then adopt a supervised learning paradigm, as in [30], [31], [32]. 20% drug pairs are considered for training (i.e., assumed to be observed), and the remaining 80% are taken for testing (i.e., masked and thus must be estimated). This means that the cardinality of the observed set $\mathcal{D}$ equals $20\%N^2$. Since the number of known interactions (i.e., 1 entries) are much less than unknown interactions (i.e., 0 entries) in matrix Y, we made sure that the training data contains at most 60% samples of known interactions for each drug, and the remaining are the samples from unknown interactions. The remaining known and unknown interactions for each drug are considered in the test data. The training and test data samples from each of the interaction categories per drug are selected randomly. Also, only one pair of interactions are kept that belong either from the upper triangle or the lower triangle of the DDI matrix. The test data comprises of 3,32,716 samples or interactions. Out of this, 3,18,714 samples correspond to class 0 (i.e., no observed drug interaction) and 14,002 correspond to class 1 (i.e., observed drug interaction) that makes the problem very challenging.

B. GRPMF and benchmarks settings

In all our experiments, GRPMF is applied with the iteration numbers $K = 10$ and $L = 5$, that appeared enough to reach stability for the inner and outer loops of our algorithm. As a post-processing, the interaction matrix R recovered using GRPMF is processed by removing the diagonal entries that indicate self interactions (interaction of the drug with itself), and are not of interest in this application. The class labels are assigned using a simple thresholding operation, such that if $|R_{ij}| \leq 0.5$, decision is class 0, otherwise this is class 1.



(a) AUPR, AUC vs. Z (b) F1 score, Accuracy vs. Z (c) AUPR, AUC vs. λ_U

(d) F1 score, Accuracy vs. λ_U (e) AUPR, AUC vs. λ_R (f) F1 score, Accuracy vs. λ_R

Fig. 2: Performance of GRPMF with different values of Z, λ_U and λ_R

The performance of the proposed method is compared against the following benchmark algorithms for DDI prediction:

- **Graph DDI [30]**: In this work, the KGs are constructed using different types of embeddings namely, RDF2Vec, TransE and TransD. These embeddings, along with the DDI matrix, are fed one by one to different machine learning techniques like Random Decision Forest (RF), Gaussian Naive Bayes (GNB), and Logistic Regression (LogR) for DDI prediction. Here, only the RDF2Vec embedding vectors with Skip-Gram is retained for comparison as it achieved the best performance in the study [30].
- **KGNN [31]**: Here, the KG results from the drugs' topological structures for DDI prediction. The KG and DDI matrix is fed to Graph Neural Network (GNN). This method focuses on drug neighborhood sampling and aggregates the entities to represent the drugs' potential neighbors in three different ways: (i) Sum, (ii) Concat, and (iii) Neighbor.
- **Conv-LSTM [32]**: In this method, KG is learned using different embeddings (e.g., ComplEx, TransE, RDF2Vec, etc.) and fed as input to the CNN and LSTM along with the input DDI matrix, to perform DDI prediction. Only ComplEx embedding is considered here for comparison as it gave the best results in [32].
- Additionally, comparisons with Graph Regularized Matrix Factorization **GRMF** [50] and **PMFG** (7) techniques are presented. The former uses the SIMCOMP drug similarity as an expert knowledge term but, in contrast with our method, it does not infer any graph (i.e., the graph is imposed from the beginning by the user). The method PMFG has been described in the beginning of the paper. Its results are generated using (10) with $\lambda_U = 0$, using the same post-processing procedure as in our method.

The hyperparameters for GRMF, PMFG and the proposed GRPMF method are tuned using grid search. The other techniques make use of the hyperparameter values mentioned in their respective works.

Two main performance metrics are used to evaluate the models, namely the Area under the ROC curve (AUC) and the Area under the PR curve (AUPR). In addition, weighted Precision, weighted Recall, weighted F1 score and Accuracy are calculated for the prediction results.

C. Comparison with benchmarks

Table I summarizes the average values of the performance metrics obtained with the benchmark techniques and the proposed method on the test data obtained using 5 randomly generated train-test splits. Since we have considered $\approx 20\%$ data for training and the rest for testing, the AUPR of all the models, especially the benchmark models employing deep learning and machine learning models are low. It can be seen from the table that among the learning based models, KGNN (Concat) performs best in terms of AUPR and AUC. The remaining learning based methods have AUC ≈ 0.5, indicating no discrimination capability between the classes. They may require more training data for improved performance. The methods based on matrix factorization work much better in this case. Among them, the proposed GRPMF method achieves the best performance across all metrics, with a considerable increase in the AUPR and AUC metrics, respectively, over KGNN (Concat). Additionally, it provides $\approx 28\%$ and 40% improvement in AUPR, and $\approx 5\%$ and 7% improvement in AUC, with respect to GRMF and PMFG respectively.

The high AUPR score of GRPMF indicates that the extra terms in the formulation (10) compared to PMFG (7) are able to learn the drug-drug interactions in an effective manner.

Due to the important class imbalance, the values of accuracy and the weighted metrics, namely, Precision, Recall, and F1 score (given in Table I), are more biased towards the majority class (i.e., class 0 in this case). This is the reason for the difference between the AUPR and AUC metrics compared to all other metrics. Confusion matrices for one test set of all the methods are presented in Fig. 1 to provide additional insights on the prediction results. It can be observed that,

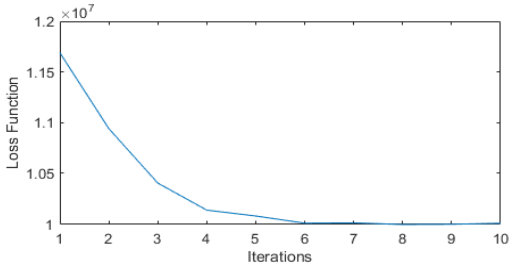


Fig. 3: Evolution of F along iterations, for GRPMF method.

while KGNN (Figs. 1 (a)-(c)) performs well in predicting the true interactions, the number of false interactions prediction is also high. The predictions of both Conv-LSTM (Fig. 1 (d)) and Graph DDI (GNB) (Fig. 1 (e)) have a low count of true drug interactions. While the other variants, LogR and RF of Graph DDI (Fig. 1 (f)) are not able to predict the true drug interactions. On the other hand, MF-based techniques appear to have better performance. The number of true positives for the observed drug interactions are more incase of GRMF (Fig. 1 (h)) compared to PMFG (Fig. 1 (g)) and hence the former has a higher AUPR and AUC score. While PMFG reports a high number of false negatives, GRMF has more false positives, for known drug interactions. Overall the proposed GRPMF method (Fig. 1 (i)) reaches the best performance compared to all other methods. The number of true positives are high and false positives are low, both for known and unknown interactions compared to other benchmark methods.

D. Ablation Study

Three main hyperparameters govern the performance of the proposed GRPMF method: (i) latent dimension Z, (ii) expert prior weight λ_U, and (iii) data fidelity weight λ_R. We now study the effect of different values of these hyperparameters on the performance of the proposed GRPMF by varying one of the hyperparameters while keeping the others fixed. Figs. 2 (a) and (b) present the plots of all the performance metrics for different values of the Z. We observe that AUPR metric is high for $Z = 20$ while the other metrics are comparable for other values of Z. Similarly, from Figs. 2 (c) and (d), we observe that $\lambda_U = 0.5$ seems to be an optimal choice across all metrics. Figs. 2 (e) and (f) present the performance variation with the change in λ_R. Since λ_R controls the equality constraint $X = UU^\top$, and as a consequence, the fidelity to the observed data Y, the performance is observed to increase with λ_R, with best value $\lambda_R = 1$. The best values for these hyperparameters were considered for generating the results for GRPMF method in Table I. Finally, Fig. 3 displays the plot of the evolution of the GRPMF loss function F along the iterations for the proposed GRPMF method with the retained settings. It can be seen that F decreases monotonically as expected, and that it reaches stability within a few iterations.

VI. Conclusion

This paper presents GRPMF, a novel matrix factorization framework for DDI prediction. Through the introduction of an original regularization strategy, our approach encodes efficiently prior expert knowledge so as to perform jointly DDI prediction and graph similarity inference. Our experimental results obtained with the DrugBank dataset demonstrates the superior performance of our approach compared to deep learning and machine learning-based methods. Although the expert knowledge prior in our experiments relied on the SIMCOMP score for drug similarity, other drug similarity features could be considered in future work. Another extension would be to incorporate an explicit constraint favoring binary entries in the sought matrix, which appeared to be beneficial in DDI, and more generally bioinformatics studies [51]. A possible avenue would be to peruse our recent formulation [52], by introducing probabilistic graph modeling terms into it.

References

[1] S. Vilar, C. Friedman, and G. Hripcsak, "Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media," *Briefings in Bioinformatics*, vol. 19, pp. 863—-877, 2018.

[2] T. Roblek, T. Vaupotic, A. Mrhar, and M. Lainscak, "Drug-drug interaction software in clinical practice: a systematic review," *European journal of clinical pharmacology*, vol. 71(2), pp. 131–142, 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0219796

[3] S. Vilar, E. Uriarte, L. Santana, T. Lorberbaum, G. Hripcsak, C. Friedman, and N. Tatonetti, "Similarity-based modeling in large-scale prediction of drug-drug interactions," *Nat Protocol*, vol. 9(9), p. 2147–63, 2014.

[4] H. Luo, P. Zhang, H. Huang, J. Huang, E. Kao, L. Shi, L. He, and L. Yang, "Ddi-cpi, a server that predicts drug-drug interactions through implementing the chemical-protein interactome," *Nucleic Acids Research*, vol. 42(Web Server issue), p. 46–52, 2014.

[5] J.-Y. Shi, H. Huang, J.-X. Li, P. Lei, Z. Y-N., and S.-M. Yiu, "Predicting comprehensive drug-drug interactions for new drugs via triple matrix factorization," *Lecture Notes in Computer Science: Bioinformatics and Biomedical Engineering*, p. 108–17, 2017.

[6] Y. Zhang, H. Wu, and J. e. a. Xu, "Leveraging syntactic and semantic graph kernels to extract pharmacokinetic drug drug interactions from biomedical literature," *BMC Systems Biology*, 2016.

[7] J. Duke, X. Han, Z. Wang, A. Subhadarshini, S. Karnik, X. Li, S. Hall, Y. Jin, J. Callaghan, and M. e. a. Overhage, "Literature based drug interaction prediction with clinical assessment using electronic medical records: novel myopathy associated drug interactions," *PLoS Compututational Biology*, 2012.

[8] "U.s. food and drug administration," http://www.fda.gov.

[9] K. Han, P. Cao, Y. Wang, F. Xie, J. Ma, M. Yu, J. Wang, Y. Xu, Y. Zhang, and J. Wan, "A review of approaches for predicting drug–drug interactions based on machine learning," *Frontiers in Pharmacology*, vol. 12, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fphar.2021.814858

[10] J. Strahl, J. Peltonen, H. Mamitsuka, and S. Kaski, "Scalable probabilistic matrix factorization with graph-based priors," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 5851–5858. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/6043

[11] Z. Z. Cheng F, "Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties," in *Journal of the American Medical Informatics Association*, 2014.

[12] E. C. Rohani N, "Drug-drug interaction predicting by neural network using integrated similarity," in *Nature Scientific Reports*, vol. 9(1), 2019.

[13] C. Yan, G. Duan, Y. Zhang, F.-X. Wu, Y. Pan, and J. Wang, "Predicting drug-drug interactions based on integrated similarity and semi-supervised learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 168–179, 2022.

[14] A. Kastrin, P. Ferk, and B. Leskošek, "Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning," *PLOS ONE*, vol. 13, pp. 1–23, 05 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0196865

[15] G. Shtar, L. Rokach, and B. Shapira, "Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures," *PLOS ONE*, vol. 14, pp. 1–21, 08 2019. [Online]. Available: https://doi.org/10.1371/journal.pone.0219796

[16] B. Xu, X. Shi, Z. Zhao, and W. Zheng, "Leveraging biomedical resources in bi-lstm for drug-drug interaction extraction," *IEEE Access*, vol. 6, pp. 33 432–33 439, 2018.

[17] P. K. Shukla, P. K. Shukla, P. Sharmau, P. Rawat, J. Samar, R. Moriwal, and M. Kaur, "Efficient prediction of drug–drug interaction using deep learning models," *IET Systems Biology*, vol. 14, pp. 211–216(5), August 2020.

[18] N. Liu, C.-B. Chen, and S. Kumara, "Semi-supervised learning algorithm for identifying high-priority drug–drug interactions through adverse event reports," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 57–68, 2020.

[19] G. Lee, C. Park, and J. Ahn, "Novel deep learning model for more accurate prediction of drug-drug interaction effects," *BMC Bioinformatics*, vol. 20, 2019. [Online]. Available: https://doi.org/10.1186/s12859-019-3013-0

[20] T. H. Vo, N. T. K. Nguyen, Q. H. Kha, and N. Q. K. Le, "On the road to explainable ai in drug-drug interactions prediction: A systematic review," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 2112–2123, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2001037022001386

[21] K. Schwarz, A. Allam, N. Perez Gonzalez, and M. Krauthammer, "Attentionddi: Siamese attention-based deep learning method for drug–drug interaction predictions," *BMC Bioinformatics*, vol. 22, no. 412, 2021. [Online]. Available: https://doi.org/10.1186/s12859-021-04325-y

[22] D. Pan, L. Quan, Z. Jin, T. Chen, X. Wang, J. Xie, T. Wu, and Q. Lyu, "Multisource attention-mechanism-based encoder–decoder model for predicting drug–drug interaction events," *Journal of Chemical Information and Modeling*, vol. 62, no. 23, pp. 6258–6270, 2022, pMID: 36449561. [Online]. Available: https://doi.org/10.1021/acs.jcim.2c01112

[23] S. Liu, Y. Zhang, Y. Cui, Y. Qiu, Y. Deng, Z. M. Zhang, and W. Zhang, "Enhancing drug-drug interaction prediction using deep attention neural networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–1, 2022.

[24] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Application of dimensionality reduction in recommender systems: A case study," in *WebKDD Workshop at the ACM SIGKKD*, 2000.

[25] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[26] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems*, vol. 20, 2008.

[27] H. Yu, K. Mao, and J. e. a. Shi, "Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization," *BMC Systems Biology*, vol. 12, 2018.

[28] J. Zhu, Y. Liu, Y. Zhang, and D. Li, "Attribute supervised probabilistic dependent matrix tri-factorization model for the prediction of adverse drug-drug interaction," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2820–2832, 2021.

[29] W. Zhang, Y. Chen, D. Li, and X. Yue, "Manifold regularized matrix factorization for drug-drug interaction prediction," *Journal of Biomedical Informatics*, vol. 88, pp. 90–97, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046418302144

[30] R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, and M. Dumontier, "Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings," *BMC Bioinformatics*, vol. 20, p. 726, 2019.

[31] X. Lin, Z. Quan, Z.-J. Wang, T. Ma, and X. Zeng, "Kgnn: Knowledge graph neural network for drug-drug interaction prediction," in *International Joint Conferences on Artificial Intelligence Organization (IJCAI-20)*, 7 2020, pp. 2739–2745.

[32] M. Rezaul Karim, M. Cochez, J. Jares, M. Uddin, O. Beyan, and S. Decker, "Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-lstm network," in *ACM-BCB 2019 - Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, Sep. 2019, pp. 113–123.

[33] Y. Koren, R. M. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009. [Online]. Available: https://doi.org/10.1109/MC.2009.263

[34] G. R. Naik, *Non-negative Matrix Factorization Techniques: Advances in Theory and Applications*. Springer, 2016.

[35] B. Recht, "A simpler approach to matrix completion," *J. Mach. Learn. Res.*, vol. 12, no. null, p. 3413–3430, dec 2011.

[36] Q. Gu, J. Zhou, and C. Ding, "Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs," in *Proceedings of the 2010 SIAM International Conference on Data Mining (SDM)*, 2010, pp. 199–210. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611972801.18

[37] A. Mongia and A. Majumdar, "Matrix completion on multiple graphs: Application in collaborative filtering," *Signal Process.*, vol. 165, no. C, p. 144–148, dec 2019. [Online]. Available: https://doi.org/10.1016/j.sigpro.2019.07.002

[38] A. Mongia, V. Jain, E. Chouzenoux, and A. Majumdar, "Deep latent factor model for predicting drug target interactions," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1254–1258.

[39] A. Mongia, S. Jain, E. Chouzenoux, and A. Majumdar, "Deepvir: Graphical deep matrix factorization for in silico antiviral repositioning-application to covid-19," *Journal of Computational Biology*, vol. 29, pp. 441–452, May 2022. [Online]. Available: http://doi.org/10.1089/cmb.2021.0108

[40] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems*, vol. 20, 2008.

[41] M. Hattori, N. Tanaka, M. Kanehisa, and S. Goto, "Simcomp/subcomp: chemical structure search servers for network analyses," *Nucleic Acids Research.*, vol. 38, 2010.

[42] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Carnegie Mellon University, USA, Tech. Rep., 1994.

[43] X. Duan, J. Li, Q. Wang, and X. Zhang, "Low rank approximation of the symmetric positive semidefinite matrix," *Journal of Computational and Applied Mathematics*, vol. 260, pp. 236–243, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377042713005359

[44] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Mathematical Programming, Series A*, vol. 137, no. 1, pp. 91–124, Aug. 2011. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00790042

[45] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2017. [Online]. Available: https://hal.sorbonne-universite.fr/hal-01517477

[46] "Drug-drug interaction network," http://snap.stanford.edu/biodata/datasets/10001/10001-ChCh-Miner.html.

[47] D. Wishart, Y. Feunang, A. Guo, E. Lo, A. Marcu, J. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, and M. Wilson, "Drugbank 5.0: a major update to the drugbank database for 2018," *Nucleic Acids Research*, 2017.

[48] "Miner: Gigascale multimodal biological network," https://github.com/snap-stanford/miner-data, 2017.

[49] "Simcomp search," https://www.genome.jp/tools/simcomp/.

[50] A. Ezzat, P. Zhao, M. Wu, X.-L. Li, and C.-K. Kwoh, "Drug-target interaction prediction with graph regularized matrix factorization," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 14, no. 3, pp. 646–656, 2016.

[51] M. Davenport, Y. Plan, E. Berg, and M. Wootters, "1-bit matrix completion," *Information and Inference*, vol. 3, 09 2012.

[52] D. Talwar, A. Mongia, E. Chouzenoux, and A. Majumdar, "Binary matrix completion on graphs: Application to collaborative filtering," *Digital Signal Processing*, vol. 122, p. 103350, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200421003894