Does adding the drug-drug similarity to drug-target interaction prediction methods make a noticeable improvement in their efficiency?

Reza Hassanzadeh (r.hassanzadeh@uma.ac.ir)
University of Mohaghegh Ardabili
Soheila Shabani-Mashcool
University of Tehran

Research Article

Keywords: Drug-Target Interaction, Drug Repositioning, Machine Learning, Drug-Drug Similarity

Posted Date: May 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1533175/v2

Abstract

Predicting drug-target interactions has become an important bioinformatics issue because it is one of the critical and preliminary stages of drug repositioning. Therefore, scientists are trying to develop more accurate computational methods for predicting drug-target interactions. These methods are usually based on machine learning or recommender systems and use biological and chemical information to improve the accuracy of predictions. In the background of these methods, there is a hypothesis that similar drugs have similar targets. So, the similarity between drugs as chemical information is added to the computational methods to improve the prediction results. The question that arises here is whether this claim is actually true? If so, what method should we use to calculate drug-drug similarities? Will we obtain the same improvement from any method we use?

Here, we want to investigate this claimed positive effect. For this purpose, we consider different types of real similarities, random similarities, four important datasets and four state-of-the-art methods. Our results show that the type of data, the method which is used to predict the interactions, and the algorithm used to calculate the similarities are all important, and it cannot be easily stated that adding drug-drug similarities can significantly improve the results. Therefore, our results could suggest a checklist for scientists who want to improve their machine learning methods.

Introduction

Most drugs fail in the early stages of a clinical trial and it takes a lot of time and cost for a drug to be successful in the market [1, 2]. These factors have led scientists to work on better and cheaper ways to find suitable drugs. One of the most effective and interesting solutions to solve these problems is drug repositioning (also called drug repurposing). It is true that drug repositioning, by eliminating the early stages of drug design, can speed up research, but it also has drawbacks. For example, determining the dosage of a drug that is considered for a new disease using drug repositioning is one of the most important challenges of this viewpoint because the drug has already been considered for another disease with a specific dose. However, this viewpoint has found its place and we have to consider it today.

One of the most important steps of drug repositioning is identifying Drug-Target Interactions (DTI), which is a difficult task if laboratory and traditional methods are used. In contrast, computational methods can be more effective both in terms of time and cost. These methods can identify or predict DTI more quickly. The computational methods are usually based on machine learning or recommender systems. In order to predict interactions, these methods first consider a mathematical model for the information in the databases and then add biological and/or chemical information to the model, according to guilt by association principle. For example, NRLMF [3], NetLapRLS [4], BLM-NII [5], WNN-GIP [6] and DT-Hybrid [7] are some of the well-known methods in this field.

NRLMF is a matrix factorization approach that predicts the probability that a drug would interact with a target. In this method, the properties of a drug and a target are represented by two latent vectors in the shared low dimensional latent space, respectively [3]. NetLapRLS is a semi-supervised learning method based on Laplacian regularized least square. NetLapRLS, by incorporating a new kernel established from the known drug-protein interaction network, is actually an improvement of the LapRLS [4, 8]. The bipartite local model (BLM) is a supervised learning approach introduced by Bleakley and Yamanishi in 2009 [9]. To improve the BLM, Mei et al. presented a simple procedure called neighbor-based interaction-profile inferring (NII) and integrated it into the existing BLM method and called it BLM-NII [5]. WNN-GIP is actually a combination of a simple weighted nearest neighbor algorithm and the GIP method [6, 10]. An example of recommender systems method introduced for DTI prediction problem is DT-Hybrid. It is a network-based interface method that extends a well-established recommendation technique by domain-based knowledge including drug and target similarity [7]. Many other algorithms have been introduced for this problem, but the algorithms mentioned are the most popular and can be considered as the state-of-the-art methods in this field.

As mentioned, the methods first model the information in databases. There are some public databases, for example, KEGG [11], PubChem [12], DrugBank [13], and ChEMBL [14] that contain information about drugs, targets, and interactions between them. Usually, all methods introduced for predicting DTI interactions use DrugBank to evaluate their results or compare them to other methods. Regardless of what algorithm each of these methods uses, they all add similarity between targets and similarity between drugs to improve the prediction. The similarities between targets (proteins) are always calculated by the Smith-Waterman method [15] and the chemical structure similarities between drugs are usually computed with SIMCOMP [16] which has been implemented in the KEGG system for searching similar chemical structures in the chemical structure databases. SIMCOMP is a graph-based method and uses a graph alignment algorithm to get a global similarity score based on the size of the common substructures between two compounds [5]. Of course, other information such as molecular fingerprints can be used to calculate similarities between drugs, but it is not usually used. There are several types of molecular fingerprints (e.g., MACCS [17], PubChem fingerprint [18], BCI fingerprints [19] and TGD [20]). PubChem fingerprints are 2D fingerprints that make a drug to be expressed by a vector and used to discover similar conformers by the PubChem database. These fingerprints are very popular and easily calculated for every drug.

The purpose of this study is not to identify the best method. Here, we want to discuss the following questions specifically for the DTI prediction problem:

- Does considering the similarity of drugs indeed improve the results of computational methods?
- Is SIMCOMP the best way to calculate drug-drug similarities in any computational method?
- Do the type and size of the dataset affect the improvement that occurs with adding drug-drug similarities?

Experiments

Datasets

Yamanishi et al. have provided four benchmark drug-target interaction datasets including Nuclear Receptors, G-Protein Coupled Receptors (GPCR), Ion Channels, and Enzymes [21]. The datasets are publicly available at http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/. The interaction were retrieved from databases KEGG BRITE [22], BRENDA [23], SuperTarget [24], and DrugBank [13]. Each database contains three types of information in the form of matrices: the drug-target interaction matrix, the drug-drug similarity matrix calculated by SIMCOMP [16], and the target-target similarity matrix obtained by Smith-Waterman method [15]. Some properties of datasets are shown in Table 1.

The abbreviations in Table 1 are as follows:

- N_D: Number of drugs.
- N_T: Number of targets.
- N_I: Number of interactions.

$$Density = \frac{N_I}{N_D \times N_T}$$

- AD_T: Average number of drugs per target.
- AT_D: Average number of targets per drug.
- D_{1T}: Percentage of drugs with only one target.
- T_{1D}: Percentage of targets with only one drug.

Table 1. The properties of the benchmark datasets.

Dataset	Nuclear Receptors	GPCR	Ion Channels	Enzymes
N_D	54	223	210	445
N_T	26	95	204	664
N_I	90	635	1476	2926
Density	0.0641	0.0299	0.0344	0.0099
AD_T	3.46	6.68	7.24	4.41
AT_D	1.67	2.85	7.03	6.58
D_{1T}	72.22%	47.53%	38.57%	39.78%
T_{1D}	30.77%	35.79%	11.27%	43.37%

Evaluation

For each data set, in addition to the default drug similarity matrix obtained by SIMCOMP, we calculated 104 other matrices including one hundred random similarity matrices, one matrix where every element is equal to one, and three matrices calculated from PubChem fingerprint using Tanimoto coefficient, Dice coefficient and Cosine similarity. For the fingerprints of two drugs A and B, the Tanimoto, Dice and Cosine similarity can be calculated as follows:

$$Tanimoto(A, B) = \frac{c}{a + b - c}$$

,

$$Cosine(A, B) = \frac{c}{\sqrt{ab}}$$

,

$$Dice(A, b) = \frac{2c}{a + b}$$

,

where, a equals the amount of bit set to 1 in A, b equals the amount of bits set to 1 in B and c equals the amount of bits set to 1 in both A and B. We considered a matrix where every element is equal to one to actually find out what happens to the results of the algorithms if the similarity of the drugs is not affected.

In order to evaluate, to make the comparison fair, we consider four state-of-the-art methods NRLMF, NetLapRLS, BLM-NII and WNN-GIP. Therefore, in short, we executed every algorithm on every dataset using every drug similarity matrix. To do this, we slightly modified the PyDTI package [3] to perform the evaluation. Like most studies in this field, results are assessed using the area under the ROC curve (AUC) and the area under the precision-recall curve (AUPR). Similar to [3, 4, 6, 10], we performed 10-fold CV for five times to evaluate the performance of the methods on datasets. Then, we calculated the average AUC and AUPR over the five repetitions. In the next section, we will illustrate the results of the evaluations.

Results And Discussion

Before discussing the results, it is necessary to state some of the abbreviations given in the tables and figures as follows:

- All-onesSim: The value obtained for the matrix where every element is equal to one.
- MeanRandoms: The average value obtained for random matrices.
- BestRandom: The best value obtained for the random matrices.
- WorstRandom: The worst value obtained for the random matrices.
- CosinePF: The value obtained for the matrix calculated by cosine similarity for the PubChem fingerprint.
- DicePF: The value obtained for the matrix calculated by Dice similarity for the PubChem fingerprint.
- TanimotoPF: The value obtained for the matrix calculated by Tanimoto similarity for the PubChem fingerprint.

The evaluation results on Enzyme, GPCR, Ion Channel and Nuclear Receptors datasets are shown in Tables 2, 3, 4 and 5, respectively. It is worth noting that the best parameters for each algorithm are obtained in [3], and we have used these parameters here as well.

Table 2
Comparing different drug-drug similarities on Enzyme dataset

	Method	All-onesSim	MeanRandoms	BestRandom	WorstRandom	CosinePF	DicePF	TanimotoPF	SIMCOMP
AUC	NRLMF	0.971239	0.968016	0.969477	0.966555	**0.976691**	0.97665	0.975809	0.97632
	BLM-NII	0.977584	0.7542	0.80648	0.718349	0.977368	0.977764	0.976215	0.969431
	NetLapRLS	0.959789	0.96367	0.964636	0.962813	0.966335	0.966613	0.968903	**0.972169**
	WNN-GIP	0.938578	0.515036	0.524567	0.507309	0.914265	0.897733	0.875283	**0.964062**
AUPR	NRLMF	0.84053	0.841717	0.845043	0.839261	0.870242	0.870329	0.870117	0.875611
	BLM-NII	0.592729	0.023396	0.034385	0.019237	0.605514	0.60798	0.535238	**0.703746**
	NetLapRLS	0.784019	0.787323	0.787526	0.787082	0.789326	0.789748	0.791864	**0.794216**
	WNN-GIP	0.476497	0.011065	0.01174	0.010693	0.256565	0.281493	0.243454	**0.69719**

Table 3
Comparing different drug-drug similarities on GPCR dataset

	Method	All-onesSim	MeanRandoms	BestRandom	WorstRandom	CosinePF	DicePF	TanimotoPF	SIMCOMP
AUC	NRLMF	0.932221	0.922694	0.929836	0.917277	0.956879	0.957188	0.95682	0.960355
	BLM-NII	**0.94386**	0.671366	0.692179	0.647643	0.934594	0.928518	0.879454	0.943664
	NetLapRLS	0.902196	0.90289	0.905388	0.896996	0.910593	0.910846	0.91363	**0.914909**
	WNN-GIP	0.872255	0.528443	0.540304	0.517898	0.804141	0.787323	0.901193	**0.933079**
AUPR	NRLMF	0.570196	0.62361	0.642159	0.60265	0.69302	0.689631	0.688301	0.702622
	BLM-NII	0.373081	0.054418	0.062578	0.046693	0.342311	0.33531	0.324491	**0.514827**
	NetLapRLS	0.606391	0.611795	0.612422	0.611115	0.613065	0.613264	**0.615776**	0.615446
	WNN-GIP	0.278136	0.033394	0.035729	0.031326	0.2326	0.230504	0.428247	**0.466361**

Table 4
Comparing different drug-drug similarities on Ion Channels dataset

	Method	All-onesSim	MeanRandoms	BestRandom	WorstRandom	CosinePF	DicePF	TanimotoPF	SIMCOMP
AUC	NRLMF	0.979234	0.975785	0.977846	0.973896	0.981475	0.980925	0.980701	0.983564
	BLM-NII	0.974675	0.702874	0.744834	0.672745	0.96044	0.958388	0.944077	**0.981287**
	NetLapRLS	0.958158	0.95734	0.957955	0.956605	0.959433	0.959498	0.959527	**0.959882**
	WNN-GIP	0.861103	0.525954	0.535912	0.516046	0.930855	0.919196	0.944477	**0.956789**
AUPR	NRLMF	0.865326	0.856477	0.863683	0.847016	0.864683	0.85956	0.858608	0.863386
	BLM-NII	0.521158	0.058516	0.068181	0.051707	0.484567	0.482101	0.636176	**0.821476**
	NetLapRLS	0.81846	0.820111	0.820284	0.819911	0.821028	0.821095	0.821819	**0.823003**
	WNN-GIP	0.34916	0.038653	0.04019	0.037466	0.53947	0.524961	0.594643	**0.667893**

In each row of tables, the best similarity matrix for each algorithm is bolded. The best AUC and AUPR are also marked with underlines. The first point about these tables is that the use of random matrices has degraded the efficiency of the methods. In fact, what the first four columns of the tables show is that ignoring the drug-drug similarities yields far better results than using inaccurate drug-drug similarities. It should be noted that the NRLMF and NetLapRLS have less tolerance than other methods in this case. Although the purpose of this study is not to identify a better method, but in most cases, the performance of NRLMF is better than other methods. Of course, this performance is due to its many parameters. In the Enzyme dataset (Table 2), the AUPR value for all methods and the AUC value for NetLapRLS and WNN-GIP methods are the best values when SIMCOMP similarity is considered. The NRLMF and BLM-NII methods obtain the best AUC value if they use the CosinePF and DicePF similarities, respectively. In the GPCR dataset (Table 3), the AUC for BLM-NII and the AUPR for NetLapRLS are the best values if they use the All-onesSim and TanimotoPF similarities, respectively. Except for these two cases, according to Table 3, the use of SIMCOMP has given the best results in all cases. Table 4 shows that, in the Ion Channels dataset, using All-onesSim for the NRLMF method leads to a better AUPR. In all other cases, it is clear that SIMCOMP is the best.

Table 5
Comparing different drug-drug similarities on Nuclear Receptors dataset

	Method	All-onesSim	MeanRandoms	BestRandom	WorstRandom	CosinePF	DicePF	TanimotoPF	SIMCOMP
AUC	NRLMF	0.889655	0.864416	0.887016	0.832639	0.937526	0.945632	0.945968	0.948522
	BLM-NII	0.775846	0.580958	0.613693	0.537734	0.797103	0.803759	0.896945	**0.905075**
	NetLapRLS	0.79702	0.802193	0.819461	0.77738	0.823197	0.824621	0.835049	**0.849627**
	WNN-GIP	0.810938	0.541304	0.591313	0.504229	0.900681	0.898618	**0.90459**	0.90394
AUPR	NRLMF	0.515368	0.499308	0.564758	0.427307	0.720545	0.728063	0.726034	0.722834
	BLM-NII	0.391072	0.123816	0.178884	0.087529	0.485113	0.495231	0.63054	**0.659326**
	NetLapRLS	0.428803	0.430737	0.437371	0.421767	0.444117	0.445235	0.454609	**0.464816**
	WNN-GIP	0.317686	0.095114	0.118977	0.079856	0.581542	0.584819	**0.590779**	0.582391

In the Nuclear Receptors dataset (Table 5), the SIMCOMP gives both the best AUC and AUPR for NetLapRLS and BLM-NII methods. The same thing happens with TanimotoPF and WNN-GIP. The AUC and AUPR values for NRLMF are the best if it uses the SIMCOMP and DicePF similarities, respectively. In summary, these tables show that in almost 94% of experiments, the use of drug-drug similarities has led to better results.

So far we have seen that drug-drug similarities can increase the accuracy of DTI predictions. But which method of calculating drug-drug similarities is more appropriate for the DTI predictions problem? The answer shown in Tables 2–5 is clearly SIMCOMP. But the results shown in these tables are obtained by parameters tuned for SIMCOMP [3]. Therefore, we randomly selected a dataset for each method and tuned the parameters of that method for all drug-drug similarities except random similarities. Nuclear Receptors, GPCR, Ion Channel and Enzyme datasets were considered for NRLMF, NetLapRLS, WNN-GIP and BLM-NII methods respectively. The results of these experiments are illustrated in Fig. 1. The use of SIMCOMP for NetLapRLS and WNN-GIP methods gives the best AUC in GPCR and Ion Channel datasets, respectively. The AUCs and AUPRs calculated in the rest of the experiments, i.e. 75% of them, show that TanimotoPF gave better results than the rest of the similarities. In general, it can be concluded that for these datasets and these methods, TanimotoPF and SIMCOMP are more appropriate than other similarities in the DTI prediction problem.

To investigate the effect of the type and size of the datasets on the values obtained in the experiments, we check the values in Tables 2 to 5 in a different way. Figures 2 to 5 are given for this purpose. In each figure, we considered a method and illustrated the values of AUC and AUPR obtained for that method over all datasets. The results for the NRLMF, BLM-NII, NetLapRLS and WNN-GIP methods are shown in Figs. 2 to 5, respectively.
The results of Figs. 2 to 5 can be summarized as follows:

- By replacing the similarities, the change in the value of AUPR is greater than that of AUC.
- Ion Channel and Enzyme datasets seem to be less dependent on similarity matrices replacement.
- In almost all figures, when the similarity matrix is replaced, the amount of AUC and AUPR changes for the Nuclear Receptors dataset is greater than what happens for other datasets. This has sometimes happened with less tolerance for the GPCR dataset.
- Compared to other methods, the NRLMF and NetLapRLS methods are less dependent on similarities and by replacing the matrices, their AUC and AUPR values change slightly.

In addition to the more changes that occur in the results on Nuclear Receptors and GPCR datasets, all methods perform worse on these two data, compared to other data. If we review Table 1 again, we find that these two datasets are smaller than the Ion Channel and Enzyme datasets, and the difference between the AD_T and AT_D criteria in these two data is a larger number. Also, the D_{1T} criterion has a larger value for these two data, especially for the Nuclear Receptors dataset. Probably, these factors have caused that the different methods cannot have better performance and less tolerance on these two datasets.

Conclusions

The meta-analysis in the DTI problem has been done in this manuscript. Four state-of-the-art methods were selected and implemented on four benchmark datasets. The first results are about the kind of drug-drug similarities and the measures. The results show that using a meaningful similarity can improve the performance of all four methods and if there is no information about a drug, it is better to consider the similarity of that drug with other drugs as a value of 1. Tables 2 to 5 indicated that SIMCOMP has acceptable results in almost all methods and all datasets but it should be noticed that these methods have some parameters which can be optimized for the different similarities. The other important

conclusion in these tables is that the best results are relevant to the NRLMF method even with the worst similarity it has the best performance and has stable performance. It shows that, in the DTI problem, some methods are significantly related to similarity measures.

In the second experiment, tuning of the parameters of all methods has done to find the best parameters related to the used similarity matrix. The absorbing part of the result is that by finding the best parameters, the performance of the methods by using the similarities has changed and TanimotoPF similarity has the best results. It should be noted, however, that the SIMCOMP was in second place by a very small margin compared to TanimotoPF.

In the last experiment, impact of the datasets was examined. For this purpose, the performance of every method on four benchmark dataset was compared. It inferred that performance of some methods is significantly related to datasets. NetLapRLS and especially NRLMF methods are less dependent on the datasets. It can be seen that they have the same performance on all datasets even if no similarity measure is used. The performance of other methods was reduced in some datasets with some similarity measures.

So, if we want to conclude, we should mention that in the DTI prediction problem, using a drug similarity matrix can improve the results, but this improvement depends on the type of algorithm, the size and type of dataset, and the type of method used to obtain the similarity matrix. This means that this improvement may be very small for one method and very desirable for another. It may work well on some datasets and not so much on another. If a method wants to improve the results by using the drug-drug similarities, it must increase the effect of the similarity matrix in some steps of its algorithm. Otherwise, it may not achieve the desired results. As a suggestion, it can be mentioned that all the experiments performed here can be done on the target-target similarities. Perhaps different, better and more useful discussions and conclusions were made about them.

Declarations

- Ethics approval and consent to participate

Not applicable.

- Consent for publication

Not applicable.

- Availability of data and materials

The datasets and codes can be freely accessed via https://github.com/Reza-HZ/DSComparisons.

- Competing interests

The authors declare that they have no competing interests.

- Funding

Not applicable.

- Authors' contributions

All authors implements the codes, analyzed the experiments, wrote the main manuscript text, and reviewed it.

- Acknowledgements

The authors would like to express thanks to Yong Liu, Min Wu, Chunyan Miao, Peilin Zhao and Xiao-Li Li for making the PyDTI package available.

References

1. Adams CP, Brantner VV (2006) Estimating the cost of new drug development: is it really $802 million?. Health affairs 25(2):420–428.
2. Krantz A (1998) Diversification of the drug discovery process. Nature biotechnology 16(13):1294–1294.
3. Liu Y, Wu M, Miao C, Zhao P, Li XL (2016) Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLoS computational biology 12(2):p.e1004760.
4. Xia Z, Wu LY, Zhou X, Wong ST (2010) September. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC systems biology 4(2):1–16.

5. Mei JP, Kwoh CK, Yang P, Li XL, Zheng J (2013) Drug–target interaction prediction by learning from local information and neighbors. Bioinformatics 29(2):238–245.
6. Van Laarhoven T, Marchiori E (2013) Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. PloS one 8(6):p.e66952.
7. Alaimo S, Pulvirenti A, Giugno R, Ferro A (2013) Drug–target interaction prediction through domain-tuned network-based inference. Bioinformatics 29(16):2004–2008.
8. Belkin M, Niyogi P, Sindhwani V (2006) Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. Journal of machine learning research 7(11).
9. Bleakley K, Yamanishi Y (2009) Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics 25(18):2397–2403.
10. Van Laarhoven T, Nabuurs SB, Marchiori E (2011) Gaussian interaction profile kernels for predicting drug–target interaction. Bioinformatics 27(21):3036–3043.
11. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28(1):27–30.
12. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J (2016) PubChem substance and compound databases. Nucleic acids research 44(D1):D1202-D1213.
13. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research 34(suppl 1):D668-D672.
14. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic acids research 40(D1):D1100-D1107.
15. Smith TF, Waterman MS (1981) Identification of common molecular subsequences. Journal of molecular biology 147(1):195–197.
16. Hattori M, Okuno Y, Goto S, Kanehisa M (2003) Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. Journal of the American Chemical Society 125(39):11853–11865.
17. Durant JL, Leland BA, Henry DR, Nourse JG (2002) Reoptimization of MDL keys for use in drug discovery. Journal of chemical information and computer sciences 42(6):1273–1280.
18. Bolton EE, Wang Y, Thiessen P A, Bryant S H (2008). PubChem: integrated platform of small molecules and biological activities. In Annual reports in computational chemistry 4:217–241.
19. Barnard JM, Downs GM (1997) Chemical fragment generation and clustering software. Journal of chemical information and computer sciences 37(1):141–142.
20. Sheridan RP, Miller MD, Underwood DJ, Kearsley SK (1996) Chemical similarity using geometric atom pair descriptors. Journal of chemical information and computer sciences 36(1):128–136.
21. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24(13):i232-i240.
22. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic acids research 34(suppl_1):D354-D357.
23. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. Nucleic acids research 32(suppl_1):D431-D433.
24. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, Ahmed J, Urdiales EG, Gewiess A, Jensen LJ, Schneider R (2007) SuperTarget and Matador: resources for exploring drug-target relationships. Nucleic acids research, 36(suppl_1):D919-D922.

Figures

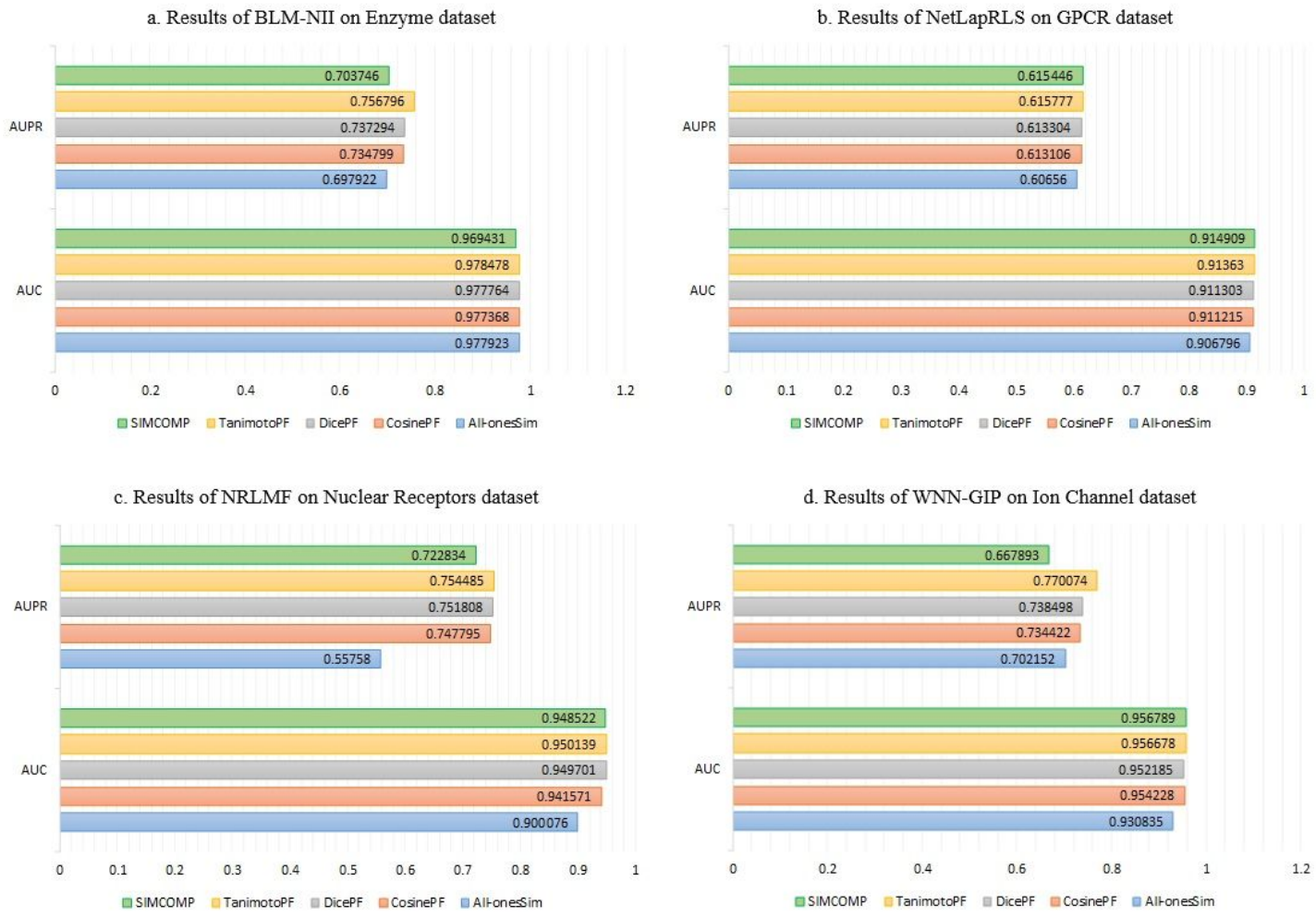
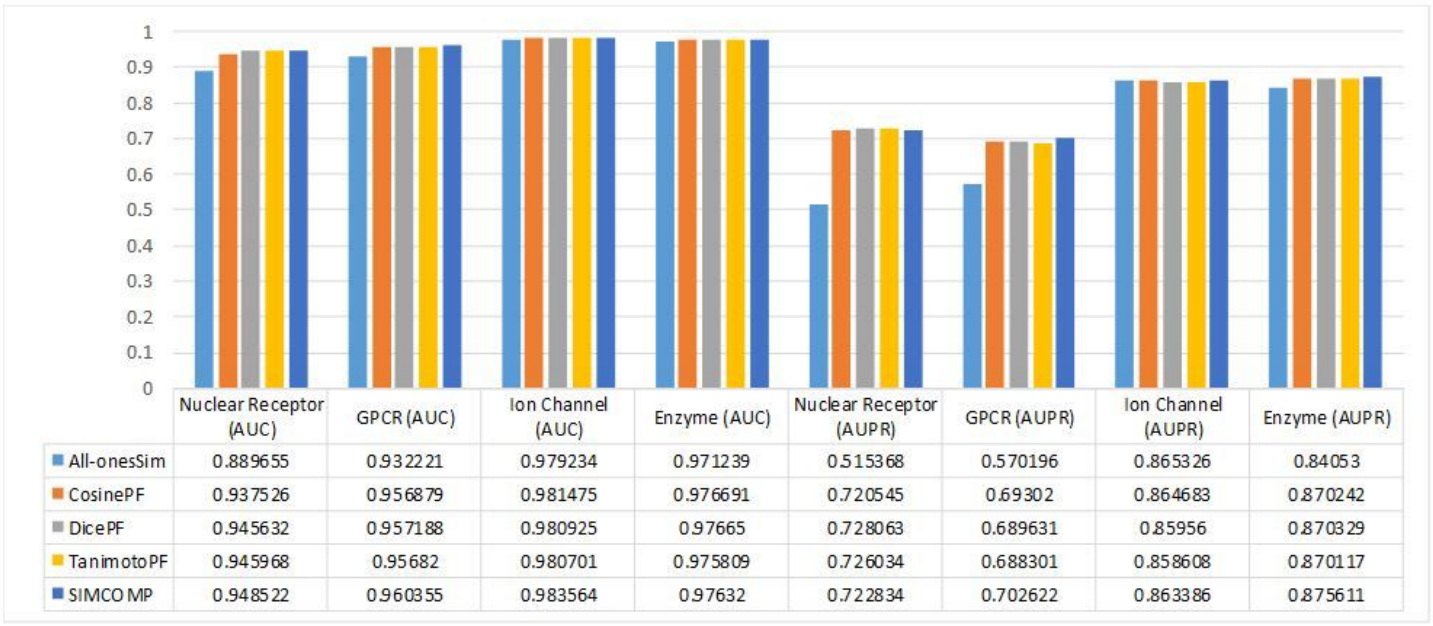


Figure 1

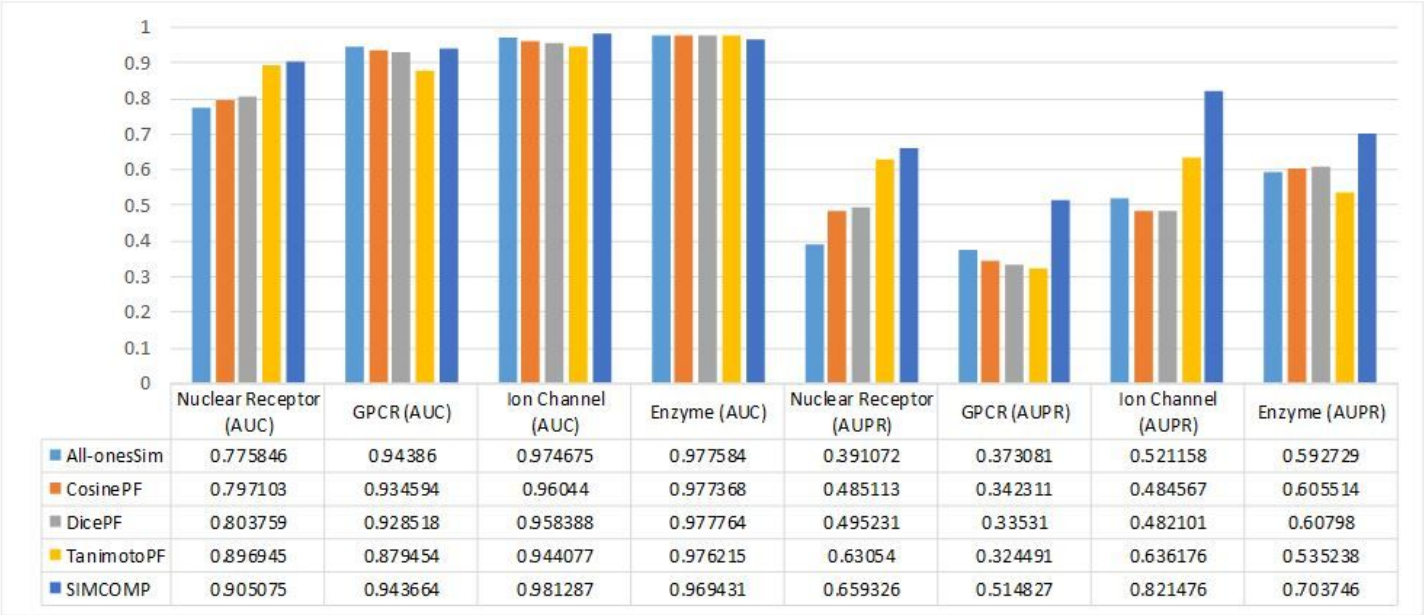
The AUC and AUPR values obtained by tuning the parameters of each method for different drug-drug similarities.



	Nuclear Receptor (AUC)	GPCR (AUC)	Ion Channel (AUC)	Enzyme (AUC)	Nuclear Receptor (AUPR)	GPCR (AUPR)	Ion Channel (AUPR)	Enzyme (AUPR)
All-onesSim	0.889655	0.932221	0.979234	0.971239	0.515368	0.570196	0.865326	0.84053
CosinePF	0.937526	0.956879	0.981475	0.976691	0.720545	0.69302	0.864683	0.870242
DicePF	0.945632	0.957188	0.980925	0.97665	0.728063	0.689631	0.85956	0.870329
TanimotoPF	0.945968	0.95682	0.980701	0.975809	0.726034	0.688301	0.858608	0.870117
SIMCOMP	0.948522	0.960355	0.983564	0.97632	0.722834	0.702622	0.863386	0.875611

Figure 2

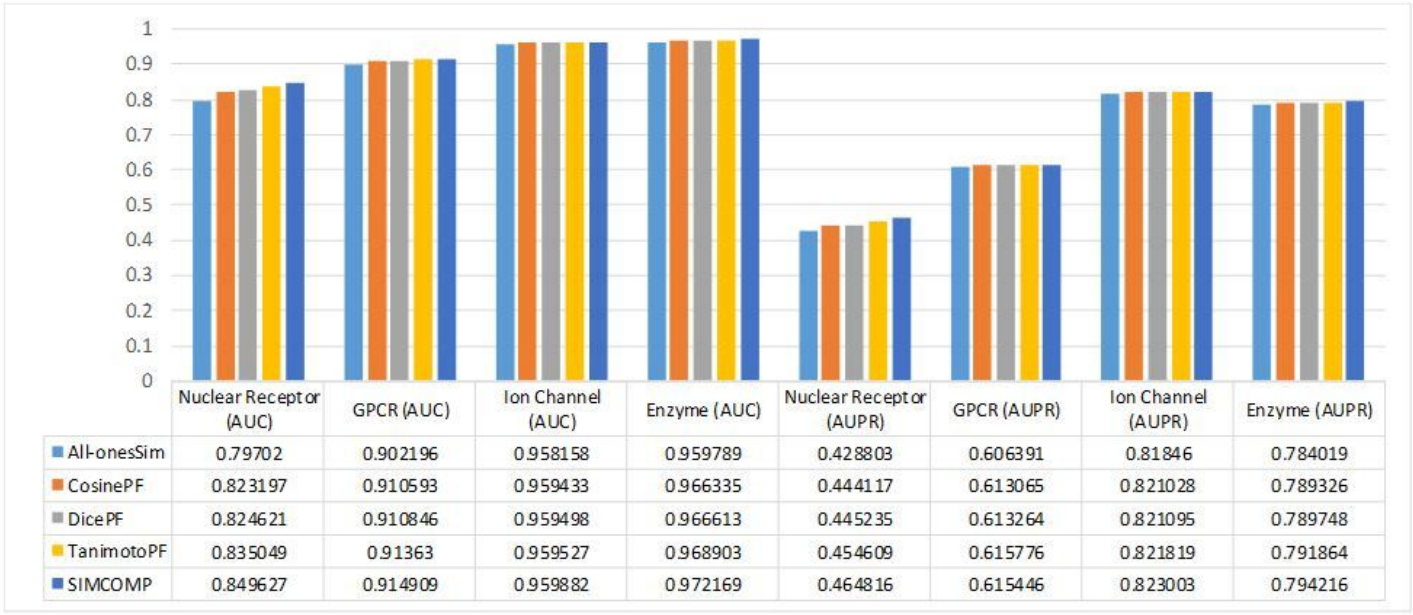
Investigating the effect of data type on the use of different drug similarities for NRLMF method.



	Nuclear Receptor (AUC)	GPCR (AUC)	Ion Channel (AUC)	Enzyme (AUC)	Nuclear Receptor (AUPR)	GPCR (AUPR)	Ion Channel (AUPR)	Enzyme (AUPR)
All-onesSim	0.775846	0.94386	0.974675	0.977584	0.391072	0.373081	0.521158	0.592729
CosinePF	0.797103	0.934594	0.96044	0.977368	0.485113	0.342311	0.484567	0.605514
DicePF	0.803759	0.928518	0.958388	0.977764	0.495231	0.33531	0.482101	0.60798
TanimotoPF	0.896945	0.879454	0.944077	0.976215	0.63054	0.324491	0.636176	0.535238
SIMCOMP	0.905075	0.943664	0.981287	0.969431	0.659326	0.514827	0.821476	0.703746

Figure 3

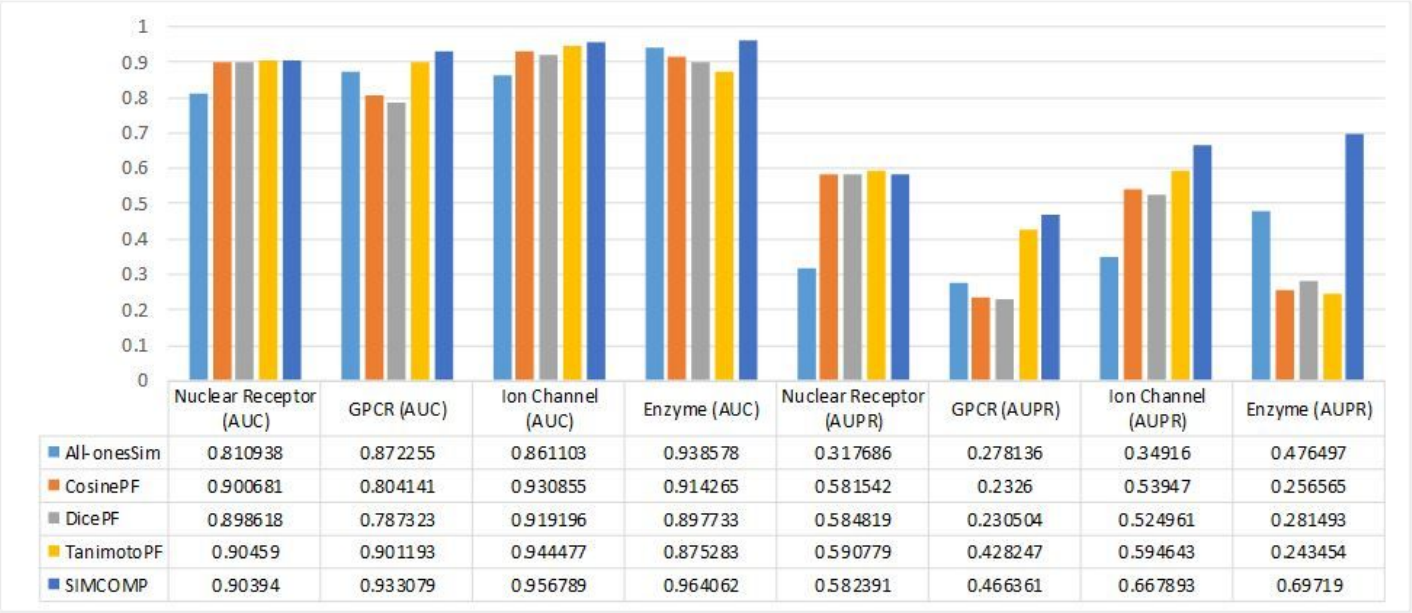
Investigating the effect of data type on the use of different drug similarities for BLM-NII method.



	Nuclear Receptor (AUC)	GPCR (AUC)	Ion Channel (AUC)	Enzyme (AUC)	Nuclear Receptor (AUPR)	GPCR (AUPR)	Ion Channel (AUPR)	Enzyme (AUPR)
All-onesSim	0.79702	0.902196	0.958158	0.959789	0.428803	0.606391	0.81846	0.784019
CosinePF	0.823197	0.910593	0.959433	0.966335	0.444117	0.613065	0.821028	0.789326
DicePF	0.824621	0.910846	0.959498	0.966613	0.445235	0.613264	0.821095	0.789748
TanimotoPF	0.835049	0.91363	0.959527	0.968903	0.454609	0.615776	0.821819	0.791864
SIMCOMP	0.849627	0.914909	0.959882	0.972169	0.464816	0.615446	0.823003	0.794216

Figure 4

Investigating the effect of data type on the use of different drug similarities for NetLapRLS method.



	Nuclear Receptor (AUC)	GPCR (AUC)	Ion Channel (AUC)	Enzyme (AUC)	Nuclear Receptor (AUPR)	GPCR (AUPR)	Ion Channel (AUPR)	Enzyme (AUPR)
All-onesSim	0.810938	0.872255	0.861103	0.938578	0.317686	0.278136	0.34916	0.476497
CosinePF	0.900681	0.804141	0.930855	0.914265	0.581542	0.2326	0.53947	0.256565
DicePF	0.898618	0.787323	0.919196	0.897733	0.584819	0.230504	0.524961	0.281493
TanimotoPF	0.90459	0.901193	0.944477	0.875283	0.590779	0.428247	0.594643	0.243454
SIMCOMP	0.90394	0.933079	0.956789	0.964062	0.582391	0.466361	0.667893	0.69719

Figure 5

Investigating the effect of data type on the use of different drug similarities for WNN-GIP method.