HAL
open science

DeepVir, Graphical Deep Matrix Factorization for In Silico Antiviral Repositioning. Application to COVID19.

Aanchal Mongia, Stuti Jain, Emilie Chouzenoux, Angshul Majumdar

To cite this version:

Aanchal Mongia, Stuti Jain, Emilie Chouzenoux, Angshul Majumdar. DeepVir, Graphical Deep Matrix Factorization for In Silico Antiviral Repositioning. Application to COVID19.. Journal of Computational Biology, 2022, 29 (5), pp.441-452. 10.36227/techrxiv.12987497.v1 . hal-02946338

HAL Id: hal-02946338

https://hal.science/hal-02946338v1

Submitted on 23 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:2009.10333v1 [cs.LG] 22 Sep 2020

DeepVir - Graphical Deep Matrix Factorization for *In Silico* Antiviral Repositioning: Application to COVID-19

Aanchal Mongia[1], Stuti Jain[3], Emilie Chouzenoux[2*] & Angshul Majumdar[3*]

[1]Dept. of CSE, IIIT - Delhi, India, 110020
[2]CVN, Inria Saclay, Univ. Paris Saclay, 91190 Gif-sur-Yvette, France
[3]Dept. of ECE, IIIT - Delhi, India, 110020
*Corresponding authors contact/Email:
emilie.chouzenoux@centralesupelec.fr, angshul@iiitd.ac.in

September 23, 2020

Abstract

This work formulates antiviral repositioning as a matrix completion problem where the antiviral drugs are along the rows and the viruses along the columns. The input matrix is partially filled, with ones in positions where the antiviral has been known to be effective against a virus. The curated metadata for antivirals (chemical structure and pathways) and viruses (genomic structure and symptoms) is encoded into our matrix completion framework as graph Laplacian regularization. We then frame the resulting multiple graph regularized matrix completion problem as deep matrix factorization. This is solved by using a novel optimization method called HyPALM (Hybrid Proximal Alternating Linearized Minimization). Results on our curated RNA drug virus association (DVA) dataset shows that the proposed approach excels over state-of-the-art graph regularized matrix completion techniques. When applied to *in silico* prediction of antivirals for COVID-19, our approach returns antivirals that are either used for treating patients or are under for trials for the same.

Contact: angshul@iiitd.ac.in

1 Introduction

The problem of matrix completion has been deployed in numerous applications of computer science and bioinformatics. One can mention, for instance, recommender systems [13], wireless sensor networks [12] and MIMO channel estimation [73]. Matrix completion also forms the foundation of many biological interaction problems including gene expression imputation [34], drug-target interaction prediction [24, 16, 85], miRNA disease association [11, 40] and gene disease association [54]. As seen in the aforementioned studies, matrix completion is turning out to one of the most promising approaches for modelling biological interactions.

Matrix completion for *in silico* drug re-positioning is an established topic [24, 84]. In such context, the matrix rows relates to drugs while the columns correspond to the diseases. The information about a drug being effective for a disease is encoded with a one into the matrix and zero, otherwise. The matrix is only partially filled since information about all drug disease associations are not known. Matrix completion amounts to filling the missing entries by relying and interpolating over the available knowledge.

Our recent work [52] introduced a dataset of antiviral drug virus association (DVA). It shows how the matrix completion framework can be used to computationally predict the drugs that could be effective against SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), the virus responsible for the ongoing pandemic, COVID-19 (COrona VIrus Disease-2019). In the aforesaid dataset, we manually curated drug-virus associations from `drugbank.ca`. We collected all the anti-viral drugs proved to be effective against viruses infecting humans along with the similarity information between the viruses (genomic structure-based similarity) and drugs (chemical structure-based similarity). The dataset contains both DNA and RNA viruses known to infect human beings. State-of-the-art off-the-shelf matrix completion techniques were then employed to solve the association prediction problem. It was found that graph regularised techniques, including similarities based on genomic structure of viruses and chemical structure of antiviral drugs, performed better than the non-regularised competitors. The most successful techniques

were graph regularized versions of matrix completion [51], binary matrix completion [48] and (shallow) matrix factorization [25].

In this present work, we propose to solve the problem of drug-virus association prediction via graph regularized deep matrix factorization. We propose a novel theoretically sounded algorithm for this task. Moreover, since our final goal is to repurpose drugs for COVID-19, an RNA virus, we have pruned our DVA dataset to consists only RNA viruses. DNA viruses are not meaningful in this context [55], hence we will not be considering them in this work. Furthermore, in addition to the graph based on chemical structure similarity of drugs [52], one more similarity graph has been built on the mode of action of the drugs. Similarly, an additional similarity graph on viruses is built, based on the human symptoms caused by the virus' infection. The current work uses the symptom based similarities along with the genomic similarities used in the prior work [52].

A prior study on graph regularised deep matrix factorization [49] was based on the alternating direction method of multipliers (ADMM) approach [76]. A major issue with ADMM is that the convergence guarantees are rather mild in such challenging non convex scenario. Furthermore, ADMM method requires the resolution of costly inner steps as well as the introduction of several intermediary variables that may be detrimental to the practical efficiency of the method. Owing to the aforesaid reasons, we propose a novel resolution scheme called HyPALM (Hybrid Proximal Alternating Linearized Minimization), relying on the recent techniques developed in [14, 7, 1]. To summarize the main contributions of our proposed work are as follows:

- An original graph structure introducing (i) symptoms based viral similarity in the virus-virus graph, (ii) mechanism-of-action based drug similarity in the drug-drug graph, leading to the so-called multi graph regularized deep matrix factorization (GRDMF) model; We call the approch GRDMF throughout this paper.
- A novel optimization method, named HyPALM, for addressing the resulting matrix factorization problem with sounded convergence guarantees on its iterates.

2 Results

2.1 Algorithm Selection

In this subsection, we study the choice of technique to solve the GRDMF problem (10). In the past work [50, 49], the proposed formulation has been solved using ADMM (alternating direction method of multipliers) [8, 37]. We introduce here a novel resolution strategy named HyPALM. In contrast with ADMM, the iterates of HyPALM are guaranteed to converge to a stationnary point of the non-convex problem. The HyPALM algorithm steps are depicted in the Methods section of the paper.

We randomly drop 10 % of association values in the drug-virus association matrix and report the AUC (Area under the ROC curve) and AUPR (Area under the precision recall curve) on the test entries (averaged over 10 runs of cross validation) after solving 2-layer and 3-layer GRDMF problem with ADMM (first two columns) and HyPALM (last two columns). As can be seen, HyPALM solution gives much better prediction results, for both 2 and 3 layers, and both evaluation metrics . Hence, owing to the better performance and the assessed convergence guarantees, we propose to retain HyPALM in the remaining of the paper for our simulations with the GRDMF model.

Model Algorithm	GRDMF-2L ADMM	GRDMF-3L ADMM	GRDMF-2L HyPALM	GRDMF-3L HyPALM
AUC	0.9076	0.9143	0.9457	0.9516
AUPR	0.6523	0.6945	0.8180	0.8038

Table 1: **AUC and AUPR obtained after running ADMM and HyPALM algorithms for DLMF matrix completion for drug-virus association prediction.**

2.2 Evaluation

We compare the performance of our method with other recent graph regularized matrix completion techniques such has graph regularized matrix completion (GRMC) [51], graph regularized binary matrix completion (GRBMC) [52] and graph regularized techniques like graph regularized matrix factorization (GRMF) [25]. The authors of [52] show

that GRMF and GRBMC perform well in predicting drug-virus association and hence estimating drugs for SARS-CoV-2. We evaluate the performance of 2-layers and 3-layers graph regularized deep matrix factorization model (GRDMF) with the similarities provided by the method from [52] and by the newly proposed models GRDMF-2L and GRDMF-3L.

To quantitatively measure and compare the performance of the proposed and other benchmark algorithms, we carry out 10-fold cross validation on the drug-virus association dataset in 3 settings and observe the standard metrics such as mean AUC and AUPR. The first setting (CV1) corresponds to randomly hiding 10 % association entries as the test set in the association matrix, while the other two settings CV2 and CV3 correspond to randomly hiding 10 % viruses and drugs in the data as test set respectively. Such a validation technique is a standard to evaluate the accuracy in association/interaction prediction problems [16, 74]. As can be observed from Table 2, both version of our algorithm (GRDMF-2L and GRDMF-3L) perform much better than the state-of-the-art in terms of AUPR and gives comparable results in terms of AUC. This metric is more important than AUC because it penalizes more highly ranked false positives so that only highly ranked drug-virus pairs in prediction would be recommended for biological or chemical tests. It can be seen that the prediction results do not improve much in the 3-layers version. Hence, we use the 2-layers GRDMF for further experiments, as it presents lower complexity.

	Model	GRMC	GRBMC	GRMF	GRDMF-2L	GRDMF-3L
CV1	AUC	0.9092	0.9168	0.8926	0.9457	**0.9516**
	AUPR	0.5022	0.55	0.6393	**0.8180**	0.8038
CV2	AUC	0.7907	0.7624	0.7805	**0.7791**	0.7595
	AUPR	0.3304	0.3076	0.3461	**0.4610**	0.4541
CV3	AUC	0.8685	0.8775	0.8841	0.8953	**0.9028**
	AUPR	0.7282	0.7042	0.6885	0.7513	**0.7549**

Table 2: **AUC and AUPR for association prediction for all techniques under the 3 cross validation settings with multiple similarities used for drugs and viruses**.

Let us analyse deeper the results in Table 2. The dataset contains more drugs (86) than viruses (23). In CV1 settings, when entries are randomly omitted, without skipping drugs and viruses our method performs the best. There is enough data and hence our method does not overfit. The results improve from one (GRMF) to two and three layers. In the CV2 setting, when viruses are omitted, the size of the dataset reduces. The proposed deeper method overfits and hence performs poorly compared to shallower techniques. CV3 also reduces the size of the dataset by omitting drugs, but since the number of drugs are significantly larger than the number of viruses, the effect is not pronounced on our proposed algorithm and we see an improved prediction.

We then measure the precision and recall at top k (Pre@k and Rec@k, $k = 3$, $k = 5$ and 7) drug recommendations for viruses while implementing Leave-one-out-cross validation (LOOCV) by hiding (i.e., leaving out) the association profile of every virus. Performing LOOCV tells us how well is the algorithm doing in terms for predicting drugs for current known viruses. This is the most realistic setting, where all the drugs and all the viruses (barring the test virus) are considered. One can see that our algorithm gives comparable precision values and exhibits the best recall at all values of k. This appears as the most realistic case since, on average, there are 4 to 5 drugs for each virus.

	Model	GRMC	GRBMC	GRMF	GRDMF-2L	GRDMF-3L
k=3	Pre@k	0.3478	0.3623	0.3623	0.3043	0.3478
	Rec@k	0.4204	0.4421	0.4639	0.4938	0.5398
k=5	Pre@k	0.2522	0.2609	0.2609	0.2522	0.2870
	Rec@k	0.4554	0.5195	0.5152	0.5956	0.6542
k=7	Pre@k	0.2174	0.2112	0.2174	0.2050	0.2360
	Rec@k	0.5382	0.5288	0.5375	0.6272	0.6892

Table 3: **Precision@k and Recall@k for association prediction for k=5 and 10 with multiple similarities used for drugs and viruses.**

The values for the GRDMF model parameters μ, k_1, k_2, k_3 and for the HyPALM parameters α, ϑ have been set by performing cross-validation on the training set (see Supplementary Material for the list of retained values). Moreover, we set the number of iterations $K = 10$ in HyPALM, which appears sufficient to reach stability of

the iterates. The time taken by the proposed approach is comparable to the state-of-the art techniques and has been reported in Table 4. All experiments have been performed on a system with Intel(R) Core(TM) i7 x64-based processor and 8 GB RAM. We also illustrate in Figure 1 the practical convergence profile of our algorithm HyPALM, on an example.

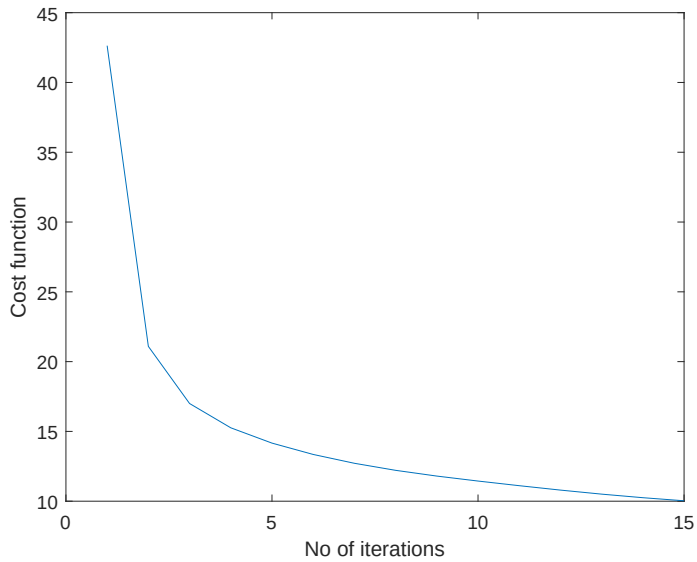

Figure 1: Evolution of the loss function F along the iterations of HyPALM, on the drug-virus association prediction for cross validation setting 1.

Model	GRMC	GRBMC	GRMF	GRDMF-2L	GRDMF-3L
Time	4.327	0.002	0.026	0.07	0.07

Table 4: **Execution times (in seconds) for association prediction for all techniques**.

2.3 Ablation study

In order to quantitatively assess the necessity/importance of each of the similarity models at use, we conducted an ablation study where we used various combinations of similarity (graph regularizations) matrices within the model to predict 10 % of the test association matrix, which were randomly hided as test set (cross validation setting CV1). The results have been reported in Table 5.

Here, $S1_d$ corresponds to chemical structure based similarity for drugs and $S2_d$ corresponds to the similarity based on mechanism of action of drugs. Similarly, $S1_v$ corresponds to the genomic structure based similarity and $S2_v$ corresponds to symptomatic similarity between the viruses.

Similarity	$(S1_d, S1_v)$	$(S1_d, S2_v)$	$(S2_d, S2_v)$	$(S2_d, S1_v)$	$(S1_d + S2_d, S1_v + S2_v)$
AUC	0.8897	0.9010	0.9229	0.9289	0.9457
AUPR	0.6581	0.6790	0.6970	0.6933	0.8180)

Table 5: **AUC and AUPR for association prediction for various combinations of similarities with DMF for all techniques under the cross validation setting CV1**.

Let us make a few observations from this study. Firstly, taking into account the new similarities gives the best prediction results. Secondly, when we use the chemical structure based drug similarity ($S1_d$) and vary the viral similarities from $S1_v$ to $S2_v$ (columns 1 and 2), we observe that symptomatic similarity ($S2_v$) yields better results than genomic similarity ($S1_v$) and hence contributes more towards the prediction task. Similar behavior appears when using the genomic similarity ($S1_v$) for viruses and varying the kind of drug similarity used (columns

1 and 4 of Table 5). In other words, our results assess that the new similarities contribute more towards the drug recommendation and greatly improve the prediction results when taken into account in combination with the previous ones. Lastly, we notice that simply replacing the older similarities with the new ones (column 3 of Table 5) improves the results. Hence, this ablation study establishes the benefits of the new similarity models that we introduced.

2.4 SARS-Cov-2 prediction

In this subsection, we predict top-10 drugs using the best performing solution GRDMF-2L and the baseline algorithms for the novel Coronavirus (SARS-CoV-2). We display the results in Table 6.

Rank	GRDMF-2L	GRMC	GRBMC	GRMF
1	Ribavirin	Remdesivir	Remdesivir	Remdesivir
2	Chloroquine	Ribavirin	Ribavirin	Ribavirin
3	Remdesivir	Umifenovir	Umifenovir	Umifenovir
4	Umifenovir	Telaprevir	Telaprevir	Boceprevir
5	Favipiravir	Boceprevir	Sofosbuvir	Telaprevir
6	Baloxavir marboxil	Palivizumab	Boceprevir	Sofosbuvir
7	Interferon alfa-2a, Recombinant	Sofosbuvir	Palivizumab	Palivizumab
8	Geldanamycin	Simeprevir	Chloroquine	Simeprevir
9	Laninamivir	Taribavirin	Simeprevir	Taribavirin
10	Dolutegravir	Chloroquine	Taribavirin	Paritaprevir

as c

Table 6: **Top-10 ranked recommendations/drugs predicted for SARS-Cov-2 by the DVA computational methods**.

As can be observed, Table 6 reports the drug recommendations for the novel Corona virus (nCov/SARS-CoV-2) using the proposed algorithm and other baselines. The results from our proposed algorithm are sensible. Ribavarin is an FDA approved antiviral being considered as a treatment for COVID19 either by itself [35] or as a part of combination [30]. Remdesivir is a promising candidate for treating COVID19 having shown efficacy against SARS-CoV-1. The antiviral has been approved for emergency use by the US FDA [20, 60]; it has been approved by the drug controller general of India (DGCI) to treat mild to moderate COVID-19 patients 1 [1]. The science behind the mechanism of Remdesivir's action for inhibiting COVID19 has been thoroughly studied [82]. [80] showed that Umifenovir (brand name Arbidol) is effective as a prophyalactic against COVID19 infection. It has also found to be effective in the treatment of COVID affected subjects [75, 75, 72, 79]. DGCI has approved it for phase 3 trials in India [2]. Favipiravir, a Russian drug for influenza, has also been undergoing trials in Japan; it has been sent to 40+ countries for COVID19 trials [3]. [63] explains the scientific reason behind the prospective success of this antiviral. Baloxavir Marboxil has undergone trial in China for treating SARS-COV2 [4]. Interferon alfa-2a in combination with Umifenovir have been found to be effective against COVID19 infected pneumonia [79]. In silico docking studies shows the potential of Baloxavir marboxil [59]; [42] suggests that this antiviral can be effective if administered within 24 hours of the onset of infection. Several in silico docking studies have shown Dolutegravir may be effective in treating COVID19 infection [4, 39]. Trials for Chloroquine and alternately hydroxychloroquine was once suspended by WHO based on a controversial article, but they were resumed within a few days of suspension [5]. This antimalarial drug has been found to be effective against corona in multiple studies [19]; the scientific reason behind the efficacy of this drug against COVID19 has been explained in [28]. Geldanamycin and Laninamivir, predicted by our algorithm has not been considered for treating COVID19 patients.

The benchmark techniques GRMC, GRBMC and GRMF have three drugs common to ours – Ribavarin, Remdesivir and Umifenovir. These three techniques also predict an anti-hepatitis C drug – Sofosbuvir; this drug was known to be effective against the coronavirus family [22]; it's mechanism of action against COVID19 has also been studied [21]. Evidence from clinical studies have corroborated the same [31]. GRMC and GRBMC also predicts chloroquine (as do we); but GRMF is unable to predict it.

[1] https://www.thehindu.com/sci-tech/health/indias-drug-regulator-grants-gilead-sciences-marketing-authorisation-for-remdesivir/article31727034.ece

[2] https://futuremedicineindia.com/umifenovir-gets-nod-for-phase-iii-clinical-trial-against-covid-19-in-india/

[3] https://www.cnbc.com/2020/05/04/fujifilms-flu-drug-favipiravir-sent-to-43-nations-for-covid-19-trials.html

[4] http://www.chictr.org.cn/showprojen.aspx?proj=49015

[5] https://www.newsbytesapp.com/timeline/world/61795/289490/who-set-to-resume-hydroxychloroquine-trials

Note that the results shown in Table 6 are ranked. The ranking is important, since it gives us a faith in the efficacy of the drugs. Our method predicts 6 drugs under trial among the top 7. The benchmarks only recommend 3 drugs in correct order. In fact, had we looked at the top 5 recommendations, only our algorithm would have recommended all of them correctly.

The benchmark algorithms predict three common drugs – Telaprevir [62] , Boceprevir [26] and Simeprevir [44, 70]. These drugs have the potential to be considered for COVID19 clinical trials. All the three benchmarks predict the antibody Palivizumab; there is no evidence in medical literature to support its usage for treating COVID19 patients or using it as a prophylactic. They also predict Taribavirin, which is similar to Ribavirin, and is used for treating Hepatitis-C; it has not been considered for clinical trials for SARS-COV2.

GRMF predicts a unique drug – Paritaprevir. An in silico docking study [62] showed that it has the potential for treating COVID19 much like Telaprevir, Boceprevir and Simeprevir.

The summary of the findings are in the following Table 7.

Recommendations	#predictions used in trial	#predictions having potential	#predictions not used/discontinued
GRDMF-2L	6	2	2
GRMC	5	3	2
GRBMC	5	3	2
GRMF	4	4	2

Table 7: **Top-10 recommendations/drugs predicted for SARS-Cov-2 by the DVA computational methods**.

Only our proposed algorithm predicts 6 drugs that are under trial (Remdesivir, Ribavarin, Umifenovir, Chloroquine, Favipiravir and Interferon alfa-2a). Two of the benchmarks predict 5 drugs that are under trial (Remdesivir, Ribavarin, Umifenovir, Sofosbuvir and Chloroquine); one of them (GRMF) misses out on Chloroquine.

It is encouraging to observe that all the techniques (proposed and benchmarks) are predicting common drugs – Remdesivir, Ribavarin, Umifenovir.

3 Conclusion

AI researchers all across the world are contributing towards fighting COVID19. This is our humble contribution towards the same goal [45, 29]. In this paper, we have proposed a novel optimization approach to solve the drug-virus association prediction problem. The contribution of this work is two-fold. Biologically, we have extended the phenomena of solving drug-virus association prediction for drug re-positioning using multiple similarities (graphs) and introduced new similarity measures for both drugs (based on mechanism of action) and viruses (based on symptoms). Technically, the proposed novel technique brings together the benefits of deep learning (using deep matrix factorization) and multiple graph learning (by regularization with multiple graphs). The algorithmic solution to the problem, relying on the novel HyPALM approach, does not require the setting of any extra hyper-parameter and has sounded convergence guarantees. We have made the software publicly available at `https://github.com/aanchalMongia/DeepVir`.

4 Methods

4.1 Background

4.1.1 Modeling Biological Interactions

A significant amount of research has been done in formulating drug re-positioning as a machine learning task. Various approaches such as network diffusion, supervised classification, neighborhood based prediction, clustering and matrix completion have been used in the past to address the said problem. Of these matrix completion techniques have turned out to be the most promising approach [24]. Various biological applications which model the problem of drug-re-positioning using matrix completion include:

- **Drug target interaction prediction:** The problem is to predict interaction score between a drug and target protein in a matrix assumed to have partial interaction data with drugs on rows and proteins on columns [43, 25, 47, 74, 51, 85].

- **Drug Disease association prediction:** One predicts the probability that a certain drug will interact with a disease or not in a partially filled drug-disease association matrix. [84, 46, 81, 48].
- **Drug-Drug interaction prediction:** The drug-drug relationship is detected through a symmetrical drug-drug matrix/network [83]. This helps predict drugs similar to the ones known to be effective against a pathogen/disease. Attempts have also been made to propose a machine learning approach for predicting drug-likeness of a molecule to distinguish potential drugs from small molecules that lack drug-like features [5].

In addition to drug re-positioning, other biological interaction problems have also be modeled as matrix completion with the common goal of discovering or improving treatment of human diseases, for example we can cite:

- **Gene disease association prediction:** Disease and gene features are used to learn genes for diseases as a gene-disease association prediction [54, 78, 36]. This assists deciphering genetic basis of human diseases and better understanding of gene functions, interactions, and pathways.
- **Protein-protein interaction prediction:** Interactions are estimated in a protein-protein network [77, 38]. Kshirsagar *et al.* [38] deploy a matrix completion variant to model interactions between host (human here) proteins and pathogen (viruses causing infectious diseases here) proteins. This helps identifying the interaction between viral proteins and the human proteins, enabling deeper understanding of infectious diseases (which may involve biologically similar pathogens).

Motivated by such works, the problem of Drug virus association prediction was considered in [52]. Although solvable by any computational approach, the authors propose to tackle it using matrix completion motivated by the success of matrix completion in the above-mentioned applications.

Among all the methodologies compared in [52], graph regularized matrix factorization based technique (GRMF) provided the best results for the validation setting where drugs are predicted for novel viruses. This impels us to come up with an improved algorithm built upon GRMF and leverage the benefit of deep learning and various kinds of metadata associated with drugs/viruses into the current method to aid drug re-positioning for viruses (including SARS-CoV-2).

4.1.2 Matrix completion

The goal of matrix completion is to recover all the entries of a matrix given a subset of known entries. Assume that Y is a masked / undersampled version of the complete matrix X. Then, the problem can be expressed as:

$$Y = M \odot X, \tag{1}$$

Here, M denotes the mask which is element-wise multiplied to X and has ones at positions where the values are known and zeros elsewhere. In general, the problem is under-determined with infinitely many solutions. However, when the sought matrix X is of low-rank, several efficient resolution methods exist. The simplest of them is the matrix factorization approach [15, 25]. Although known to work empirically, a theoretical understanding of factorization based completion is relatively recent [68]. In this approach to matrix completion, X is factored into a product of two low-rank matrices U and V. Formally, this is expressed as:

$$Y = M \odot (UV). \tag{2}$$

The factor matrices (U and V) are obtained by minimizing the following objective function:

$$\text{minimize}_{U,V} \, ||Y - M \odot (UV)||_F^2, \tag{3}$$

with $\|\cdot\|_F$ the Frobenius norm. The setting of the size of the latent factors U and V allow to impose rank constraints on their product.

Motivated by the success of deep learning in numerous fields [64, 67, 61, 53], the shallow (two factor) models have been extended to deeper versions. The general factorization problem when the matrix is completely observed have been proposed in [41, 71]. The solution to matrix completion via deep factorization has been proposed very recently, in [53]. This can be formally expressed as follows, in the 2-layers case:

$$\text{minimize}_{U_1,U_2,V} \, ||Y - M \odot (U_1 U_2 V)||_F^2. \tag{4}$$

Hereagain, the size of the latent factors U_1, U_2 and V control the rank of the resulting product. One can extent to deeper layer formulations (3-layers and 4-layers) by factorizing X into 4 and 5 matrices respectively. The first

factor matrix U_1 (of size $m \times k_1$) contains the representation of the row elements while the last factor matrix V (of size $k_2 \times n$) is associated to its column representations in the latent space. The values for k_1 and k_2 here represent the number of latent factors, and are typically set using cross-validation.

Matrix factorization yields non-convex optimization problems which may be difficult to solve. A more direct approach to low rank matrix completion is to directly estimate matrix X under low-rank constraints. Since the minimization of the rank leads to NP hard problems, one usually replaces it by its convex proxy, the nuclear norm [9, 10], defined as the sum of singular values. The problem formulation is then given as follows:

$$\text{minimize}_X \, ||Y - M \odot X||_F^2 + ||X||_*. \tag{5}$$

In regular matrix completion, the entries in the unknown matrix are assumed to be real-valued. In communications and signal processing, this is actually never the case as those values are almost always quantized. This led to the quantized matrix completion problem [23, 6]. One extreme case is binary matrix completion [17] where the matrix entries are represented by only one bit. Binary matrix completion is particularly appropriate for modeling biological interactions, so unsurprisingly studies in the past have used it for microbe disease association [65].

In all the aforesaid approaches, there was no scope for incorporating associated metadata, to constrain the completion problem and reduce the indeterminacy. In problems of biological imputations, there are often several sources of associated metadata. For example, in drug target interaction, one can compute the SIMCOMP score between the drugs based on their chemical structure as a drug-drug similarity measure [27]. Similarly, similarities between target protein can be found a priori by computing the Smith-waterman score between the amino acid sequences of the target proteins [66]. Studies have shown that the incorporation of associated metadata can improve the prediction results considerably [18]. There can be several approaches to model the associated metadata, in the framework of matrix completion. One of the most promising approaches is to model the associations as graphs. This led to weighted graph regularised matrix factorization [25]. The formulation was expressed as:

$$\text{minimize}_{U,V} \, ||Y - M \odot (UV)||_F^2 + 2\mu \, \text{tr}(U^\top L_r U) + 2\mu \, \text{tr}(V L_c V^\top). \tag{6}$$

Hereabove, $\mu \geq 0$ is the parameter penalizing the graph regularization terms, and $\text{tr}(\cdot)$ denotes the trace of a matrix. Furthermore, the cost function in (10) incorporates Laplacian weight matrices L_r and L_c imposing prior correlation between row entities and column entities, respectively. Those matrices are obtained from known similarity factors S_r and S_c, according to

$$L_r = D_r - S_r, \quad L_c = D_c - S_c. \tag{7}$$

D_r and D_c are diagonal degree matrices for the corresponding similarity matrices S_r or S_c, that is, their i-th diagonal entry equals $D_r^{ii} = \sum_j S_r^{ij}$ and $D_c^{ii} = \sum_j S_c^{ij}$.

We recently proposed a deep formulation for (10) in [50] which amounts to solving:

$$\text{minimize}_{U_1,U_2,V} \, ||Y - M \odot (U_1 U_2 V)||_F^2 + \mu_1 \, \text{tr}(U_1^\top L_r U_1) + \mu_2 \, \text{tr}(V L_c V^\top). \tag{8}$$

Graph regularised version of nuclear norm minimization was also proposed in [33] and was employed for drug target interactions [74]. The formulation is given by:

$$\text{minimize}_X \, ||Y - M \odot X||_F^2 + \lambda ||X||_* + \mu_1 \, \text{tr}(X^\top L_r X) + \mu_2 \, \text{tr}(X L_c X^\top) \tag{9}$$

All the aforementioned graph regularised techniques have been compared in our prior work [52] in the context of drug-virus association prediction.

4.2 Problem Formulation

In this work, we assume the unknown matrix X to be a complete drug-virus association matrix, with drugs along the rows (m drugs) and viruses along the columns (n viruses). It is a binary matrix with each entry x_{ij} either equal to zero, representing no proven association between the corresponding drug and virus or to one, denoting that the ith drug is known to be effective against the jth virus. We set Y as the observed matrix, corresponding to a sampled version from X. The information of the known association positions is encoded in the masking binary operator M.

In our proposed technique, multi-graph regularization is incorporated in the deep matrix factorization formulation with the aim to incorporate the metadata associated with the drugs and viruses in the form of similarity information as shown below:

$$\text{minimize}_{U_1,U_2,V} \, ||Y - M \odot (U_1 U_2 V)||_F^2 + 2\mu \, \text{tr}(U_1^\top L_d U_1) + 2\mu \, \text{tr}(V L_v V^\top), \quad \text{s.t.} \quad U_1 U_2 V \geq 0. \tag{10}$$

Hereabove,

$$L_d = \sum_{\ell=1}^{n_d} L_d^\ell, \quad L_v = \sum_{\ell=1}^{n_v} L_v^\ell. \tag{11}$$

with n_d and n_v the number of available similarity matrices $(S_d^\ell)_\ell$ and $(S_v^\ell)_\ell$, for row entities (i.e. drugs) and column entities (i.e., viruses). The sizes of matrices U_1, U_2 and V are $m \times k_1, k_1 \times k_2$ and $k_2 \times n$ respectively, where $k_1 \geq 1$ and $k_2 \geq 1$ are the numbers of latent factors assumed to be involved. Note that our formulation relies on the observation that computing the Laplacians associated to the sums $\sum_{\ell=1}^{n_d} S_d^\ell$ and $\sum_{\ell=1}^{n_v} S_v^\ell$ is actually equals to the sum of all the individual Laplacians involved [51], given by (11).

In our previous work on drug-virus association prediction [52], we only consider one kind of similarity for drugs (based on the chemical structure of the drugs) and viruses (based on the genomic sequence of the viruses), so that n_d and n_v were equal to one. In this work, we propose to integrate a second type of similarities for drugs and for viruses by taking into account the mechanism of action of drugs and symptoms of viruses, so that $n_d = n_v = 2$.

The sought drug-virus association matrix X to be recovered is assumed to be a positive entry matrix with a low-rank structure as similar drugs are known to affect biological systems (target pathways) in a similar fashion by having a similar mechanism of action [32], which motivates its formulation as the product of rectangular matrices, $X = U_1 U_2 V$ and the addition of the positivity constraint on it.

4.3 Data preprocessing

We deploy the sparsification method to process the similarity matrices (as done in [51, 25]) using p-nearest neighbor graph which is obtained by taking into account only the similarity values which correspond to the nearest neighbors for each drug/virus. p here is set by performing cross-validation on the training set.

4.4 Hybrid Proximal Alternating Linearized Minimization

The minimization problem (10) is non-convex. The multi-linear structure of the involved operators motivates us for using the alternating minimization algorithm from [14]. This algorithm has been shown to be very powerful, in several applications such as seismic signal recovery [58], phase retrieval [56] and hyperspectral imaging [57]. It can be seen as an alternating Majorization-Minimization (MM) technique [69], including the powerful proximity operator [3] which allows to stabilize and speed up the convergence. In order to apply this method, one needs to introduce explicitly the variable X in the problem formulation, as follows:

$$\text{minimize}_{X,U_1,U_2,V}\, F(X, U_1, U_2, V) \quad \text{s.t.} \quad X \geq 0, \tag{12}$$

with

$$F(X, U_1, U_2, V) = ||Y - M \odot X||_F^2 + \vartheta ||X - U_1 U_2 V||_F^2 + 2\mu\, \text{tr}(U_1^\top L_d U_1) + 2\mu\, \text{tr}(V L_v V^\top). \tag{13}$$

The advantage is to remove the complicated nonlinear constraint $U_1 U_2 V \geq 0$ while not deteriorating the solution (for $\vartheta > 0$ sufficiently large). Function F is then the sum of convex terms involving only a single parameter, and a quadratic, thus differentiable, non convex coupling term weighted by ϑ. We are thus in the framework depicted in [14] (see also [7] for a particular case with two variables), so that we can apply the block coordinate variable metric forward-backward to solve Problem (12). This amounts to updating the four variables X, U_1, U_2 and V sequentially, combining gradient (possibly with preconditioning) and proximal updates (see [3] for more details about the proximity operator). As emphasized in [1], the two first quadratic terms can actually be processed either through proximal or gradient descent updates, without altering the convergence of the method. This nice flexibility allows us to propose the following alternating hybrid scheme, that we called HyPALM for the resolution of (12):

- (update of X) Here, we perform a gradient step over the first quadratic term of (12), followed by a proximal step over the remaining terms:

$$B \leftarrow X - \alpha M \odot (M \odot X - Y), \tag{14}$$

$$X \leftarrow \text{minimize}_Z\, \vartheta \|Z - U_1 U_2 V\|_F^2 + \varsigma \|Z - B\|_F^2 \quad \text{s.t.} \quad Z \geq 0, \tag{15}$$

$$= \max\left\{ \frac{\varsigma B + \vartheta U_1 U_2 V}{\varsigma + \vartheta}, 0 \right\}. \tag{16}$$

  Hereabove, $\varsigma > 0$ is a parameter of the algorithm, and α is a stepsize that must be chosen within the interval $]0, 2[$ to secure the convergence of the method. The maximum operation in (16) is taken entrywise, so it is equivalent to a simple cropping of the negative values.

- (update of U_1) We solve the following minimization problem, equivalent to evaluate the proximity operator of F restricted to the variable U_1, at the previous version of U_1 denoted $\overline{U_1}$:

$$U_1 \leftarrow \text{argmin}_{U_1} F(X, U_1, U_2, V) + \varsigma \|U_1 - \overline{U_1}\|_F^2, \tag{17}$$
$$= \text{argmin}_{U_1} \vartheta ||X - U_1 U_2 V||_F^2 + 2\mu \,\text{tr}(U_1^\top L_d U_1) + \varsigma \|U_1 - \overline{U_1}\|_F^2. \tag{18}$$

By taking the gradient of the above expression with respect to U_1 equals to zero, we obtain:

$$(2\mu L_d + \varsigma I) U_1 + \vartheta U_1 (U_2 V)(U_2 V)^\top = X (U_2 V)^\top + \varsigma \overline{U_1}, \tag{19}$$

with I the identity matrix. System (19) takes the form of a Sylvester equation for which efficient solvers are available.

- (update of U_2) Similarly to the latter update, we compute:

$$U_2 \leftarrow \text{argmin}_{U_2} F(X, U_1, U_2, V) + \varsigma \|U_2 - \overline{U_2}\|_F^2, \tag{20}$$
$$= \text{argmin}_{U_2} \vartheta ||X - U_1 U_2 V||_F^2 + \varsigma \|U_2 - \overline{U_2}\|_F^2, \tag{21}$$

with $\overline{U_2}$ the previous version of U_2. This leads to the resolution of the Sylvester equation:

$$\vartheta U_2 V V^\top + \varsigma (U_1^\top U_1)^{-1} U_2 = \vartheta (U_1^\top U_1)^{-1} U_1^\top X V + \varsigma (U_1^\top U_1)^{-1} \overline{U_2}. \tag{22}$$

- (update of V) Hereagain, we solve the minimization problem:

$$V \leftarrow \text{argmin}_V F(X, U_1, U_2, V) + \varsigma \|V - \overline{V}\|_F^2, \tag{23}$$
$$= \text{argmin}_V \vartheta ||X - U_1 U_2 V||_F^2 + 2\mu \,\text{tr}(V L_v V^\top) + \varsigma \|V - \overline{V}\|_F^2, \tag{24}$$

with $\overline{V}$ the previous version of V. This leads to solving the Sylvester equation:

$$\vartheta (U_1 U_2)^\top U_1 U_2 V + V (2\mu L_v + \varsigma I) = \vartheta (U_1 U_2)^\top X + \varsigma \overline{V}. \tag{25}$$

Iterating over the above four steps, with the simplified setting $\varsigma = 1$, leads to Algorithm 1, where we denote $\text{sylv}\{A, B, C\}$ the solution of the Sylvester equation $AX + XB = C$. The convergence of the iterates to a stationary point of Problem (12) is guaranteed as a consequence of the theoretical analysis in [14, 1].

Algorithm 1 HyPALM

Set: $\mu, k_1, k_2, \alpha, L_d, L_v, \vartheta, K$
Initialize $X^{(0)} = Y$ and $(U_1^{(0)}, U_2^{(0)}, V^{(0)})$ (using SVD initialization)
For $k = 1, 2, \ldots, K$

$X^{(k)} = \max\left\{\frac{B + \vartheta U_1^{(k-1)} U_2^{(k-1)} V^{(k-1)}}{1+\vartheta}, 0\right\}$ with $B = X^{(k-1)} + \alpha(M \odot (Y - M \odot (X^{(k-1)})))$

$U_1^{(k)} = \text{sylv}\{2\mu L_d + I, \vartheta U_2^{(k-1)} V^{(k-1)} (U_2^{(k-1)} V^{(k-1)})^\top, \vartheta X^{(k)} (U_2^{(k-1)} V^{(k-1)})^\top + U_1^{(k-1)}\}$

$U_2^{(k)} = \text{sylv}\{((U_1^{(k)})^\top U_1^{(k)})^{-1}, \vartheta V^{(k-1)} (V^{(k-1)})^\top, ((U_1^{(k)})^\top U_1^{(k)})^{-1} U_2^{(k-1)} + \vartheta ((U_1^{(k)})^\top U_1^{(k)})^{-1} (U_1^{(k)})^\top X^{(k)} (V^{(k-1)})^\top\}$

$V^{(k)} = \text{sylv}\{\vartheta (U_1^{(k)} U_2^{(k)})^\top U_1^{(k)} U_2^{(k)}, 2\mu L_v + I, \vartheta (U_1^{(k)} U_2^{(k)})^\top X^{(k)} + V^{(k-1)}\}$

End For
Return: $X^{(K)}$

It should be noted that the solution for deeper layers (3-layers, 4-layers, etc) of HyPALM method for the GRDMF problem can easily be obtained in a similar fashion to the 2-layers framework shown above.

4.5 Similarity computation

In this subsection, we describe in detail how each of the new similarities between drugs and viruses is obtained.

For drugs, we relied on the chemical structure-based SIMCOMP scores [27] to represent drug-drug similarity as in [52]. We also integrate an additional type of similarity by finding cosine similarities between one hot encoded representation of drug class based on mechanism of action of the drug. Similarly, we used the genomic structure-based $d2^*$ distance based on ONF (Oligonucleotide frequency) measure [2] as similarities between the viruses. To obtain the second similarity measure, we calculate the cosine similarity between one hot encoded representations of symptomatic profiles of viruses i.e. symptoms caused by the virus. Both the drug class information and the file encoding which symptoms are caused by the virus (metadata used to generate new similarity measures) are available as supplementary. Hence, the drug similarity matrix (of size 86×86) and virus similarity matrix (of size 23×23) are fixed and encode the metadata available.

References

[1] Feriel Abboud, Emilie Chouzenoux, Jean-Christophe Pesquet, Jean-Hugues Chenot, and Louis Laborelli. A hybrid alternating proximal method for blind video restoration. In *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)*, pages 1811–1815, Lisboa, Portugal, 1-5 Sep. 2014.

[2] Nathan A Ahlgren, Jie Ren, Yang Young Lu, Jed A Fuhrman, and Fengzhu Sun. Alignment-free oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Research*, 45(1):39–53, 2017.

[3] Heinz H Bauschke, Patrick L Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[4] MA Beg and F Athar. Anti-hiv and anti-hcv drugs are the putative inhibitors of rna-dependent-rna polymerase activity of nsp12 of the sars cov-2 (covid-19). *Pharm Pharmacol Int J*, 8(3):163–172, 2020.

[5] Wiktor Beker, Agnieszka Wołos, Sara Szymkuć, and Bartosz A Grzybowski. Minimal-uncertainty prediction of general drug-likeness based on Bayesian neural networks. *Nature Machine Intelligence*, 2(8):457–465, 2020.

[6] Sonia A Bhaskar. Probabilistic low-rank matrix completion from quantized measurements. *The Journal of Machine Learning Research*, 17(1):2131–2164, 2016.

[7] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[8] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Johnathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 2010.

[9] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, 2009.

[10] Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[11] Xing Chen, Lei Wang, Jia Qu, Na-Na Guan, and Jian-Qiang Li. Predicting mirna–disease association based on inductive matrix completion. *Bioinformatics*, 34(24):4256–4265, 2018.

[12] Jie Cheng, Qiang Ye, Hongbo Jiang, Dan Wang, and Chonggang Wang. Stcdg: An efficient data gathering algorithm based on matrix completion for wireless sensor networks. *IEEE Transactions on Wireless Communications*, 12(2):850–861, 2012.

[13] Yao Cheng, Liang Yin, and Yong Yu. Lorslim: low rank sparse linear methods for top-n recommendations. In *2014 IEEE International Conference on Data Mining*, pages 90–99. IEEE, 2014.

[14] Emilie Chouzenoux, Jean-Christophe Pesquet, and Audrey Repetti. A block coordinate variable metric forward-backward algorithm. *Journal of Global Optimization*, 66(3):457–485, 2016.

[15] Murat Can Cobanoglu, Chang Liu, Feizhuo Hu, Zoltan N Oltvai, and Ivet Bahar. Predicting drug-target interactions using probabilistic matrix factorization. *Journal of chemical information and modeling*, 53(12):3399–3409, 2013.

[16] Yan-Fen Dai and Xing-Ming Zhao. A survey on the computational approaches to identify drug targets in the postgenomic era. *BioMed Research International*, 2015.

[17] Mark A Davenport, Yaniv Plan, Ewout Van Den Berg, and Mary Wootters. 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3):189–223, 2014.

[18] Hao Ding, Ichigaku Takigawa, Hiroshi Mamitsuka, and Shanfeng Zhu. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Briefings in Bioinformatics*, 15(5):734–747, 2013.

[19] YJ Duan, Q Liu, SQ Zhao, F Huang, L Ren, L Liu, and YW Zhou. The trial of chloroquine in the treatment of corona virus disease 2019 (covid-19) and its research progress in forensic toxicology. *Fa yi xue za zhi*, 36(2), 2020.

[20] Richard T Eastman, Jacob S Roth, Kyle R Brimacombe, Anton Simeonov, Min Shen, Samarjit Patnaik, and Matthew D Hall. Remdesivir: A review of its discovery and development leading to emergency use authorization for treatment of covid-19. *ACS Central Science*, 2020.

[21] Abdo A Elfiky. Ribavirin, remdesivir, sofosbuvir, galidesivir, and tenofovir against SARS-CoV-2 rna dependent rna polymerase (RdRp): A molecular docking study. *Life Sciences*, page 117592, 2020.

[22] Abdo A Elfiky, Samah M Mahdy, and Wael M Elshemey. Quantitative structure-activity relationship and molecular docking revealed a potency of anti-hepatitis c virus drugs against human corona viruses. *Journal of Medical Virology*, 89(6):1040–1047, 2017.

[23] Ashkan Esmaeili and Farokh Marvasti. A novel approach to quantized matrix completion using huber loss measure. *IEEE Signal Processing Letters*, 26(2):337–341, 2019.

[24] Ali Ezzat, Min Wu, Xiao-Li Li, and Chee-Keong Kwoh. Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Briefings in Bioinformatics*, 2018.

[25] Ali Ezzat, Peilin Zhao, Min Wu, Xiao-Li Li, and Chee-Keong Kwoh. Drug-target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 14(3):646–656, 2017.

[26] Lifeng Fu, Fei Ye, Yong Feng, Feng Yu, Qisheng Wang, Yan Wu, Cheng Zhao, Huan Sun, Baoying Huang, Peihua Niu, et al. Both boceprevir and gc376 efficaciously inhibit sars-cov-2 by targeting its main protease. *Nature Communications*, 11(1):1–8, 2020.

[27] Masahiro Hattori, Nobuya Tanaka, Minoru Kanehisa, and Susumu Goto. Simcomp/subcomp: chemical structure search servers for network analyses. *Nucleic Acids Research*, 38(suppl_2):W652–W656, 2010.

[28] Tony Y Hu, Matthew Frieman, and Joy Wolfram. Insights from nanomedicine into chloroquine efficacy against covid-19. *Nature Nanotechnology*, 15(4):247–249, 2020.

[29] Yipeng Hu, Joseph Jacob, Geoffrey JM Parker, David J Hawkes, John R Hurst, and Danail Stoyanov. The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. *Nature Machine Intelligence*, pages 1–3, 2020.

[30] Ivan Fan-Ngai Hung, Kwok-Cheung Lung, Eugene Yuk-Keung Tso, Raymond Liu, Tom Wai-Hin Chung, Man-Yee Chu, Yuk-Yung Ng, Jenny Lo, Jacky Chan, Anthony Raymond Tam, et al. Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with covid-19: an open-label, randomised, phase 2 trial. *The Lancet*, 395(10238):1695–1704, 2020.

[31] Rodrigo Jácome, José Alberto Campillo-Balderas, Samuel Ponce de León, Arturo Becerra, and Antonio Lazcano. Sofosbuvir as a potential alternative to treat the SARS-CoV-2 epidemic. *Scientific Reports*, 10(1):1–5, 2020.

[32] Erkhembayar Jadamba and Miyoung Shin. A systematic framework for drug repositioning from integrated omics and drug phenotype profiles using pathway-drug network. *BioMed Research International*, 2016, 2016.

[33] Vassilis Kalofolias, Xavier Bresson, Michael Bronstein, and Pierre Vandergheynst. Matrix completion on graphs. *arXiv preprint arXiv:1408.1717*, 2014.

[34] Arnav Kapur, Kshitij Marwah, and Gil Alterovitz. Gene expression prediction using low-rank matrix completion. *BMC bioinformatics*, 17(1):243, 2016.

[35] Jahan S Khalili, Hai Zhu, Nga Sze Amanda Mak, Yongqi Yan, and Yi Zhu. Novel coronavirus treatment with ribavirin: Groundwork for an evaluation concerning covid-19. *Journal of medical virology*, 2020.

[36] Sebastian Köhler, Sebastian Bauer, Denise Horn, and Peter N Robinson. Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*, 82(4):949–958, 2008.

[37] N. Komodakis and J. Pesquet. Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, 32(6):31–54, 2015.

[38] Meghana Kshirsagar, Keerthiram Murugesan, Jaime G Carbonell, and Judith Klein-Seetharaman. Multi-task matrix completion for learning protein interactions across diseases. *Journal of Computational Biology*, 24(6):501–514, 2017.

[39] Vannajan Sanghiran Lee, Wei Lim Chong, Sri Devi Sukumaran, Pivarat Nimmanpipug, Vengadesh Letchumanan, Bey Hing Goh, Learn-Han Lee, Sharifuddin Md Zain, and Noorsaadah Abd Rahman. Computational screening and identifying binding interaction of anti-viral and anti-malarial drugs: Toward the potential cure for sars-cov-2. *Progress in Drug Discovery & Biomedical Science*, 3(1), 2020.

[40] Jian-Qiang Li, Zhi-Hao Rong, Xing Chen, Gui-Ying Yan, and Zhu-Hong You. Mcmda: Matrix completion for mirna-disease association prediction. *Oncotarget*, 8(13):21187, 2017.

[41] Zechao Li and Jinhui Tang. Weakly supervised deep matrix factorization for social image understanding. *IEEE Transactions on Image Processing*, 26(1):276–288, 2017.

[42] Marc Lipsitch, Stanley Perlman, and Matthew K Waldor. Testing covid-19 therapies to prevent progression of mild disease. *The Lancet Infectious Diseases*, 2020.

[43] Yong Liu, Min Wu, Chunyan Miao, Peilin Zhao, and Xiao-Li Li. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS computational biology*, 12(2):e1004760, 2016.

[44] Ho Sing Lo, Kenrie PY Hui, Hei-Ming Lai, Khadija Shahed Khan, Simranjeet Kaur, Zhongqi Li, Anthony KN Chan, Hayley Hei-Yin Cheung, Ka Chun Ng, John Chi Wang Ho, et al. Simeprevir suppresses SARS-CoV-2 replication and synergizes with remdesivir. *bioRxiv*, 2020.

[45] Miguel Luengo-Oroz, Katherine Hoffmann Pham, Joseph Bullock, Robert Kirkpatrick, Alexandra Luccioni, Sasha Rubel, Cedric Wachholz, Moez Chakchouk, Phillippa Biggs, Tim Nguyen, et al. Artificial intelligence cooperation to support the global response to covid-19. *Nature Machine Intelligence*, pages 1–3, 2020.

[46] Huimin Luo, Min Li, Shaokai Wang, Quan Liu, Yaohang Li, and Jianxin Wang. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics*, 34(11):1904–1912, 2018.

[47] Hafez Eslami Manoochehri and Mehrdad Nourani. Predicting drug-target interaction using deep matrix factorization. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE, 2018.

[48] Aanchal Mongia, Emilie Chouzenoux, and Angshul Majumdar. Computational prediction of drug-disease association based on graph-regularized one bit matrix completion. *bioRxiv*, 2020. `https://www.biorxiv.org/content/10.1101/2020.04.02.020891v1.abstract`.

[49] Aanchal Mongia and Angshul Majumdar. Drug-target interaction prediction using multi-graph regularized deep matrix factorization. *bioRxiv*, page 774539, 2019.

[50] Aanchal Mongia and Angshul Majumdar. Deep matrix completion on graphs: Application in drug target interaction prediction. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020)*, pages 1324–1328. IEEE, 2020.

[51] Aanchal Mongia and Angshul Majumdar. Drug-target interaction prediction using multi graph regularized nuclear norm minimization. *Plos One*, 15(1):e0226484, 2020.

[52] Aanchal Mongia, Sanjay Kr Saha, Emilie Chouzenoux, and Angshul Majumdar. A computational approach to aid clinicians in selecting anti-viral drugs for covid-19 trials. *arXiv preprint arXiv:2007.01902*, 2020.

[53] Aanchal Mongia, Debarka Sengupta, and Angshul Majumdar. DeepMc: deep matrix completion for imputation of single-cell RNA-seq data. *Journal of Computational Biology*, 2019.

[54] Nagarajan Natarajan and Inderjit S Dhillon. Inductive matrix completion for predicting gene-disease associations. *Bioinformatics*, 30(12):i60–i68, 2014.

[55] IP O'Carroll and A Rein. Viral nucleic acids. *Encyclopedia of Cell Biology*, page 517, 2016.

[56] Audrey Repetti, Emilie Chouzenoux, and Jean-Christophe Pesquet. A nonconvex regularized approach for phase retrieval. In *Proceedings of the 21st IEEE International Conference on Image Processing (ICIP 2014)*, pages 1753–1757, Paris, France, 27-30 Oct. 2014.

[57] Audrey Repetti, Emilie Chouzenoux, and Jean-Christophe Pesquet. A preconditioned forward-backward approach with application to large-scale nonconvex spectral unmixing problems. In *Proceedings of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, pages 1498–1502, Firenze, Italy, 4-9 May 2014.

[58] Audrey Repetti, Mai Quam Pham, Laurent Duval, Emilie Chouzenoux, and Jean-Christophe Pesquet. Euclid in a taxicab: Sparse blind deconvolution with smoothed l1/l2 regularization. *IEEE Signal Processing Letters*, 22(5):539–543, 2015.

[59] Suban K Sahoo and Seshu Vardhan. Computational evidence on repurposing the anti-influenza drugs baloxavir acid and baloxavir marboxil against covid-19. *arXiv preprint arXiv:2009.01094*, 2020.

[60] Ameet Sarpatwari, Anna Kaltenboeck, and Aaron S Kesselheim. Missed opportunities on emergency remdesivir use. *JAMA*, 324(4):331–332, 2020.

[61] Patrick Schramowski, Wolfgang Stammer, Stefano Teso, Anna Brugger, Franziska Herbert, Xiaoting Shao, Hans-Georg Luigs, Anne-Katrin Mahlein, and Kristian Kersting. Making deep neural networks right for the right scientific reasons by interacting with their explanations. *Nature Machine Intelligence*, 2(8):476–486, 2020.

[62] Bhumi Shah, Palmi Modi, and Sneha R Sagar. In silico studies on therapeutic agents for covid-19: Drug repurposing approach. *Life Sciences*, page 117652, 2020.

[63] Ashleigh Shannon, Barbara Selisko, NTT Le, J Huchting, F Touret, G Piorkowski, V Fattorini, F Ferron, E Decroly, C Meier, et al. Favipiravir strikes the sars-cov-2 at its achilles heel, the rna polymerase. *bioRxiv*, 2020.

[64] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19:221–248, 2017.

[65] Jian-Yu Shi, Hua Huang, Yan-Ning Zhang, Jiang-Bo Cao, and Siu-Ming Yiu. Bmcmda: a novel model for predicting human microbe-disease associations via binary matrix completion. *BMC bioinformatics*, 19(9):85–92, 2018.

[66] Temple F Smith, Michael S Waterman, et al. Identification of common molecular subsequences. *Journal of molecular biology*, 147(1):195–197, 1981.

[67] Richard Socher, Yoshua Bengio, and Christopher D Manning. Deep learning for nlp (without magic). In *Tutorial Abstracts of ACL 2012*, pages 5–5. 2012.

[68] Ruoyu Sun and Zhi-Quan Luo. Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579, 2016.

[69] Ying Sun, Prabhu Babu, and Daniel P Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, 65(3):794–816, 2017.

[70] Alfonso Trezza, Daniele Iovinelli, Annalisa Santucci, Filippo Prischi, and Ottavia Spiga. An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors. *Scientific reports*, 10(1):1–8, 2020.

[71] George Trigeorgis, Konstantinos Bousmalis, Stefanos Zafeiriou, and Björn W Schuller. A deep matrix factorization method for learning attribute representations. *IEEE transactions on pattern analysis and machine intelligence*, 39(3):417–429, 2017.

[72] Naveen Vankadari. Arbidol: A potential antiviral drug for the treatment of SARS-CoV-2 by blocking the trimerization of viral spike glycoprotein? *International Journal of Antimicrobial Agents*, page 105998, 2020.

[73] Evangelos Vlachos, George C Alexandropoulos, and John Thompson. Massive mimo channel estimation for millimeter wave systems via matrix completion. *IEEE Signal Processing Letters*, 25(11):1675–1679, 2018.

[74] Minhui Wang, Chang Tang, and Jiajia Chen. Drug-target interaction prediction via dual laplacian graph regularized matrix completion. *BioMed Research International*, 2018, 2018.

[75] Xi Wang, Ruiyuan Cao, Huanyu Zhang, Jia Liu, Mingyue Xu, Hengrui Hu, Yufeng Li, Lei Zhao, Wei Li, Xiulian Sun, et al. The anti-influenza virus drug, arbidol is an efficient inhibitor of SARS-CoV-2 in vitro. *Cell Discovery*, 6(1):1–5, 2020.

[76] Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of admm in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78(1):29–63, 2019.

[77] Olgica Milenkovic Wei Dai and Narayanan Santhanam Prasad. Low-rank matrix completion for inference of protein-protein interaction networks. In *AIP Conference Proceedings*, volume 1281, pages 1531–1534. American Institute of Physics, 2010.

[78] Xuebing Wu, Rui Jiang, Michael Q Zhang, and Shao Li. Network-based global inference of human disease genes. *Molecular Systems Biology*, 4(1):189, 2008.

[79] Ping Xu, Jianping Huang, Zhao Fan, Wendi Huang, Minghua Qi, Xuwen Lin, Weidong Song, and Li Yi. Arbidol/ifn-α2b therapy for patients with corona virus disease 2019: a retrospective multicenter cohort study. *Microbes and Infection*, 22(4-5):200–205, 2020.

[80] Chunguang Yang, Chunjin Ke, Daoyuan Yue, Wengang Li, Zhiquan Hu, Wei Liu, Shuhua Hu, Shaogang Wang, and Jihong Liu. Effectiveness of arbidol for covid-19 prevention in health professionals. *Frontiers in Public Health*, 8:249, 2020.

[81] Mengyun Yang, Huimin Luo, Yaohang Li, and Jianxin Wang. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics*, 35(14):i455–i463, 2019.

[82] Wanchao Yin, Chunyou Mao, Xiaodong Luan, Dan-Dan Shen, Qingya Shen, Haixia Su, Xiaoxi Wang, Fulai Zhou, Wenfeng Zhao, Minqi Gao, et al. Structural basis for inhibition of the rna-dependent rna polymerase from SARS-CoV-2 by remdesivir. *Science*, 2020.

[83] Wen Zhang, Yanlin Chen, Dingfang Li, and Xiang Yue. Manifold regularized matrix factorization for drug-drug interaction prediction. *Journal of Biomedical Informatics*, 88:90–97, 2018.

[84] Wen Zhang, Xiang Yue, Weiran Lin, Wenjian Wu, Ruoqi Liu, Feng Huang, and Feng Liu. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC bioinformatics*, 19(1):1–12, 2018.

[85] Shuangjia Zheng, Yongjian Li, Sheng Chen, Jun Xu, and Yuedong Yang. Predicting drug–protein interaction using quasi-visual question answering system. *Nature Machine Intelligence*, 2(2):134–140, 2020.