Drug-Target Interaction Prediction with Graph Regularized Matrix Factorization

Ali Ezzat, Peilin Zhao, Min Wu, Xiao-Li Li, and Chee-Keong Kwoh

Abstract—Experimental determination of drug-target interactions is expensive and time-consuming. Therefore, there is a continuous demand for more accurate predictions of interactions using computational techniques. Algorithms have been devised to infer novel interactions on a global scale where the input to these algorithms is a drug-target network (i.e. a bipartite graph where edges connect pairs of drugs and targets that are known to interact). However, these algorithms had difficulty predicting interactions involving *new* drugs or targets for which there are no known interactions (i.e. "orphan" nodes in the network). Since data usually lie on or near to low-dimensional non-linear manifolds, we propose two matrix factorization methods that use graph regularization in order to learn such manifolds. In addition, considering that many of the non-occurring edges in the network are actually unknown or missing cases, we developed a preprocessing step to enhance predictions in the "new drug" and "new target" cases by adding edges with intermediate interaction likelihood scores. In our cross validation experiments, our methods achieved better results than three other state-of-the-art methods in most cases. Finally, we simulated some "new drug" and "new target" cases and found that GRMF predicted the left-out interactions reasonably well.

Index Terms—Drug-target interaction prediction, matrix factorization, graph regularization, manifold learning.

1 Introduction

Drug development is a time-consuming and expensive process that is plagued with the problem known as the high attrition rate [1]. This led to the practitioners' great interest in drug repositioning (reusing already available drugs for new indications) due to its potential to reduce the time, cost, risk and effort inherent in developing new drugs [2]. Of great importance to drug repositioning efforts are online biological databases that store and maintain information on already known drugs and drug-target interactions; examples of such databases include KEGG [3], DrugBank [4], ChEMBL [5] and STITCH [6]. However, aside from those interactions that are stored in online databases, much more interactions still remain to be discovered, which motivated the development of a wide variety of computational techniques that predict new drug-target interactions. Such computational techniques help predict, with reasonable confidence, new undiscovered interactions for further experimental investigation and confirmation, which thus greatly facilitates the drug development process.

From these computational techniques, this paper is interested in *global-scale drug-target interaction prediction* which can benefit the drug development process by identifying previously unknown targets for known drugs as well as off-targets that may lead to undesired side-effects [7]. A recent overview of such methods is provided in [8].

The earlier methods for drug-target interaction prediction used to involve docking simulations [9] or ligand-based approaches [10]. However, there are disadvantages to each of these techniques: for docking simulations, the 3D structure of the target protein must be available, which is often not the case; for ligand-based approaches, a problem arises if few or no ligands are known for the target protein.

Due to these difficulties, more attention has been given to the so-called *chemogenomic* approaches where information from both the drug and target sides are used simultaneously to improve predictions. One of the pioneering chemogenomic methods, [11], computed the pairwise chemical structure similarities between drugs as well as the pairwise genomic sequence similarities between targets (proteins). The computed pairwise similarities, along with a given drug-target network (where interacting drugs and targets are connected with edges), were then used as inputs to a kernel regression-based method to infer new interactions.

Using the same inputs, a *bipartite local model* that uses support vector machines was proposed to predict drug-target interactions [12]. For each drug-target pair, two models are trained independently, one from the drug side and the other from the target side. The two models are then used to give two predictions from which the final prediction result is obtained by an aggregating function. Afterwards, another bipartite local model that uses regularized least squares (RLS_{avg}) was proposed as well as another method (RLS_{kron}) where the two models from the drug and target sides were merged into one by taking the Kronecker product of the drug and target similarity matrices [13].

Moreover, other researchers started addressing the problem of predicting interactions for *new* drugs or targets for which there are no known interactions [14]. They proposed NII (Neighbor-based Interaction-profile Inferring) which reinforces the training procedure for new drugs or targets by deriving temporary interaction profiles for them.

- *Ali Ezzat and Chee-Keong Kwoh are with the School of Computer Engineering, Nanyang Technological University, Singapore 639798. Emails: aly002@ntu.edu.sg, asckkwoh@ntu.edu.sg*
- *Peilin Zhao, Min Wu and Xiao-Li Li are with the Institute for Infocomm Research (I^2R), A*Star, Singapore 138632. Emails: zhaop@i2r.a-star.edu.sg, wumin@i2r.a-star.edu.sg, xlli@i2r.a-star.edu.sg*

Manuscript received August 23, 2015; revised January 6, 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2016.2530062, IEEE/ACM Transactions on Computational Biology and Bioinformatics

The base prediction algorithm (RLS_{avg} in this case) would then be able to give better prediction results. Shortly after NII, a similar procedure called WNN (Weighted Nearest Neighbors) having the same goal as NII was proposed to augment RLS_{kron} [15]. Results for methods using either NII or WNN have shown that such preprocessing steps are indeed beneficial to the prediction results.

Matrix factorization techniques have also been used to predict interactions recently. Such techniques decompose the matrix representing the drug-target network into multiple low-rank matrices consisting of latent (or hidden) features that are assumed to govern the drug-target interactions. Two examples of such techniques are a Bayesian matrix factorization method, KBMF2K [16], and a collaborative matrix factorization method, CMF [17].

We observe that how to improve the prediction results for *new* drugs or targets is a key task in drug discovery. However, it is a challenging task as new drugs do not have known interactions with any targets, and new targets do not have known interactions with any drugs. While some existing methods have been proposed to improve the prediction performance for new drugs or targets, their results show that there is still room for improvement [14], [15], [17].

Based on our observation that many of the missing edges are actually unknown interactions, we designed a preprocessing step that adds edges with intermediate interaction likelihood scores to assist with prediction. Additional motivation for this preprocessing step was provided by [18] where the importance of nearest-neighbor information in drug-target interaction prediction was emphasized – the authors were interested in predicting ligands for orphan targets, and their experiments have shown that ligands of orphan targets could be predicted with reasonable accuracy whenever close neighbors (with known ligands) of those orphan targets were available, regardless of the kernels used for ligands or targets. Note that, like ligands, many drugs are small molecules, so this applies to drug-target interactions as well.

After running the preprocessing step, we use matrix factorization to predict drug-target interactions. However, unlike CMF and KBMF2K, we use *graph regularization* to prevent overfitting. In graph regularization, the similarity matrices are sparsified beforehand by keeping only the similarity values to the nearest neighbors for each drug/target. By doing so, graph regularization is able to learn a manifold on which (or near to which) the data are assumed to lie. Since it was shown in previous work (e.g. [19], [20], [21]) that data usually lies on (or near to) a manifold, learning such a manifold is expected to give more accurate results.

To evaluate our proposed method, we used cross validation to compare it with three other state-of-the-art methods, namely BLM-NII [14], RLS-WNN [15] and CMF [17]. In addition, we computationally simulated a *new target* case and a *new drug* case (by leaving their respective interactions out), and tested one of our proposed methods on these cases to investigate its ability to predict the left-out interactions.

The remainder of this paper is organized as follows. The datasets used in our work are described in Section 2 as well as the notations that are used throughout the rest of the paper. Section 3 includes brief descriptions of three competing state-of-the-art methods, followed by Section 4 which describes our proposed methods. We then display the experimental results of our work and provide relevant discussion in Section 5. Finally, we end with a conclusion in Section 6.

TABLE 1: Drugs, targets and interactions in each dataset

Datasets	NR	GPCR	IC	E
Drugs	54	223	210	445
Targets	26	95	204	664
Interactions	90	635	1476	2926

2 DATA

We use the same four datasets introduced in [11] which correspond to four different target protein types, namely *nuclear receptors* (NR), *G protein-coupled receptors* (GPCR), *ion channels* (IC) and *enzymes* (E). Table 1 contains some simple statistics for the four datasets.

Each dataset contains three matrices: $Y \in \mathbb{R}^{n \times m}$, $S^d \in \mathbb{R}^{n \times n}$ and $S^t \in \mathbb{R}^{m \times m}$. The matrix Y is the adjacency matrix encoding the drug-target interactions with n drugs as rows and m targets as columns, where Y_{ij} is 1 if drug d_i and target t_j are known to interact and 0 otherwise. The matrix S^d represents the drug pairwise chemical structure similarities and the matrix S^t denotes the target pairwise genomic sequence similarities. Drug similarities were obtained using SIMCOMP [22], and the target similarities are normalized Smith-Waterman scores [23].

3 RELATED WORK

3.1 BLM-NII

BLM-NII is the method presented in [14] for predicting drug-target interactions. It uses a bipartite local model (RLS_{avg} [13]) as the base algorithm and augments it with NII which derives temporary interaction profiles for new drugs or targets to assist with the prediction.

For each drug-target pair $(d_i, t_j) \in Y$, if the drug d_i is new (i.e. has no known interactions with any target), NII is used to infer its profile by considering its chemical similarity to all other drugs. The interaction profile for a new drug d_i (i^{th} row vector of Y) is defined as

$$Y(d_i) = \sum_{p=1}^{n} S^d(d_i, d_p) Y(d_p) \quad (1)$$

which is then normalized via min-max normalization as follows

$$Y(d_i) = \frac{Y(d_i) - min(Y(d_i))}{max(Y(d_i)) - min(Y(d_i))}. \quad (2)$$

The GIP (Gaussian interaction profile) kernel is then used to obtain a drug network similarity matrix from the drug profiles. The network similarity between two drugs d_1 and d_2 is computed as $exp(-\gamma \|Y(d_1) - Y(d_2)\|^2)$ where γ is a parameter. The drug network similarity matrix is then

linearly combined with the drug chemical similarity matrix S^d to give the final drug similarity matrix $\tilde{S}^d$. To get the prediction from the target side, a least squares solution is obtained as

$$\hat{Y}(t_j) = \tilde{S}^d(\tilde{S}^d + \sigma I)^{-1}Y(t_j) \quad (3)$$

where σ is a (Tikhonov) regularization parameter and $Y(t_j)$ is the j^{th} column vector of Y.

After a prediction from the drug side is similarly obtained, predictions from the drug and target sides are combined using an aggregating function to give the final prediction as

$$p_{ij} = g(p^d_{ij}, p^t_{ij}) \quad (4)$$

where p_{ij} is the final prediction score of the drug-target pair involving drug d_i and target t_j. In [14], the aggregating function was $g = max$. However, from our in-house experiments, we found avg (i.e. $p_{ij} = (p^d_{ij} + p^t_{ij})/2$) to give better results, which is what we used to obtain the results reported in this paper.

3.2 RLS-WNN

We refer to the work presented in [15] as RLS-WNN. It uses RLS_{kron} from [13] as its base algorithm and augments it with WNN, a procedure that is similar to and has the same goal as NII.

For every new drug d, WNN is used to infer its interaction profile as

$$Y(d) = \sum_{i=1}^{n} w_i Y(d_i) \quad (5)$$

where d_1 to d_n are the drugs sorted in descending order based on their similarity to d, and $w_i = \eta^{i-1}$ where η is a decay term with $\eta < 1$. Note that, alternatively, every new target t may have its interaction profile inferred by WNN as

$$Y(t) = \sum_{j=1}^{m} w_j Y(t_j). \quad (6)$$

After all new drugs (or targets) are updated by WNN, the GIP kernel is used to ultimately get the final drug and target similarity matrices $\tilde{S}^d$ and $\tilde{S}^t$. A least squares solution to get the predictions is obtained as

$$vec(\hat{Y}) = K(K + \sigma I)^{-1}vec(Y) \quad (7)$$

where $vec(Y)$ is a column vector containing all the drug-target pairs, σ is a (Tikhonov) regularization parameter, and $K = \tilde{S}^d \otimes \tilde{S}^t$ is a kernel over drug-target pairs that was obtained by taking the Kronecker product of $\tilde{S}^d$ and $\tilde{S}^t$.

3.3 CMF

CMF is a matrix factorization method presented in [17] which minimizes the objective function

$$\begin{aligned} \min_{A,B} \quad & \|W \odot (Y - AB^T)\|^2_F \\ & + \lambda_l(\|A\|^2_F + \|B\|^2_F) \\ & + \lambda_d\|S^d - AA^T\|^2_F \\ & + \lambda_t\|S^t - BB^T\|^2_F \end{aligned} \quad (8)$$

where $\odot$ is the element-wise product, and $W \in \mathbb{R}^{n\times m}$ is a weight matrix such that $W_{ij} = 0$ if Y_{ij} is unknown and $W_{ij} = 1$ otherwise.

The first line is the WLRA (Weighted Low-Rank Approximation) term that tries to find the latent feature matrices A and B that reconstruct Y. The second line is the Tikhonov regularization term. The third and fourth lines are regularization terms that require latent feature vectors of similar drugs/targets to be similar *and* latent feature vectors of dissimilar drugs/targets to be dissimilar, respectively.

Denoting the objective function in Equation 8 as L and letting a_i and b_j be the i^{th} and j^{th} row vectors of A and B, respectively, two alternative update rules (one for updating a_i and one for updating b_j) were derived by setting $\frac{\partial L}{\partial a_i} = 0$ and $\frac{\partial L}{\partial b_j} = 0$. Alternating least squares is then used to run the update rules alternatingly until convergence. Finally, the predicted matrix for drug-target interactions is then obtained by multiplying A and B.

4 METHODS

Here, we present our proposed method for tackling the drug-target interaction prediction problem. It consists of two steps:

(i) WKNKN (weighted K nearest known neighbors), a preprocessing step that transforms the binary values in the given drug-target matrix, Y, into interaction likelihood values;

(ii) GRMF (graph regularized matrix factorization), a matrix factorization technique for predicting drug-target interactions. A variant of GRMF called WGRMF (weighted GRMF) is also proposed.

4.1 Weighted K nearest known neighbors (WKNKN)

The given drug-target matrix $Y \in \mathbb{R}^{n\times m}$ has n drug rows and m target columns. The i^{th} row in Y, denoted as $Y(d_i)$, is the interaction profile for drug d_i. Similarly, the j^{th} column in Y, denoted as $Y(t_j)$, is the interaction profile for target t_j. A drug (or target) being *known* means that it has at least one interaction in its profile, while it being *new* means that it has no interactions in its profile. Many of the non-interactions (or 0's) in Y are unknown cases that could potentially be true interactions (i.e. they are false negatives). Thus, we propose WKNKN as a preprocessing step to estimate the interaction likelihoods for these unknown cases based on their *known* neighbors. That is, assuming Y_{ij} equals 0, WKNKN replaces it by a continuous value in the range 0 to 1 with the following three steps:

1) Horizontal Direction Update: Get weighted average of the corresponding values in the profiles of the K known drugs nearest to d_i (the weights are the similarities of d_i to these nearest neighbors).
2) Vertical Direction Update: Get weighted average of the corresponding values in the profiles of the K known targets nearest to t_j (the weights are the similarities of t_j to these nearest neighbors).
3) Final update: Replace the $Y_{ij} = 0$ by taking the average of the above two values, representing the overall likelihood that d_i and t_j interact.

function WKNKN (Y, S^d, S^t, K, η)
  $Y_d = Y_t = 0$ ▷ initialize two matrices
  for $d \leftarrow 1$ **to** n **do**
    $dnn = KNearestKnownNeighbors(d, S^d, K)$
    for $i \leftarrow 1$ **to** K **do**
      $w_i = \eta^{i-1} S^d(d, dnn_i)$
    end for
    $Z_d = \sum_{i=1}^{K} S^d(d, dnn_i)$ ▷ normalization term
    $Y_d(d) = \frac{1}{Z_d} \sum_{i=1}^{K} w_i Y(dnn_i)$
  end for
  for $t \leftarrow 1$ **to** m **do**
    $tnn = KNearestKnownNeighbors(t, S^t, K)$
    for $j \leftarrow 1$ **to** K **do**
      $w_j = \eta^{j-1} S^t(t, tnn_j)$
    end for
    $Z_t = \sum_{j=1}^{K} S^t(t, tnn_j)$ ▷ normalization term
    $Y_t(t) = \frac{1}{Z_t} \sum_{j=1}^{K} w_j Y(tnn_j)$
  end for
  $Y_{dt} = (Y_d + Y_t)/2$
  $Y = max(Y, Y_{dt})$
  return Y
end function

Fig. 1: WKNKN algorithm

Figure 1 contains the pseudocode that describes the above procedure in detail. η is a decay term where $\eta \leq 1$, and $KNearestKnownNeighbors()$ returns the K nearest known neighbors in descending order based on their similarities to d_i or t_j.

Note that to infer the interaction likelihood for drug-target pairs, WKNKN does not simply use the K nearest neighbors; it uses the K nearest *known* neighbors, which is reasonable since *known* neighbors, having additional interaction information, would contribute more than *new* neighbors, whose interaction profiles are all 0's.

4.2 Graph-regularized matrix factorization (GRMF)

4.2.1 Sparsification of the similarity matrices

Sparsification of similarity matrices is a technique that has been used before with graph regularization [24]. In this work, we derived a p-nearest neighbor graph from each of the drug and target similarity matrices, S^d and S^t. That is, given the drug similarity matrix S^d, a p-nearest neighbor graph N is generated as

$$\forall i, j, \qquad N_{ij} = \begin{cases} 1 & j \in \mathcal{N}_p(i) \ \ \& \ \ i \in \mathcal{N}_p(j) \\ 0 & j \notin \mathcal{N}_p(i) \ \ \& \ \ i \notin \mathcal{N}_p(j) \\ 0.5 & otherwise \end{cases} \tag{9}$$

where $\mathcal{N}_p(i)$ is the set of p nearest neighbors to drug d_i. N is then used to sparsify the similarity matrix S^d as

$$\forall i, j, \qquad \hat{S}^d_{ij} = N_{ij} S^d_{ij}. \tag{10}$$

This results in a sparse similarity matrix for drugs. The same procedure is done for the target similarity matrix S^t.

The graph regularization described in Section 4.2.3 helps to learn manifolds for the drug and target spaces whereabout the data is assumed to lie; an assumption when learning manifolds (called the *local invariance assumption*) is that points close to each other in the original space should also be close to each other in the learned manifold, which is achieved by obtaining the p-nearest neighbor graphs that preserve the local geometries of the original data [24].

4.2.2 Low-rank approximation

GRMF depends on the basic idea of low rank approximation (LRA) which decomposes the drug-target matrix $Y \in \mathbb{R}^{n \times m}$ into two low-rank latent feature matrices $A \in \mathbb{R}^{n \times k}$ (for drugs) and $B \in \mathbb{R}^{m \times k}$ (for targets) which minimize the LRA objective

$$\min_{A,B} \|Y - AB^T\|_F^2 \tag{11}$$

where $\|.\|_F$ is the Frobenius norm and k is the number of latent features in A and B. Note that k here is different from K of WKNKN.

4.2.3 Regularization

To prevent overfitting and increase generalization capability, Tikhonov and graph regularization terms are added, giving GRMF's objective function

$$\begin{aligned} \min_{A,B} \quad & \|Y - AB^T\|_F^2 \\ & + \lambda_l (\|A\|_F^2 + \|B\|_F^2) \\ & + \lambda_d \sum_{i,r=1}^{n} \hat{S}^d_{ir} \|a_i - a_r\|^2 \\ & + \lambda_t \sum_{j,q=1}^{m} \hat{S}^t_{jq} \|b_j - b_q\|^2 \end{aligned} \tag{12}$$

where λ_l, λ_d and λ_t are positive parameters, a_i and b_j are the i^{th} and j^{th} rows of A and B, respectively, and n and m are the numbers of drugs and targets, respectively. GRMF's objective function balances several goals. The first term requires the model to approximate the matrix Y. The second term is the Tikhonov regularization which minimizes the norms of both A and B. The third term is for drug graph regularization which minimizes the distance between latent feature vectors of two neighboring drugs. The final term is for target graph regularization. The above equation can be rewritten as

$$\begin{aligned} \min_{A,B} \quad & \|Y - AB^T\|_F^2 \\ & + \lambda_l (\|A\|_F^2 + \|B\|_F^2) \\ & + \lambda_d Tr(A^T \mathcal{L}_d A) \\ & + \lambda_t Tr(B^T \mathcal{L}_t B) \end{aligned} \tag{13}$$

where $Tr(\cdot)$ is the trace of a matrix, $\mathcal{L}_d = D^d - \hat{S}^d$ and $\mathcal{L}_t = D^t - \hat{S}^t$ are the graph Laplacians for $\hat{S}^d$ and $\hat{S}^t$, respectively, and $D^d_{ii} = \sum_r \hat{S}^d_{ir}$ and $D^t_{jj} = \sum_q \hat{S}^t_{jq}$ are diagonal matrices. For more details on the rewriting of the graph regularization terms, please refer to [25].

Furthermore, since normalized graph Laplacians are known to perform better than their unnormalized versions in many cases [26], the graph Laplacians $\mathcal{L}_d$ and $\mathcal{L}_t$ are replaced by their normalized counterparts

$\tilde{\mathcal{L}_d} = (D^d)^{-1/2}\mathcal{L}_d(D^d)^{-1/2}$ and $\tilde{\mathcal{L}_t} = (D^t)^{-1/2}\mathcal{L}_t(D^t)^{-1/2}$ giving

$$\begin{aligned}\min_{A,B} \quad & \|Y - AB^T\|_F^2 \\ & +\lambda_l(\|A\|_F^2 + \|B\|_F^2) \\ & +\lambda_d Tr(A^T\tilde{\mathcal{L}_d}A) \\ & +\lambda_t Tr(B^T\tilde{\mathcal{L}_t}B). \end{aligned} \quad (14)$$

Intuitively, minimizing GRMF's objective function would rank the interactions above the non-interactions by forcing their prediction scores to tend to 1 and 0, respectively. However, since many of the 0's are likely interactions that are not discovered yet, this may lead to unsatisfactory results. In this work, WKNKN is applied to the datasets prior to running GRMF technique to address this problem in advance.

4.2.4 Initialization of A and B

The first step in the GRMF procedure (see Figure 2) is to initialize A and B. To do so, we adapted a method inspired by the procedure used in [27]. Specifically, we decomposed $Y \in \mathbb{R}^{n\times m}$ into $U \in \mathbb{R}^{n\times k}$, $S_k \in \mathbb{R}^{k\times k}$ and $V \in \mathbb{R}^{m\times k}$ such that US_kV^T is the closest k-rank approximation to Y where U and V are matrices having orthonormal columns, and S_k is a diagonal matrix containing the k largest singular values. The maximum possible number of singular values in Y is $\min(n, m)$, thus $k_{max} = \min(n, m)$. Finally, we get the square root of S_k and let $A = US_k^{1/2}$ and $B = VS_k^{1/2}$.

While other possible initialization methods for A and B exist [28], we found this method to be quite robust in that results obtained by it were highly reproducible relative to other methods.

4.2.5 Alternating least squares

Next, alternating least squares is used to obtain the solution. Denoting the objective function in Equation 14 as L, we derive the following two alternative update rules (by setting $\frac{\partial L}{\partial A} = 0$ and $\frac{\partial L}{\partial B} = 0$) which are run alternatingly until convergence

$$A = (YB - \lambda_d\tilde{\mathcal{L}_d}A)(B^TB + \lambda_l I_k)^{-1} \quad (15)$$

$$B = (Y^TA - \lambda_t\tilde{\mathcal{L}_t}B)(A^TA + \lambda_l I_k)^{-1} \quad (16)$$

4.2.6 Weighted graph-regularized matrix factorization (WGRMF)

WGRMF is another variant of GRMF that has a weight matrix W identical to that used in CMF. The point behind the weight matrix W is to prevent unknown instances (for which interaction information is not available) from contributing to the determination of the latent feature matrices A and B which reconstruct the drug-target matrix Y. Including W makes the objective function in Equation 14 look as follows

$$\begin{aligned}\min_{A,B} \quad & \|W \odot (Y - AB^T)\|_F^2 \\ & +\lambda_l(\|A\|_F^2 + \|B\|_F^2) \\ & +\lambda_d Tr(A^T\tilde{\mathcal{L}_d}A) \\ & +\lambda_t Tr(B^T\tilde{\mathcal{L}_t}B). \end{aligned} \quad (17)$$

function GRMF ($Y, S^d, S^t, K, \eta, k, \lambda_l, \lambda_d, \lambda_t$)
  $Y = WKNKN(Y, S^d, S_g^t, K, \eta)$ ▷ optional
  $[U, S, V] = SVD(Y, k)$
  $A = US_k^{1/2}$
  $B = VS_k^{1/2}$
  Compute $\tilde{\mathcal{L}_d}, \tilde{\mathcal{L}_t}$ from $\hat{S}^d, \hat{S}^t$
  repeat
    $A = (\ YB\ - \lambda_d\tilde{\mathcal{L}_d}A)(B^TB + \lambda_l I_k)^{-1}$
    $B = (Y^TA - \lambda_t\tilde{\mathcal{L}_t}B)(A^TA + \lambda_l I_k)^{-1}$
  until convergence
  $\hat{Y} = AB$
  return $\hat{Y}$
end function

Fig. 2: GRMF algorithm

Denoting the objective function in Equation 17 as L and letting a_i and b_j be the i^{th} and j^{th} row vectors of A and B, respectively, we set $\frac{\partial L}{\partial a_i} = 0$ and $\frac{\partial L}{\partial b_j} = 0$ to derive the following two update rules that are run alternatingly until convergence

$$\forall i = 1 \ldots n,$$
$$a_i = (\sum_{j=1}^{m} W_{ij}Y_{ij}b_j - \lambda_d(\tilde{\mathcal{L}_d})_{i*}A)(\sum_{j=1}^{m} W_{ij}b_j^Tb_j + \lambda_l I_k)^{-1} \quad (18)$$

$$\forall j = 1 \ldots m,$$
$$b_j = (\sum_{i=1}^{n} W_{ij}Y_{ij}a_i - \lambda_t(\tilde{\mathcal{L}_t})_{j*}B)(\sum_{i=1}^{n} W_{ij}a_i^Ta_i + \lambda_l I_k)^{-1} \quad (19)$$

where $(\tilde{\mathcal{L}_d})_{i*}$ and $(\tilde{\mathcal{L}_t})_{j*}$ are the i^{th} and j^{th} row vectors of $\tilde{\mathcal{L}_d}$ and $\tilde{\mathcal{L}_t}$, respectively. Note that the above update rules differ from GRMF's update rules (Equations 15 and 16); GRMF's rules are *matrix-wise* update rules, while those of WGRMF are *row-wise* update rules.

5 Results

5.1 Cross validation experiments

We performed experiments to compare the existing techniques BLM-NII, RLS-WNN and CMF with our proposed method. Specifically, we conducted 5 repetitions of 10-fold cross validation (CV) for each of the methods, both with and without WKNKN as a preprocessing step. In each repetition of 10-fold CV, Y was divided into 10 folds and each fold, in turn, was left out as the test set while the remaining 9 folds were treated as the training set. In previous studies (e.g. [11], [14], [15]), the Area Under the Precision-Recall curve (AUPR) [29] was employed as the main metric for performance evaluation. As such, AUPR was also used in this work as our evaluation metric. Furthermore, AUPR heavily penalizes non-interactions that are highly ranked, which is desirable here because, in practice, we do not want incorrect predictions to be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2016.2530062, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 2: AUPR results for interaction prediction under CV_d

Methods	NR	GPCR	IC	E
BLM-NII	0.410 (0.043)	0.233 (0.010)	0.201 (0.009)	0.167 (0.011)
RLS-WNN	0.519 (0.025)	0.363 (0.008)	0.319 (0.011)	0.386 (0.014)
CMF	0.482 (0.034)	0.406 (0.008)	0.350 (0.008)	0.375 (0.007)
GRMF	0.517 (0.025)	0.369 (0.011)	0.341 (0.016)	0.349 (0.012)
WGRMF	0.520 (0.025)	0.408 (0.010)	0.364 (0.018)	**0.404 (0.014)**
WKNKN	0.529 (0.015)	0.399 (0.011)	0.352 (0.014)	0.388 (0.011)
WKNKN+BLM-NII	0.514 (0.023)	0.386 (0.005)	0.350 (0.014)	0.385 (0.005)
WKNKN+RLS-WNN	0.523 (0.030)	0.395 (0.008)	0.352 (0.014)	0.385 (0.007)
WKNKN+CMF	0.515 (0.032)	0.409 (0.010)	0.350 (0.015)	0.385 (0.004)
WKNKN+GRMF	**0.542 (0.028)**	0.404 (0.011)	0.356 (0.014)	0.390 (0.010)
WKNKN+WGRMF	0.528 (0.033)	**0.410 (0.012)**	**0.369 (0.017)**	0.401 (0.013)

Best AUPR result in each column is **bold**. Standard deviations are given in (parentheses).

recommended by the prediction algorithm (i.e. the AUPR metric heavily punishes highly ranked false positives [29]). In our experiments, we calculated an AUPR score for each 10-fold CV repetition, and the final AUPR score was the average over 5 such repetitions.

To test different aspects of the prediction methods, we performed CV under two scenarios described in [30]:

1) CV_d, where entire drug interaction profiles are left out to be used as the test set;
2) CV_t, where entire target interaction profiles are left out to be used as the test set.

Given an interaction prediction method, CV_d tests its ability to predict interactions for new drugs while CV_t tests its ability to predict interactions for new targets.

Tables 2 and 3 contain the results from the above-mentioned cross validation scenarios, and these results are explained in the coming subsections. Corresponding plots of the precision-recall curves are also provided in GRMF's website:
http://www1.i2r.a-star.edu.sg/~xlli/GRMF/index.html

5.1.1 Parameter settings

We performed cross validation on the training set for setting GRMF's parameters, namely k (rank of matrices A and B), λ_l, λ_d and λ_t. Using grid search, the best parameter combination is obtained from the values: $k \in \{50, 100\}$, $\lambda_l \in \{2^{-2}, 2^{-1}, 2^0, 2^1\}$, $\lambda_d/\lambda_t \in \{0, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. Note that, as mentioned in Section 4, k could not be more than $min(n, m)$; e.g. if the value being tested for k is 50 and $min(n, m) < 50$, we then set $k = min(n, m)$. WKNKN's parameters (K and η) were also set using grid search. As for p, it was set to $p = 5$.

In the case of WGRMF, for instances (d_i, t_j) in Y that were in the test set, we set $W_{ij} = 0$ in the weight matrix W, which means that such instances would not contribute to the updating of the latent feature matrices A and B which will be used to reconstruct the final predictions matrix $\hat{Y}$.

For the other methods, we strived to set the parameters to their optimal values wherever possible. In the end, all parameters were set to their default values (which were found to be already optimal) with the exception of η which was set as 0.7 in RLS-WNN. For CMF, the parameters were obtained as mentioned in [17].

5.1.2 Interaction prediction under CV_d

Results under CV_d are shown in Table 2. In the NR dataset, GRMF and WGRMF are comparable with RLS-WNN. While CMF is better than GRMF in the other three datasets, WGRMF is consistently better than CMF in all datasets, though the results are almost the same in the GPCR dataset. Without the use of WKNKN as a preprocessing step, WGRMF is the superior method.

Furthermore, the additional weight matrix W in WGRMF provided a good boost to the prediction performance as compared with GRMF (except in the NR dataset). In addition, thanks to the graph regularization term in WGRMF, it displayed an obvious improvement over CMF overall, supporting the usefulness of manifold learning.

After applying WKNKN, in all datasets, all of GRMF's results greatly improved after applying WKNKN. On the other hand, the improvements in WGRMF after applying WKNKN were not as pronounced; there was some improvement in the NR dataset and slight improvements in the GPCR and IC datasets, while the AUPR score actually decreased in the E dataset. In the end, WGRMF is still the superior method. In addition, GRMF generally displayed comparable results to WGRMF, and in the NR dataset, GRMF did better than WGRMF.

5.1.3 Interaction prediction under CV_t

Results under CV_t are shown in Table 3. The first thing that can be observed is that AUPR scores under CV_t are generally higher than those under CV_d. That is, with absent target interaction profiles, the different methods can still achieve good prediction performance, whereas their performance is much decreased if drug interaction profiles are hidden instead. This observation implies that target sequence similarity is more reliable and informative than drug chemical similarity, a conclusion reached in previous work [13].

In the NR dataset, RLS-WNN is the best method, while in the rest of the datasets, WGRMF is the superior method. As in the CV_d results, due to the additional weight matrix W, WGRMF is better than GRMF on all datasets (except on the NR dataset where the results are the same).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2016.2530062, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 3: AUPR results for interaction prediction under CV_t

Methods	NR	GPCR	IC	E
BLM-NII	0.418 (0.022)	0.447 (0.024)	0.634 (0.008)	0.583 (0.021)
RLS-WNN	0.468 (0.030)	0.547 (0.025)	0.746 (0.004)	0.761 (0.015)
CMF	0.379 (0.020)	0.540 (0.028)	0.751 (0.014)	0.740 (0.014)
GRMF	0.423 (0.032)	0.567 (0.027)	0.745 (0.008)	0.763 (0.020)
WGRMF	0.423 (0.017)	0.574 (0.027)	0.801 (0.008)	0.801 (0.018)
WKNKN	0.465 (0.034)	0.572 (0.027)	0.787 (0.010)	0.792 (0.016)
WKNKN+BLM-NII	0.460 (0.032)	0.607 (0.027)	0.794 (0.005)	**0.814 (0.017)**
WKNKN+RLS-WNN	0.471 (0.028)	0.603 (0.020)	0.806 (0.007)	0.809 (0.017)
WKNKN+CMF	0.434 (0.029)	0.557 (0.021)	0.742 (0.015)	0.772 (0.014)
WKNKN+GRMF	**0.500 (0.028)**	**0.615 (0.023)**	**0.815 (0.010)**	0.807 (0.016)
WKNKN+WGRMF	0.446 (0.015)	0.585 (0.027)	0.799 (0.007)	0.798 (0.018)

Best AUPR result in each column is **bold**. Standard deviations are given in (parentheses).

Furthermore, WGRMF achieved better results than CMF thanks to the graph regularization terms.

After applying WKNKN, in the NR, GPCR and IC datasets, GRMF is the superior method. In contrast to GRMF whose results greatly improved after applying WKNKN, the AUPR values did not increase as much for WGRMF (in fact, there were decreases in the IC and E dataset). In the GPCR, IC and E datasets, the performances of BLM-NII and RLS-WNN are comparable to that of GRMF, especially in the E dataset.

Interestingly, with WKNKN as a preprocessing method, GRMF did better than WGRMF on all datasets, especially on the NR dataset. Note that WGRMF's weight matrix W forces the instances of the test set to not contribute in the updating of the latent feature matrices A and B (that are used to obtain predictions later); this weight matrix W indeed improves the results when WKNKN is not being used. However, after applying WKNKN which updates the drug-target matrix Y, the weight matrix W also forces the test set instances (despite their *updated* values) to not contribute to the predictions in WGRMF. On the other hand, GRMF does not have a weight matrix W so these instances (and their updated values) were not ignored. Since target information is more important than drug information under CV_t and is generally more reliable, this has made GRMF's results better than those of WGRMF.

5.1.4 Interaction prediction with WKNKN

In tables 2 and 3, WKNKN was tested by itself and added as a baseline method. Interestingly, WKNKN produced results that are quite competitive; under both CV_d and CV_t, WKNKN set a challenging baseline that was difficult for most methods to surpass. Only WGRMF was able to consistently produce results that are better than those of WKNKN (except in the NR dataset).

Furthermore, while WKNKN generally improved the results of the different methods (as shown in the last 5 rows in Tables 2 and 3), many of the results were still not better than those of WKNKN itself. However, for WKNKN+GRMF and WKNKN+WGRMF, improvements over WKNKN itself were realized in all datasets (as shown in the last 2 rows of each table).

5.1.5 Sensitivity Analysis

The optimal parameter values for k, λ_l, λ_d and λ_t are determined automatically via cross validation on the training set *during* the execution of GRMF (refer to Section 5.1.1); these values may differ from one fold to another depending on the training set used in each, which complicates the task of plotting sensitivity analysis results for the parameters.

In order to study the roles of λ_l, λ_d and λ_t, we compared GRMF's results from Tables 2 and 3 against variants of GRMF where we set $\lambda_l = 0$ for the first variant, $\lambda_d = 0$ for the second variant and $\lambda_t = 0$ for the third. Results for these GRMF variants under CV_d and CV_t are reported in Tables 4 and 5, respectively. It was found that setting $\lambda_l = 0$ negatively impacts the performance under both CV_d and CV_t. As for λ_d, setting it to 0 negatively impacts results under CV_d, but not so much under CV_t. Vice versa for λ_t, setting it to 0 negatively impacts results under CV_t, but not so much under CV_d. This means that λ_d is important under CV_d, while λ_t is important under CV_t.

As for k, it was found that the higher the value of k, the better the prediction performance. However, due to computational complexity reasons, we set the value of k to

TABLE 4: AUPR results for GRMF variants under CV_d

Methods	NR	GPCR	IC	E
GRMF	**0.517**	**0.369**	**0.341**	**0.349**
GRMF ($\lambda_l = 0$)	0.105	0.293	0.263	0.264
GRMF ($\lambda_d = 0$)	0.122	0.049	0.034	0.011
GRMF ($\lambda_t = 0$)	0.516	0.369	0.341	0.349

Best AUPR result in each column is **bold**.

TABLE 5: AUPR results for GRMF variants under CV_t

Methods	NR	GPCR	IC	E
GRMF	**0.423**	**0.567**	**0.745**	**0.763**
GRMF ($\lambda_l = 0$)	0.094	0.417	0.603	0.671
GRMF ($\lambda_d = 0$)	0.411	0.560	0.741	0.761
GRMF ($\lambda_t = 0$)	0.089	0.039	0.037	0.011

Best AUPR result in each column is **bold**.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2016.2530062, IEEE/ACM Transactions on Computational Biology and Bioinformatics

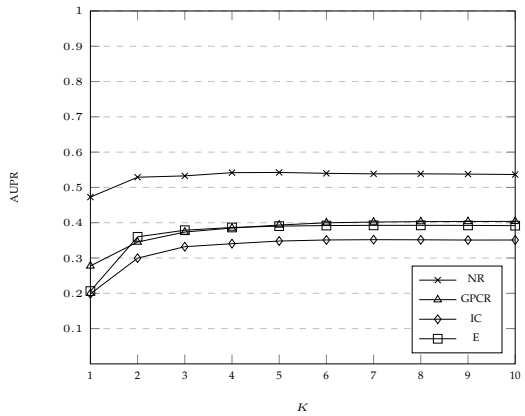


Fig. 3: Sensitivity analysis for K under CV_d

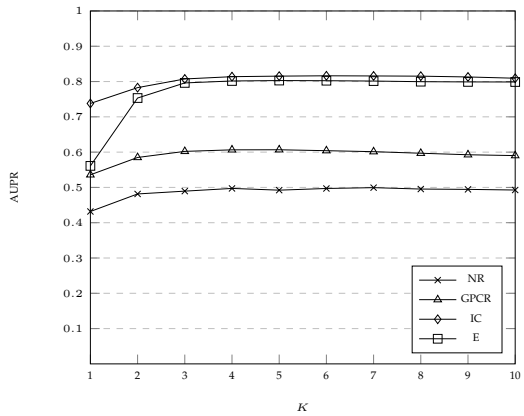


Fig. 5: Sensitivity analysis for K under CV_t

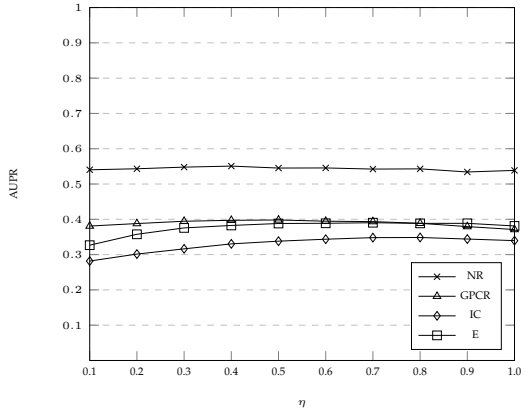


Fig. 4: Sensitivity analysis for η under CV_d

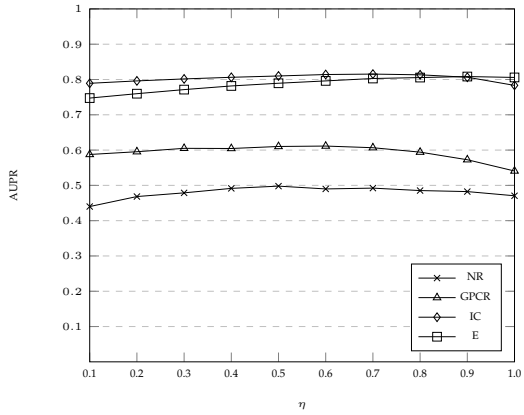


Fig. 6: Sensitivity analysis for η under CV_t

a maximum of 100. In addition, the results were found not to improve much beyond $k = 100$.

On the other hand, the parameters K and η (from WKNKN) are fixed *before* running GRMF. Sensitivity analyses are provided for these two parameters in Figures 3, 4, 5 and 6.

5.1.6 General Comments

Following are some remaining findings and conclusions reached from our cross validation experiments:

- Contrary to the other datasets, the results of the NR dataset under CV_t are lower than those under CV_d. In [30], this was concluded to be due to the small size of the NR dataset (see Table 1) which is causing the results to be unstable. We complement this by saying that target information is more important for inference under CV_t, and since the number of targets in the NR dataset is very small, this means that there are few reliable targets to infer from (as opposed to CV_d where there are relatively more drugs to infer from). Thus, the prediction performance is negatively affected by the lack of a sufficient number of neighboring targets.
- While the prediction scores may differ between datasets, this may not be solely due to the type of interactions in each dataset (each dataset contains a different type of interactions involving a specific kind of target proteins) but also due to the number of drugs and targets in each dataset and the reliability of the drug and target similarities used as well as the cross validation scenario being performed. For example, in both the NR and GPCR dataset, the drug-to-target ratio is higher than in the IC and E datasets (see Table 1); we believe this is why all methods achieved higher AUPR for the NR and GPCR datasets than for the IC and E datasets under CV_d (where drug information is more important). In addition, it seems that the same drug-to-target ratio caused the opposite to take place under CV_t (where target information is more important) – all methods achieved higher AUPR for the IC and E datasets than for the NR and GPCR datasets. In conclusion, it is important to take into account the drug-to-target ratio while studying the prediction performance of any method on the different datasets.
- By observing the results of our proposed methods in Tables 2 and 3, modeling the manifold structures of the drug and target spaces (via the drug and target graph regularization terms, respectively) was shown to improve prediction performance in terms of AUPR, indicating the effectiveness of the proposed graph regularization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2016.2530062, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 6: Predicted drugs for TRPV6, IC dataset

Rank	Drug	Drug KEGG id
1	**Verapamil**	**D02356**
2	**Gabapentin**	**D00332**
3	**Phenytoin**	**D00512**
4	**Nisoldipine**	**D00618**
5	**Nifedipine**	**D00437**
6	**Nitrendipine**	**D00629**
7	**Cinnarizine**	**D01295**
8	**Pimozide**	**D00560**
9	**Halothane**	**D00542**
10	Diazoxide	D00294
11	Nimodipine	D00438
12	Verapamil hydrochloride	D00619
13	Terfenadine	D00521
14	Isradipine	D00349
15	Felodipine	D00319
16	Methoxyflurane	D00544
17	L-Proline	D00035
18	Amiloride hydrochloride	D00649
19	Mibefradil dihydrochloride	D05024
20	Penfluridol	D02630

Known interactions are in **bold**

TABLE 7: Predicted targets for Hexobarbital, E dataset

Rank	Target	Target KEGG id
1	**cytochrome P450 CYP19A1**	**hsa1588**
2	**cytochrome P450 CYP4Z1**	**hsa199974**
3	**cytochrome P450 CYP4X1**	**hsa260293**
4	**cytochrome P450 CYP4B1**	**hsa1580**
5	**cytochrome P450 CYP2J2**	**hsa1573**
6	hydroxymethylglutaryl-CoA reductase HMGCR	hsa3156
7	**cytochrome P450 CYP2S1**	**hsa29785**
8	**cytochrome P450 CYP2F1**	**hsa1572**
9	**cytochrome P450 CYP2B6**	**hsa1555**
10	**cytochrome P450 CYP1B1**	**hsa1545**
11	**cytochrome P450 CYP3A43**	**hsa64816**
12	**cytochrome P450 CYP2A6**	**hsa1548**
13	**cytochrome P450 CYP3A5**	**hsa1577**
14	**cytochrome P450 CYP3A7**	**hsa1551**
15	**cytochrome P450 CYP2C18**	**hsa1562**
16	**cytochrome P450 CYP2A13**	**hsa1553**
17	**cytochrome P450 CYP2E1**	**hsa1571**
18	**cytochrome P450 CYP17A1**	**hsa1586**
19	**cytochrome P450 CYP1A1**	**hsa1543**
20	**cytochrome P450 CYP2C19**	**hsa1557**

Known interactions are in **bold**

- Finally, an advantage of GRMF over WGRMF that is not spoken in the results is that GRMF runs much faster than WGRMF because GRMF uses *matrix-wise* update rules for A and B (Equations 15 and 16), while WGRMF is forced to use *row-wise* update rules (Equations 18 and 19) due to the weight matrix W.

5.2 Predicting novel interactions

In this section, we simulate some cases where an arbitrary drug or target is left out (i.e. its interactions are removed from the original dataset) to see if its interactions would be predicted successfully. We particularly focused on drugs and targets for which the similarity to the nearest neighbor (according to S^d and S^t, respectively) is low as they represent the tougher cases to predict interactions for; according to [18], nearest neighbor information has high influence on the prediction performance, so if the nearest known neighbor to a new drug or target is not so near, it may be difficult to accurately predict interactions for it.

From the IC dataset, we left out the target protein – transient receptor potential cation channel, TRPV6 (KEGG id: hsa55503) – for which there are 9 known interactions in Y. After GRMF was run on the modified dataset, all drugs were sorted in descending order of how likely they would interact with TRPV6; the top 20 predicted interactions for TRPV6 are given in Table 6. The 9 interactions of TRPV6 were predicted successfully in the top 20. RLS-WNN and CMF were also able to predict these interactions in their top 20. As for BLM-NII, it was able to predict 7 of the 9 interactions, and the predicted interactions were not as highly ranked as in the other methods' predictions.

The same procedure was done for the drug – Hexobarbital (KEGG id: D01071) – by leaving it out of the E dataset before running GRMF; the top 20 predictions are given in Table 7. 19 out of the top 20 predicted interactions were known interactions in Y. A similar prediction performance was displayed by RLS-WNN, while both BLM-NII and CMF were each able to predict only 1 known interaction in their top 20. Tables of predicted interactions for TRPV6 and Hexobarbital using BLM-NII, RLS-WNN and CMF are provided as supplementary material at GRMF's website.

We remind the reader that the above two cases (i.e. TRPV6 and Hexobarbital) are considered difficult cases. Specifically, the similarity of TRPV6 to its nearest neighboring target (according to S^t) is as low as 0.05, while the similarity of Hexobarbital to its nearest neighboring drug (according to S^d) is 0.35 which is also quite low. According to these cases, it was shown that GRMF performs reasonably well, which was also confirmed by other experiments that have been conducted (results added to GRMF's website). In conclusion, GRMF is generally able to predict challenging interactions.

6 Conclusion

In this paper, we presented two matrix factorization methods, GRMF and WGRMF, for drug-target interaction prediction. Both of them implicitly perform manifold

learning via graph regularization. WGRMF differs from GRMF in that it additionally has a weight matrix W that prevents unknown instances (i.e. drug-target pairs for which interaction information is not available) from contributing to the final predictions. Experiments were conducted using two different types of cross validations, namely CV_d (drug side CV) and CV_t (target side CV), to compare our methods with three other competing state-of-the-art methods. In most of the cases, the top results belonged to either GRMF or WGRMF, showing that manifold learning resulted in improvements in prediction performance.

Moreover, we developed a preprocessing step, WKNKN, which transforms all the 0's in the given drug-target matrix into interaction likelihood values. This is important as many 0's in the matrix actually correspond to unknown cases or missing values rather than confirmed non-interactions. It was observed that WKNKN improves results dramatically for all methods but with exceptions for those methods that contain a weight matrix W (i.e. CMF and WGRMF). Furthermore, when WKNKN is applied, results of GRMF became comparable to those of WGRMF, which is an added bonus because GRMF actually runs much faster due to its *matrix-wise* update rules.

We then performed a couple of experiments to investigate GRMF's ability to predict novel interactions for new drugs and targets. Specifically, the entire profile of a known drug or target was left out, and then GRMF was run to see how many of its known interactions would get predicted. GRMF's ability to predict new interactions in such cases was successfully confirmed in the experiments.

As future work, other sources of information may be used to further improve prediction performance. For example, instead of using just one kernel (or similarity matrix) for drugs and one for targets, multiple kernels may be used for either drugs or targets in a multiple kernel learning algorithm [31].

7 Acknowledgements

This work is supported by the Agency for Science, Technology and Research (A*STAR), Singapore.

References

[1] N. Novac, "Challenges and opportunities of drug repositioning," *Trends in pharmacological sciences*, vol. 34, no. 5, pp. 267–272, 2013.

[2] M. Hurle, L. Yang, Q. Xie, D. Rajpal, P. Sanseau, and P. Agarwal, "Computational drug repositioning: from data to therapeutics," *Clinical Pharmacology & Therapeutics*, vol. 93, no. 4, pp. 335–341, 2013.

[3] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, "Kegg for integration and interpretation of large-scale molecular data sets," vol. 40, no. D1, pp. D109–D114, 2012.

[4] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, and D. S. Wishart, "Drugbank 3.0: a comprehensive resource for 'omics' research on drugs," vol. 39, no. suppl 1, pp. D1035–D1041, 2011.

[5] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, and J. P. Overington, "Chembl: a large-scale bioactivity database for drug discovery," 2011.

[6] M. Kuhn, D. Szklarczyk, S. Pletscher-Frankild, T. H. Blicher, C. von Mering, L. J. Jensen, and P. Bork, "Stitch 4: integration of protein chemical interactions with user data," vol. 42, no. D1, pp. D401–D407, 2014.

[7] L. Xie, S. L. Kinnings, L. Xie, and P. E. Bourne, *Predicting the Polypharmacology of Drugs: Identifying New Uses through Chemoinformatics, Structural Informatics, and Molecular Modeling-Based Approaches*. John Wiley & Sons, Inc.

[8] Z. Mousavian and A. Masoudi-Nejad, "Drug-target interaction prediction via chemogenomic space: learning-based methods," *Expert opinion on drug metabolism & toxicology*, vol. 10, no. 9, pp. 1273–1287, 2014.

[9] A. C. Cheng, R. G. Coleman, K. T. Smyth, Q. Cao, P. Soulard, D. R. Caffrey, A. C. Salzberg, and E. S. Huang, "Structure-based maximal affinity model predicts small-molecule druggability," *Nature biotechnology*, vol. 25, no. 1, pp. 71–75, 2007.

[10] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, and B. K. Shoichet, "Relating protein pharmacology by ligand chemistry," *Nature biotechnology*, vol. 25, no. 2, pp. 197–206, 2007.

[11] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug–target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.

[12] K. Bleakley and Y. Yamanishi, "Supervised prediction of drug–target interactions using bipartite local models," *Bioinformatics*, vol. 25, no. 18, pp. 2397–2403, 2009.

[13] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug–target interaction," *Bioinformatics*, vol. 27, no. 21, pp. 3036–3043, 2011.

[14] J.-P. Mei, C.-K. Kwoh, P. Yang, X.-L. Li, and J. Zheng, "Drug-target interaction prediction by learning from local information and neighbors," *Bioinformatics*, vol. 29, no. 2, pp. 238–245, 2013.

[15] T. van Laarhoven and E. Marchiori, "Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile," *PloS one*, vol. 8, no. 6, p. e66952, 2013.

[16] M. Gönen, "Predicting drug–target interactions from chemical and genomic kernels using bayesian matrix factorization," *Bioinformatics*, vol. 28, no. 18, pp. 2304–2310, 2012.

[17] X. Zheng, H. Ding, H. Mamitsuka, and S. Zhu, "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1025–1033.

[18] A. M. Wassermann, H. Geppert, and J. Bajorath, "Ligand prediction for orphan targets using support vector machines and various target-ligand kernels is dominated by nearest neighbor effects," *Journal of Chemical Information and Modeling*, vol. 49, no. 10, pp. 2155–2167, 2009, pMID: 19780576.

[19] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," vol. 290, no. 5500, pp. 2319–2323, 2000.

[20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," vol. 290, no. 5500, pp. 2323–2326, 2000.

[21] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems 14*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, 2002, pp. 585–591.

[22] M. Hattori, N. Tanaka, M. Kanehisa, and S. Goto, "Simcomp/subcomp: chemical structure search servers for network analyses," *Nucleic acids research*, p. gkq367, 2010.

[23] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *Journal of molecular biology*, vol. 147, no. 1, pp. 195–197, 1981.

[24] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, 2011.

[25] Q. Gu, J. Zhou, and C. H. Ding, "Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs," in *SDM*. SIAM, 2010, pp. 199–210.

[26] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *The Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.

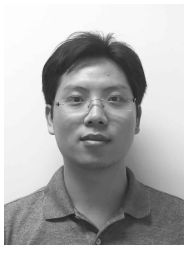
[27] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Application of dimensionality reduction in recommender system – a case study," in *ACM WebKDD Workshop*, 2000.

[28] A. N. Langville, C. D. Meyer, R. Albright, J. Cox, and D. Duling, "Algorithms, initializations, and convergence for the nonnegative matrix factorization," *CoRR*, vol. abs/1407.7299, 2014.

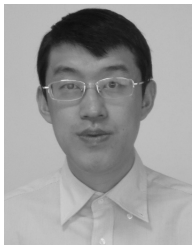
[29] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 233–240.
[30] T. Pahikkala, A. Airola, S. Pietilä, S. Shakyawar, A. Szwajda, J. Tang, and T. Aittokallio, "Toward more realistic drug–target interaction predictions," vol. 16, no. 2, pp. 325–337, 2015.
[31] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *The Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

Ali Ezzat is a Ph.D. student at the School of Computer Engineering in Nanyang Technological University (NTU), Singapore. He received the B.Sc. degree in computer science from Ain Shams University, Egypt, in 2006 and the M.Sc. degree in bioinformatics from Nanyang Technological University, Singapore, in 2013. He also worked as an IT professional at United OFOQ, Egypt, from 2008 to 2011. At present, he is affiliated with the Biomedical Informatics Graduate Lab at NTU and the Data Analytics department at the Institute for Infocomm Research (I2R), A*STAR, Singapore. His research interests include machine learning, data mining and bioinformatics.

Peilin Zhao is currently a Research Scientist at the Institute for Infocomm Research (I2R), A*STAR, Singapore. His research Interests are Machine Learning and its applications to Big Data Analytics, etc. He was a visiting scholar at the Big Data Lab, Baidu Research from Jan 2014 to Jun 2014, and a Postdoc Fellow with Prof. Tong Zhang in the Department of Statistics, Rutgers University, from Sep 2013 to Jun 2014. During these periods, his main research topic was on Stochastic Optimization. He received his PHD degree (with thesis topic focused on Online Learning theory and methods) from School of Computer Engineering, Nanyang Technological University and his bachelor degree from Department of Mathematics, Zhejiang University. In his research areas, he has published over 40 papers in top venues, including JMLR, AIJ, ICML, NIPS, KDD, etc. He has been invited as a PC member or reviewer for many international conferences and journals in his area.



Min Wu is currently a Research Scientist in the Data Analytics Department at the Institute for Infocomm Research (I2R) under the Agency for Science, Technology and Research (A*STAR), Singapore. He received the B.Eng. from the University of Science and Technology of China (USTC), China in 2006 and his Ph.D. degree from Nanyang Technological University, Singapore in 2011. He received the best paper award in DASFAA 2015 and won the IJCAI contest 2015 on repeated buyers prediction after sales promotion. His current research interests include machine learning, data mining and bioinformatics.



Xiaoli Li is currently a machine learning lab head at the Institute for Infocomm Research, A*STAR, Singapore. He also holds adjunct associate professor positions at the National University of Singapore and Nanyang Technological University. His research interests include data mining, machine learning and bioinformatics. He has served as a PC member/workshop chair/session chair in leading data mining related conferences (including KDD, ICDM, SDM, PKDD/ECML, PAKDD, WWW, AAAI, and CIKM) and as an editor of bioinformatics-related books. In 2005, he received the Best Paper Award in the 16th International Conference on Genome Informatics (GIW 2005). In 2011, he received the Best Paper Runner-Up Award in the 16th International Conference on Database Systems for Advanced Applications (DASFAA 2011). Xiaoli has published more than 120 papers, including top tier conferences such as KDD, ICDM, SDM, PKDD/ECML, ICML, IJCAI, AAAI, ACL, EMNLP, SIGIR, CIKM, UbiCom, etc. as well as some top tier journals such as IEEE Transactions TKDE, Bioinformatics and IEEE Transactions on Reliability. Please refer to http://www1.i2r.a-star.edu.sg/ xlli/ for more details.

Chee-Keong Kwoh received the bachelor's degree in electrical engineering (first class) and the master's degree in industrial system engineering from the National University of Singapore in 1987 and 1991, respectively. He received the PhD degree from the Imperial College of Science, Technology and Medicine, University of London, in 1995. He has been with the School of Computer Engineering, Nanyang Technological University (NTU) since 1993. He is the programme director of the "MSc in Bioinformatics" programme at NTU. His research interests include data mining, soft computing and graph-based inference; applications areas include bioinformatics and biomedical engineering. He has done significant research work in his research areas and has published many quality international conferences and journal papers. He is an editorial board member of the International Journal of Data Mining and Bioinformatics; the Scientific World Journal; Network Modeling and Analysis in Health Informatics and Bioinformatics; Theoretical Biology Insights; and Bioinformation. He has been a guest editor for many journals such as Journal of Mechanics in Medicine and Biology, the International Journal on Biomedical and Pharmaceutical Engineering and others. He has often been invited as an organizing member or referee and reviewer for a number of premier conferences and journals including GIW, IEEE BIBM, RECOMB, PRIB, BIBM, ICDM and iCBBE. He is a member of the Association for Medical and Bio-Informatics, Imperial College Alumni Association of Singapore. He has provided many services to professional bodies in Singapore and was conferred the Public Service Medal by the President of Singapore in 2008.