A Matrix Factorization-based Drug-virus Link Prediction Method for SARS-CoV-2 Drug Prioritization

Anonymous Authos

Anonymous Institute

Abstract. Matrix factorization (MF) has been widely used in drug discovery for link prediction, which aims to reveal new drug-target links by integrating drug-drug and target-target similarity information with a drug-target interaction matrix. The MF method is based on the assumption that similar drugs share similar targets and *vice versa*. However, one major disadvantage is that the similarity matrices used in MF models mostly only describe structural similarity which does not entirely represent the similarity between drugs or targets. In this work, we have developed a similarity fusion enhanced MF model to incorporate different types of similarity for novel drug-target link prediction. We have applied the proposed model on a drug-virus association dataset for COVID, and compared the performance with other existing MF models developed for this disease. The results show that the similarity fusion method can provide more information for drug-drug and virus-virus similarity, and hence improve the performance of MF models. Finally, we provide the top 10 drugs as prioritized by our proposed model, and discuss how they have been supported by literature.

Keywords: COVID · drug discovery · matrix factorization · machine learning

1 Introduction

The COVID-19 pandemic [5], caused by Severe Acute Respiratory Syndrome (SARS)-associated coronavirus, SARS-CoV-2, was first reported in Wuhan, China in December 2019, and has spread to most of the countries in the world. The pandemic is considered as a serious global health issue due to its high transmittability and risk of death. By April 2022, World Health Organization (WHO) [1] has reported more than 510 million COVID cases and over 6 million deaths worldwide. Research has shown that a significant proportion of COVID patients can continue to experience a long-term symptoms. [3]. Today quite a few COVID-19 vaccines have been approved for emergency to prevent the transmission of COVID, however effectiveness is still under investigation [6]. At the moment, there is no approved drugs for the treatment of COVID as drug development is a long term process which involves screening hits, compound optimization, and many phases of clinical trials.

Recently, computational methods have been applied to accelerate the drug discovery process [20], and many computational approaches have been employed in COVID-19 drug discovery [23]. Matrix factorization (MF) [15] is a method that decomposes a matrix into a product of two latent matrices in lower dimension. The principle of using MF methods in the development of drugs is to consider the drug target interaction prediction problem as a recommendation system problem, based on the assumption that similar drugs tend to share similar targets, and *vice versa*.

MF methods have been used widely in drug discovery. Zhang *et al.* have proposed a similarity constrained matrix factorization network (SCMFDD) [33] for drug-disease association prediction. The results suggest that SCMFDD outperforms the logistic regression model PREDICT [8], achieving 0.33 in AUPR and 0.92 in AUC. Zhang *et al.* have developed a feature-derived graph regularized matrix factorization (FGRMF) [32] for drug side-effect prediction and reported that FGRMF outperforms restricted Boltzmann machine and collaborative filtering method [34] 0.08 in AUPR and 0.003 in AUC. Zhang *et al.* have introduced a manifold regularized matrix factorization (MRMF) [31] for drug-drug interaction prediction. MRMF is compared with neighbor recommender and random walk [24], and outperforms the two models 0.03 in AUPR and 0.01 in AUC.

MF methods have been widely used in COVID drug prioritization as well [21,27,27,19]. Mongia *et al* [22]have developed a drug-virus interaction dataset, and applied different MF models on it, including Matrix Completion (MC), MF, Deep Matrix Factorization (DMF), Graph Regularize Matrix Factorization (GRMF), Graph Regularized Matrix Completion (GRMC), and Graph Regularized Binary Matrix Completion (GRBMC). Results show that GRMF, GRMC and GRBMC have achieved the best performance.

In this work, we report a similarity fusion network enhanced GRMF model, aiming to discover novel Drug-Virus Interactions (DVI) for COVID-19. Employing various evaluation and validation metrics, we demonstrate that the similarity fusion enhanced GRMF model improves the performance of MF models, and performs better than other MF models that have been applied for COVID. We then select the top 10 prioritized drugs from each model that can be potentially used for COVID treatment, and provide supporting evidence from the literature.

2 Materials

2.1 DVA dataset

The model to prioritize drugs for SARS-CoV-2 is applied to the Drug Virus Association (DVA) [22] dataset, which is a manually curated dataset for COVID containing 216 anti-virual drug-virus associations, comprising 121 drugs and 39 viruses.

2.2 Similarity measurement for drugs and viruses

As a single similarity method may be limited in describing the biological information of drugs and viruses [13] and to provide a more comprehensive description,

we compute the similarity of drugs and targets from multiple perspectives, i.e. besides intrinsic biological properties such as structure-based similarity, we also consider the associated biological entities for measuring the similarity of drugs and viruses, such as pathway similarity, gene expression profile similarity, etc.

Structure-based similarity Following the key concept that chemically similar drugs tend to interact with similar targets [29], two strategies are chosen to measure the structure-based drug-drug similarity, i.e. the SIMCOMP score [11] and fingerprint similarity, respectively. The SIMCOMP score is a graph-based method for comparing chemical structures, where pairwise similarity between molecular structures is computed by converting structures into graphs and then finding the maximal common substructures. In our work SIMCOMP is computed through the KEGG API [14]. Fingerprint similarity describes the presence or absence of substructures in a molecule, and is computed by the Tanimoto coefficient (T_c similarity) [28] (in Equation 1) between the two Extended-connectivity fingerprints (ECFP) [26]:

$$T_c = \frac{|A \cap B|}{|A \cup |B} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{1}$$

Gene expression similarity Gene expression patterns indicate the interaction profile of a drug [12]. We have obtained the druggable genome data from DGIdb [9], and generated gene expression vectors describing the presence of gene names for each drug. The similarity of gene expression is assessed by Pearson's correlation coefficient:

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{2}$$

Sequence-based similarity Virus sequences are obtained from KEGG database, and the virus sequence similarity matrix was measured by computing the d2* Oligonucleotide frequency (ONF) [2] dissimilarity (at $k = 6$) between pairs of viral genome sequences through VirHostMatcher [2].

Pathway similarity Viruses that share similar infectious pathways may be inhibited by similar drugs. We download the pathways for each virus, and convert each pathway into a set of involved genes through KEGG, then compute the T_c similarity for each gene set.

3 Methodology

3.1 Problem description

Given a set of drugs $D = \{d_1, d_2, ..., d_m\}$ and viruses $V = \{v_1, v_2, ..., v_n\}$, where m and n represent the number of drugs and viruses, their interaction Y is a

$m \times n$ binary matrix where $y_{ij} = 1$ indicates interaction between drug d_i and virus v_j while $y_{ij} = 0$ means no interaction. S_d is a $m \times m$ matrix representing the similarity between each drug-drug pair, while S_v is a $n \times n$ matrix representing the similarity between each virus-virus pair. Similarity scores in each matrix are normalized in the range [0, 1]. The objective is to decompose matrix Y to obtain novel (i.e. unknown) DVIs.

3.2 Similarity Fusion

Using a nonlinear equation [17] for similarity fusion is one of the common methods for similarity integration:

$$S_{ij} = 1 - \prod_{k=1}^{n} (1 - s_{ij}^k) \tag{3}$$

where n denotes the number of similarity metrics; s_{ij}^k denotes the the similarity between the i-th and j-th drugs or viruses in the k-th measurement. Here, we use this concept to integrate the metrics outlined above.

3.3 GRMF model for novel drug-virus link prediction

To predict novel drug-virus link, we adopt GRMF [7] and WGRMF [7] on the DVA dataset and the drug-drug and virus-virus similarity fusion matrices. An overview of the workflow is shown in Figure 1. Multiple similarity matrix of drug-drug and virus-virus pairs are computed and integrated as described in **Similarity measurement for drugs and viruses** and **Similarity Fusion**. The similarity fusion matrices are then passed through a pre-processing PNN [7] algorithm which sparsifies the resulting similarity matrix by only keeping the most similar p neighbors for each drug or target and discarding the rest.

The interaction matrix is also pre-processed through a WKNKN [7] algorithm where the DVI values are re-calculated with respect to the interaction profile of a similar drug/virus, in order to reduce the sparsity of the DVI matrix. The pre-processed DVI matrix is then decomposed into two latent matrices in lower dimension through Singular Value Decomposition (SVD) [4], as described below:

$$Y_{m \times n} = u_{n \times n} s_{n \times m} v_{m \times m}^T \tag{4}$$

$$U = u s^{\frac{1}{2}} \tag{5}$$

$$V = v s^{\frac{1}{2}} \tag{6}$$

The objective of GRMF is described as follows:

$$\begin{aligned} \min_{U,V} \quad & \left\| Y - U V^T \right\|_F^2 \\ & + \lambda_l (\|U\|_F^2 + \|V\|_F^2) \\ & + \lambda_d Tr(U^T \tilde{\mathcal{L}}_d U) \\ & + \lambda_v Tr(V^T \tilde{\mathcal{L}}_v V) \end{aligned} \tag{7}$$

where $\|.\|_F^2$ is the Frobenius norm; λ_l, λ_d, and λ_t are positive parameters; $Tr(\cdot)$ is the trace of a matrix, $\mathcal{L}_d$ and $\mathcal{L}_v$ are the graph Laplacians for S_d and S_v. By setting $\frac{\partial L}{\partial U} = 0$ and $\frac{\partial L}{\partial V} = 0$he latent matrices U and V are the updated as follows:

$$U = (YV - \lambda_d \tilde{\mathcal{L}}_d U)(V^T V + \lambda_v I_k)^{-1} \tag{8}$$

$$V = (Y^T U - \lambda_t \tilde{\mathcal{L}}_v V)(U^T U + \lambda_v I_k)^{-1} \tag{9}$$

where k is the number of latent features in U and V. The new DVI matrix is formalized as follows:

$$Y' = UV^T \tag{10}$$

Finally, the new DVI matrix Y' is utilized to derive novel drug-virus links. We have developed the python versions of GRMF and WGRMF, and the software together with the integrated features implemented here are available at https://github.com/yutongLi1997/Similarity-Fusion-GRMF.

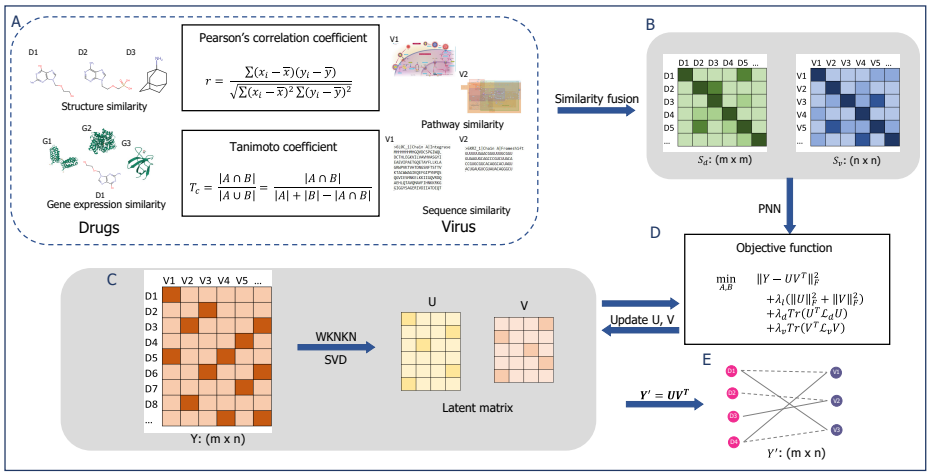


Fig. 1: The flowchart of a similarity fusion network enhanced GRMF model. (A) Different similarity measurements for drug-drug pairs and virus-virus pairs are computed, including structure similarity, gene expression similarity, sequence similarity and pathway similarity. (B) Multiple similarity matrices are integrated into one matrix through a nonlinear equation for similarity fusion. (C) The DVI matrix is pre-processed and decomposed into two lower dimension matrices U and V. (D) The pre-processed drug-drug and virus-virus similarity matrices are used to update the latent matrices U and V. (E) After the pre-set number of iterations has been reached, U and V compose into the predicted DVI matrix.

3.4 Experimental setting and evaluation Metrics

We followed the evaluation steps discussed in [22] to compare the performance with three different variants of 10-fold cross-validation setting for different purposes of discovering new drug-virus pair, discover new drugs, or discover new viruses:

cv_p: to select 10% of the associations between drugs and viruses;
cv_d: to select 10% of the drugs;
cv_v: to select 10% of the viruses.

A train/test mask is then applied to Y where the selected set are set to be 0 in training and 1 in testing.

Area Under the Receiver Operating Characteristic curve (AUC) [10] and Area Under the Precision-Recall curve (AUPR) are selected as evaluation metrics as they are both robust for evaluating imbalanced classes. Both AUC and AUPR can range from 0 to 1. AUC over 0.5 suggests that the model performs better than a random classifier. However the baseline of AUPR is equal to the fraction of positives, and thus AUPR over 0.04 suggests better performance than random classifier.

4 Results

4.1 Model Tuning

For hyper-parameter tuning, we run grid search to determine the best set of hyper-parameters. The detailed parameter setting is as follows:

$k \in \{5, 10, 15, 20, 25, 30, 35, 40\}$;
$\lambda_l \in \{0.25, 0.5, 1, 2\}$;
$\lambda_d, \lambda_v \in \{0, 0.0001, 0.001, 0.01, 0.1\}$.

GRMF and WGRMF achieve best performance at $k = 20$, $\lambda = 0.25$, and $\lambda_d, \lambda_v = 0.1$, which maximizes the contribution of similarity matrices in the loss function, and allows the model to exploit the contribution to its largest extent. This supports our research aim to explore which similarity fusion network can enhance MF model performance. For the other comparison models, we chose the optimal value recommended in the relevant reference.

4.2 Effectiveness of similarity fusion

An example demonstrating the change of similarity values through similarity fusion is given in Figure 2. The integrated similarity matrix has larger values and is more populated than the single similarity matrix, showing that the similarity fusion method can reduce the sparsity of a matrix and thus provide a richer representation of the interaction network.

We have validated the effectiveness of using the similarity fusion method by comparing the performance of GRMF and WGRMF trained with multiple

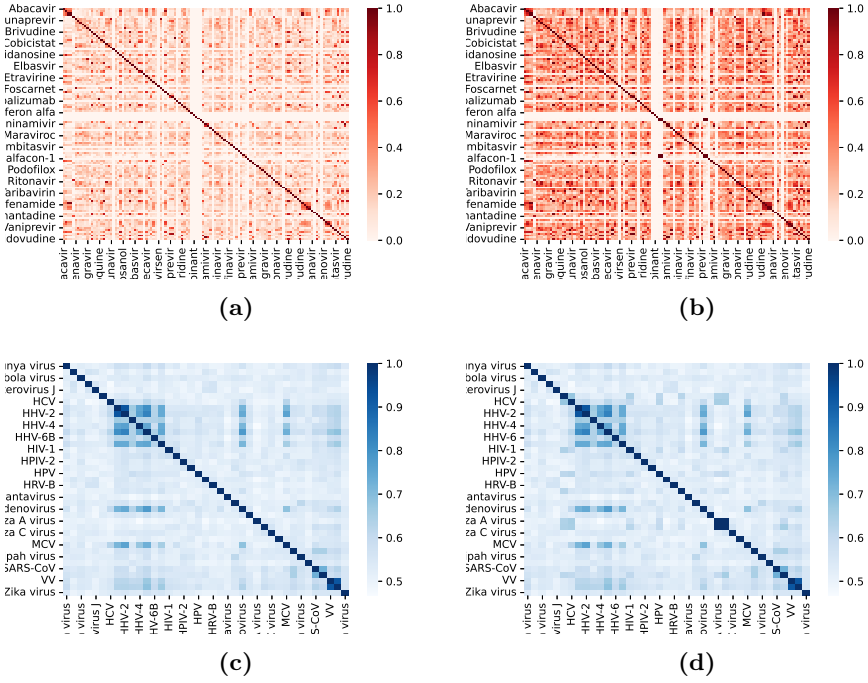
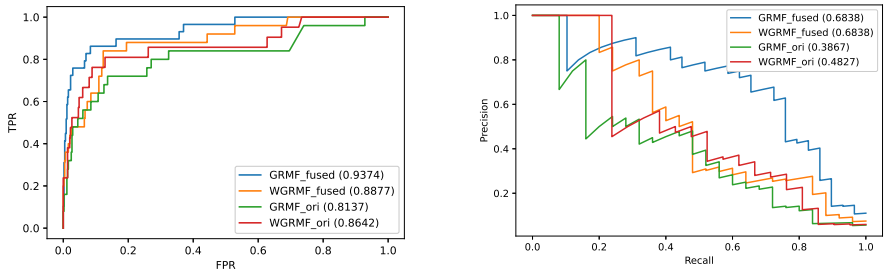


Fig. 2: The heatmap of similarity matrix before and after similarity fusion. (a) and (c) show the single similarity matrix, (b) and (d) show the integrated similarity matrix.

similarity matrices and a single similarity matrix. Figure 3 shows the AUC and AUPR curve of GRMF and WGRMF trained with single and integrated similarity matrices in one test set predicting new drug-virus pair. Both models perform better when having the similarity fusion step, where GRMF and WGRMF are improved 0.12 and 0.02 in AUC, 0.3 and 0.2 in AUPR, respectively, indicating the effectiveness of the similarity fusion network in the enhancement of model performance.



(a) **(b)**

Fig. 3: The drug-virus pairwise performance comparison of GRMF and WGRMF using multiple and single similarity matrix on test set. (a) The AUC curve of the comparison models. (b) The AUPR curve of the comparison models

The complete summary of a 10-fold cross-validation comparison across three cv settings is shown in Table 1. Overall, the similarity fusion network enhanced models perform better than the baseline models, and the GRMF model performs better than WGRMF. Moreover, the models based on similarity fusion networks are relatively more stable than the baseline models, according to the standard deviation across the 10-fold cross validation. In particular, the integrated matrix improves the performance more in predicting new drugs and viruses than new DVI.

4.3 Performance comparison

We have compared the performance of similarity fusion enhanced GRMF and WGRMF with another improved GRMF model, as well as other MF methods that have been used in drug prioritization for SARS-CoV-2, namely, GRDMF [21], GRGMF [35], IRNMF [27], and SCPMF [19]. We have also performed the above models on both integrated similarity matrix and single similarity matrix, to further validate the effectiveness of similarity fusion method. The results are shown in Figure 4.

Overall, the similarity fusion network enhanced GRMF achieves the best performance in AUC score across all three cv settings. For the task of predicting new

Table 1: Cross validation results for GRMF and WGRMF on single and integrated similarity matrix under 3 cross-validation settings

	Metrics	**Single similarity**		**Similarity fusion**	
		GRMF	WGRMF	GRMF	WGRMF
cv_p	AUC	0.859 (±0.057)	0.856 (±0.041)	**0.900 (±0.033)**	0.893 (0.0.029)
	AUPR	0.440 (±0.110)	0.451 (±0.097)	**0.503 (±0.123)**	0.466 (±0.087)
cv_t	AUC	0.776 (±0.078)	0.776 (±0.091)	**0.831 (±0.062)**	0.810 (±0.052)
	AUPR	0.422 (±0.165)	0.441 (±0.144)	0.513 (±0.142)	**0.520 (±0.139)**
cv_v	AUC	0.734 (±0.145)	0.680 (±0.183)	**0.751 (±0.130)**	0.734 (±0.144)
	AUPR	0.241 (±0.153)	**0.322 (±0.178)**	0.270 (±0.220)	0.275 (±0.198)

DVI, GRMF outperforms GRDMF, WGRMF, GRGMF, IRNMF, and SCPMF 0.03, 0.01, 0.14, 0.14, and 0.25, respectively. For predicting new drugs, it outperforms the five models 0.05, 0.02, 0.13, 0.1, and 0.17, respectively. For predicting new viruses, GRMF outperforms the five models 0.08, 0.02, 0.16, 0.15, and 0.14, respectively. GRGMF performs the best in AUPR.

From a general perspective, most models are improved by the similarity fusion method. For predicting new DVI, GRMF, GRDMF, WGRMF, GRGMF, IRNMF, and SCPMF have improved by 0.05, 0.03, 0.04, 0.001, 0.03, and 0.09, respectively. For predicting new drugs, the above models, excluding GRGMF, have improved by 0.06, 0.02, 0.04, 0.02, and 0.02, respectively. For predicting new viruses, the above models have improved by 0.02, 0.01, 0.05, 0.04, 0.01, and 0.05, respectively. This indicates that the similarity fusion network works well with general MF methods, and can enhance the model performance by providing more information for link prediction.

4.4 COVID drug prioritization

Finally, we repurposed existing anti-viral drugs with the similarity fusion network enhanced GRMF and WGRMF for the treatment of SARS-CoV-2, as listed in Table 2. The drugs prioritized by both models show a certain agreement, including Ribavirin, Umifenovir, Baloxavir marboxil, Zanamivir, Lamivudine, Favipiravir, Tenofovir alafenamide, and Triazavirin. Some of the drugs have already been investigated by other research groups, and have shown great potency for COVID patients. Umifenovir exhibits statistically significant efficacy for mild-asymptomatic patients [25]. Favipiravir is considered to have strong possibility for treating mild-to-moderate illness COVID patients [18]. Peramivir is proposed to be a candidate for the treatment of COVID-19 and other infections related cytokine storm syndrome [30]. Triazavirin is suggested to effectively block both the entry of the pathogen into a host cell and its replication [16].

5 Conclusion

In this study, we have developed a similarity fusion enhanced GRMF model to incorporate more information from drugs and viruses when predicting new drug-

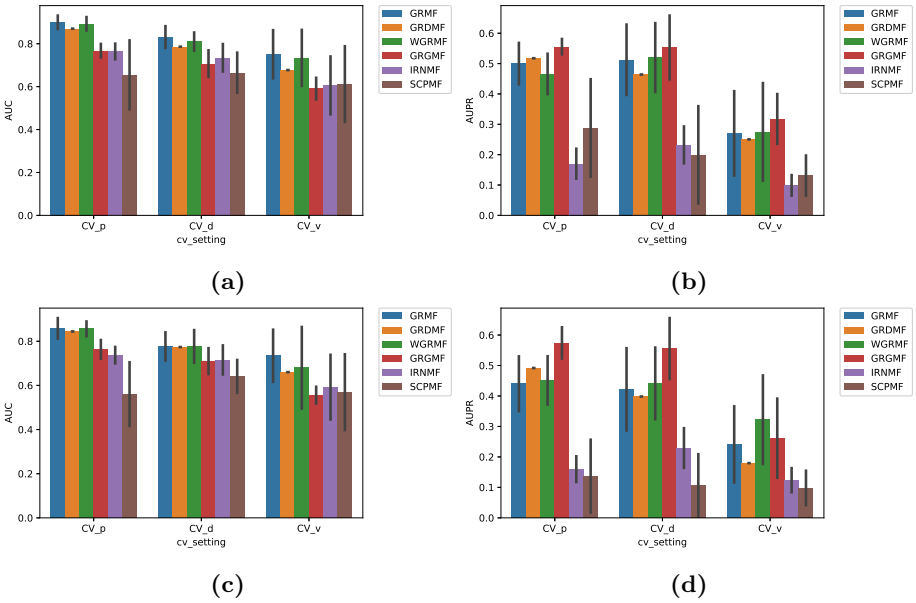


Fig. 4: Cross-validation results for model comparison on three cv settings. (a) and (b) show the mean and standard deviation of AUC and AUPR for similarity fusion enhanced GRMF, GRDMF, WGRMF, GRGMF, IRNMF, and SCPMF. (c) and (d) show the mean and standard deviation of AUC and AUPR for the baseline model of the above models.

Table 2: Drug Prioritization for SARS-CoV-2

GRMF	**WGRMF**
Ribavirin	Umifenovir
Umifenovir	Baloxavir marboxil
Baloxavir marboxil	Ribavirin
Zanamivir	Lamivudine
Favipiravir	Tenofovir alafenamide
Triazavirin	Elvucitabine
Oseltamivir	Tenofovir disoproxil fumarate
Laninamivir	Zanamivir
Peramivir	Favipiravir
Tenofovir alafenamide	Triazavirin

virus links. To evaluate the performance of the model in repurposing drugs for COVID-19 treatment, we have applied the proposed model, as well as other MF models that have been developed for COVID, on the DVA dataset and run a 10-fold cross-validation on three different cv settings. The results show that the similarity fusion method is able to improve the performance of MF models in general, and the proposed model achieves the best performance among all the MF models that have been developed for COVID. We have then suggested the top 10 drugs that the model has prioritized for the treatment of COVID, some of which have shown the potency and been suggested by other research as well.

References

1. Who coronavirus (covid-19) dashboard, `https://covid19.who.int/`
2. Ahlgren, N.A., Ren, J., Lu, Y.Y., Fuhrman, J.A., Sun, F.: Alignment-free oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic acids research **45**(1), 39–53 (2017)
3. Aiyegbusi, O.L., Hughes, S.E., Turner, G., Rivera, S.C., McMullan, C., Chandan, J.S., Haroon, S., Price, G., Davies, E.H., Nirantharakumar, K., et al.: Symptoms, complications and management of long covid: a review. Journal of the Royal Society of Medicine **114**(9), 428–442 (2021)
4. Björnsson, H., Venegas, S.: A manual for eof and svd analyses of climatic data. CCGCR Report **97**(1), 112–134 (1997)
5. Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W.C., Wang, C.B., Bernardini, S.: The covid-19 pandemic. Critical reviews in clinical laboratory sciences **57**(6), 365–388 (2020)
6. Dolgin, E.: Omicron is supercharging the covid vaccine booster debate. Nature **10** (2021)
7. Ezzat, A., Zhao, P., Wu, M., Li, X.L., Kwoh, C.K.: Drug-target interaction prediction with graph regularized matrix factorization. IEEE/ACM transactions on computational biology and bioinformatics **14**(3), 646–656 (2016)
8. Gottlieb, A., Stein, G.Y., Ruppin, E., Sharan, R.: Predict: a method for inferring novel drug indications with application to personalized medicine. Molecular systems biology **7**(1), 496 (2011)
9. Griffith, M., Griffith, O.L., Coffman, A.C., Weible, J.V., McMichael, J.F., Spies, N.C., Koval, J., Das, I., Callaway, M.B., Eldred, J.M., et al.: Dgidb: mining the druggable genome. Nature methods **10**(12), 1209–1210 (2013)
10. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology **143**(1), 29–36 (1982)
11. Hattori, M., Tanaka, N., Kanehisa, M., Goto, S.: Simcomp/subcomp: chemical structure search servers for network analyses. Nucleic acids research **38**(suppl_2), W652–W656 (2010)
12. Hizukuri, Y., Sawada, R., Yamanishi, Y.: Predicting target proteins for drug candidate compounds based on drug-induced gene expression data in a chemical structure-independent manner. BMC medical genomics **8**(1), 1–10 (2015)
13. Huang, L., Luo, H., Li, S., Wu, F.X., Wang, J.: Drug–drug similarity measure and its applications. Briefings in Bioinformatics **22**(4), bbaa265 (2021)

14. Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., et al.: Kegg for linking genomes to life and the environment. Nucleic acids research **36**(suppl_1), D480–D484 (2007)
15. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
16. Kováč, I.M.J.Č.G., Hudecová, M.P.L.: Triazavirin might be the new hope to fight severe acute respiratory syndrome coronavirus 2 (sars-cov-2). Ceska a Slovenska farmacie: casopis Ceske farmaceuticke spolecnosti a Slovenske farmaceuticke spolecnosti **70**(1), 18–25 (2021)
17. Liu, H., Sun, J., Guan, J., Zheng, J., Zhou, S.: Improving compound–protein interaction prediction by building up highly credible negative samples. Bioinformatics **31**(12), i221–i229 (2015)
18. Manabe, T., Kambayashi, D., Akatsu, H., Kudo, K.: Favipiravir for the treatment of patients with covid-19: a systematic review and meta-analysis. BMC infectious diseases **21**(1), 1–13 (2021)
19. Meng, Y., Jin, M., Tang, X., Xu, J.: Drug repositioning based on similarity constrained probabilistic matrix factorization: Covid-19 as a case study. Applied soft computing **103**, 107135 (2021)
20. Mohamed, K., Yazdanpanah, N., Saghazadeh, A., Rezaei, N.: Computational drug discovery and repurposing for the treatment of covid-19: a systematic review. Bioorganic chemistry **106**, 104490 (2021)
21. Mongia, A., Jain, S., Chouzenoux, E., Majumdar, A.: Deepvir: Graphical deep matrix factorization for in silico antiviral repositioning—application to covid-19. Journal of Computational Biology (2022)
22. Mongia, A., Saha, S.K., Chouzenoux, E., Majumdar, A.: A computational approach to aid clinicians in selecting anti-viral drugs for covid-19 trials. Scientific reports **11**(1), 1–12 (2021)
23. Muratov, E.N., Amaro, R., Andrade, C.H., Brown, N., Ekins, S., Fourches, D., Isayev, O., Kozakov, D., Medina-Franco, J.L., Merz, K.M., et al.: A critical overview of computational approaches employed for covid-19 drug discovery. Chemical Society Reviews (2021)
24. Park, K., Kim, D., Ha, S., Lee, D.: Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks. PloS one **10**(10), e0140816 (2015)
25. Ramachandran, R., Bhosale, V., Reddy, H., Atam, V., Faridi, M., Fatima, J., Shukla, V., Khan, Z.A., Khan, H., Singh, V., et al.: Phase iii, randomized, double-blind, placebo controlled trial of efficacy, safety and tolerability of antiviral drug umifenovir vs standard care of therapy in non-severe covid-19 patients. International Journal of Infectious Diseases **115**, 62–69 (2022)
26. Rogers, D., Hahn, M.: Extended-connectivity fingerprints. Journal of chemical information and modeling **50**(5), 742–754 (2010)
27. Tang, X., Cai, L., Meng, Y., Xu, J., Lu, C., Yang, J.: Indicator regularized non-negative matrix factorization method-based drug repurposing for covid-19. Frontiers in Immunology **11**, 3824 (2021)
28. Tanimoto, T.T.: Elementary mathematical theory of classification and prediction (1958)
29. Vilar, S., Hripcsak, G.: The role of drug profiles as similarity metrics: applications to repurposing, adverse effects detection and drug–drug interactions. Briefings in Bioinformatics **18**(4), 670–681 (2017)

30. Zhang, C.x., Tu, Y., Kong, D.p., Li, Y., Chen, D.g., Zhang, W.n., Su, L., Zhuang, C.l., Wang, Z.b.: Peramivir, an anti-influenza virus drug, exhibits potential anti-cytokine storm effects. bioRxiv (2020)
31. Zhang, W., Chen, Y., Li, D., Yue, X.: Manifold regularized matrix factorization for drug-drug interaction prediction. Journal of biomedical informatics **88**, 90–97 (2018)
32. Zhang, W., Liu, X., Chen, Y., Wu, W., Wang, W., Li, X.: Feature-derived graph regularized matrix factorization for predicting drug side effects. Neurocomputing **287**, 154–162 (2018)
33. Zhang, W., Yue, X., Lin, W., Wu, W., Liu, R., Huang, F., Liu, F.: Predicting drug-disease associations by using similarity constrained matrix factorization. BMC bioinformatics **19**(1), 1–12 (2018)
34. Zhang, W., Zou, H., Luo, L., Liu, Q., Wu, W., Xiao, W.: Predicting potential side effects of drugs by recommender methods and ensemble learning. Neurocomputing **173**, 979–987 (2016)
35. Zhang, Z.C., Zhang, X.F., Wu, M., Ou-Yang, L., Zhao, X.M., Li, X.L.: A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. Bioinformatics **36**(11), 3474–3481 (2020)