HAL open science

Antibiotic Bacteria Interaction: Dataset and Benchmarking

Sayantika Chatterjee, Angshul Majumdar, Emilie Chouzenoux

To cite this version:

Sayantika Chatterjee, Angshul Majumdar, Emilie Chouzenoux. Antibiotic Bacteria Interaction: Dataset and Benchmarking. 2023. hal-04376907

HAL Id: hal-04376907

https://inria.hal.science/hal-04376907v1

Preprint submitted on 7 Jan 2024

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.

Distributed under a Creative Commons Attribution 4.0 International License

Antibiotic Bacteria Interaction: Dataset and Benchmarking

Sayantika Chatterjee[a,1], Angshul Majumdar[b,2], Emilie Chouzenoux[c,3]

[a] Electronics and Communication Department, Indraprastha Institute Of Information Technology, Newdelhi, Delhi, India
[b] Institute of Advanced Intelligence, TCG CREST, Kolkata, India
[c] OPIS, Inria Saclay, CVN, University of Paris Saclay, France

Abstract

This study introduces a dataset for drug-bacteria associations (DBA) that affects humans. Our contribution extends beyond merely curating the association matrix; we also conduct genomic similarity computations for all known bacteria impacting humans and assess structural similarities for all antibiotic drugs. This is a first of a kind dataset serving the fundamental goal of streamlining antibiotic repurposing for bacterial infections. we conducted thorough benchmarking on this dataset using state-of-the-art drug repurposing models. Furthermore, we applied this dataset to three case studies involving drug-resistant bacterial infections. Our results indicate that the dataset coupled with the algorithms are indeed effective for repurposing drugs with the potential for treating drug resistant bacterial infections.

Keywords

Drug-Bacteria Association

Drug repositioning

Bacteria genomic similarity

Structural similarity for drugs

Benchmarked algorithms

Multidrug-resistant bacterial strains

[1]Email:sayantika21175@iiitd.ac.in
[2]Email:angshul@iiitd.ac.in
[3]Email:emilie.chouzenoux@inria.fr

Introduction

The central objective of this research is the curation of a comprehensive dataset named Drug-Bacteria Association (DBA). The dataset resource not only captures associations between drugs and bacteria but also provides essential information about the structural similarity of drugs and the genomic similarity of bacteria. The primary motivation behind curating this database is to harness its potential for drug repositioning, particularly in the context of antibiotics.

Drug repositioning, also referred to as drug repurposing or reprofiling, involves the exploration of new applications for existing drugs. In practice, it can be used in three scenarios:

- Investigating the effects of existing drugs on different diseases
- Identifying existing targets for new drugs
- Identifying existing drugs for new diseases

Drug repositioning presents numerous advantages in contrast to traditional drug discovery methods. Firstly, as existing drugs have already undergone rigorous safety and efficacy testing, the need for extensive clinical trials is minimized. Consequently, the cost and time required for drug development can be significantly reduced. Secondly, since the safety of existing drugs for human use has already been established, the likelihood of encountering severe side effects is notably diminished.

Drug repositioning has witnessed numerous remarkable success stories, among which one of the most widely recognized is the utilization of Aspirin for heart protection. Originally developed as a headache reliever [1], this drug has demonstrated its efficacy in safeguarding the heart. Another notable example involves Propranolol, initially designed as a beta-blocker to address hypertension [2], but subsequently repurposed with great success in the prevention of migraines [3]. These instances highlight the potential of repurposing existing drugs to address new medical conditions.

The emergence of COVID-19 brought drug repositioning to the forefront of scientific exploration. Given the novelty of the virus, developing a completely new drug to combat this infection would have been a time-consuming process. Consequently, repurposing existing drugs became the primary option in the short term [4–6]. The field of drug repositioning, particularly in the context of viral infections, experienced a surge in research efforts. New methodologies, such as networksMedicine [7] and artificial intelligence [8], were specifically developed to tackle this challenge. Additionally, an extensive database focusing on drug-virus associations was meticulously curated [9].

Two extensive datasets are available for general-purpose drug repositioning, namely from sources [10] (covering 313 diseases and 593 drugs) and [11] (encompassing 313 diseases and 663 drugs). It is worth noting that the latter dataset is an expanded version of the former, including additional entries. These datasets are not specific to any particular branch of diseases or drugs, making them suitable for testing novel algorithms designed for drug repurposing purposes. However these datasets are not focussed and does not solve any specific problem, and can only be used for benchmarking algorithms.

According to the World Health Organization (WHO), viral and bacterial infections are predominantly categorized as neglected diseases[4]. As previously mentioned, a specific database tailored for viral infections has been curated in [9]. Consequently, in this study, our focus lies in curating a comprehensive dataset specifically for drug-bacteria associations. The primary component of this database is a matrix that aligns bacteria along the columns and drugs along the rows. Within this matrix, a value of 1 indicates that a drug is effective against the corresponding bacterial infection. The

[4] https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases

information has been meticulously curated from the antibiotic policy established by the All India Institute of Medical Sciences[5]. In addition to the main association matrix, two other matrices are associated with it. The first matrix represents the similarity between drugs and was curated using DrugBank[6] and KEGG[7]. The second matrix depicts the similarity between bacteria, computed and curated using the Genome-Genome-Distance Calculator[8].

The second major contribution of this study involves the rigorous benchmarking of drug repurposing algorithms on our curated dataset. It is worth noting that datasets for drug repurposing are often limited in size, typically consisting of only a few thousand entries. Consequently, deep learning approaches tend to yield suboptimal results due to the scarcity of data. Through our experimentation, we discovered that matrix completion-based techniques, which does not rely on training, yield by far the best results. This finding is corroborated by a recent survey conducted in the field [12]. It is important to recognize that contemporary network-based methods bear relevance to matrix completion techniques [13–15]. Therefore, for the remainder of this study, we will adopt the nomenclature and approach of matrix completion.

The final contribution of this study centers on a set of case studies. While the use of antibiotics for non-complicated bacterial infections is widely acknowledged, challenges arise when bacteria develop resistance to these drugs. It is crucial to emphasize that antimicrobial resistance is a highly concerning global issue [16, 17]. Consequently, our research efforts concentrated on predicting appropriate antibiotics for drug-resistant isolates.

Methods

Data set: The information about the bacterial infections and the corresponding effective drugs are taken from the antibiotic guidelines 2018 of Christian Medical College, Vellore[9] and the antibiotic policy manual of AIIMS (All India Institute of Medical Sciences). In these two manuals, bacterial infections including the causative organisms (bacteria) and the effective drugs or antibiotics are specified.

As mentioned in the introduction, our approach for drug-bacteria association is based on matrix completion. In this approach, antibiotics are represented along the rows, while bacteria are represented along the columns (or vice versa). Following the provided guidelines, we assigned a value of 1 to indicate the effectiveness of a drug against a particular bacterial infection, thus representing true positives in the matrix. However, it is important to acknowledge that the information regarding drugs known to be ineffective against specific bacterial infections were not available in the guidelines. This limitation is not exclusive to our data set, as it is a common challenge in various biological interaction problems, including drug-target interaction [18], drug-disease interaction [10, 11], drug-drug interaction [19], drug-virus interaction [9], and others. The resulting matrix has dimensions of 53 x 61, with 53 antibiotics and 61 bacteria. Standard matrix completion cannot incorporate metadata in its framework. However, there is a host of information available for both drugs and bacteria. Past research in biological interaction where metadata is available indictae that graphical matrix completion methods yield very good results. In this class of techniques, the graphs incorporate the associated metadata.

[5] https://www.aiimsjodhpur.edu.in/quick%20docs/Antibiotic%20Policy%20AIIMS%20JDH%202018.pdf

[6] https://go.drugbank.com/categories/DBCAT000104

[7] https://www.genome.jp/tools/simcomp/simcomp_help.html

[8] https://www.dsmz.de/services/online-tools/genome-to-genome-distance-calculator-ggdc

[9] https://github.com/sayantika21175/Thesis_2023/blob/main/Antibiotic%20guidelines%20for%20adults%202018-%20CMC%20vellore.pdf

To incorporate relevant metadata for the drugs and bacteria, we obtain all the necessary information on drugs from the DrugBank vocabulary[10]. From there, we extract the unique DrugBank ID for the antibiotics. However, instead of utilizing the raw data for the drugs, we focus on the structural similarity between them. This follows from a prior work on drug virus association [9, 75]. This similarity is computed using the SIMCOMP score, which requires access to the Kyoto Encyclopedia of Genes and Genomes (KEGG) [20]. Each KEGG object is assigned a unique identifier known as the KEGG ID (kid). The linking between KEGG and DrugBank can be found in the linking file[11] [21]. In cases where certain drugs were missing from the file, we manually supplemented the information to establish the connection between each antibiotic and its corresponding DrugBank ID with the appropriate KEGG ID in the KEGG Compound/KEGG Drug database[12] of the KEGG.

Similarly, we also need to compute the similarity between bacteria, focusing on their genomic similarity. To accomplish this, we acquired the full genome sequences in FASTA format from the National Center for Biotechnology Information (NCBI) genome database[13] [22]. It is important to note that we specifically considered the primary strain for each bacteria, resulting in a total of 61 genome sequences. These sequences serve as input for calculating the similarity among bacteria.

Similarity computation

Analogies among drugs and bacteria are computed using the chemical information of drugs and the genome sequences of bacteria. These similarities prove valuable in designing the drug repurposing tool discussed in this study.

Structural similarity for drugs

The drug data includes DrugBank IDs and their corresponding KEGG IDs. These DrugBank IDs are then combined with the corresponding KEGG Compound IDs to create a new dataset that includes DrugBank ID, KEGG Compound ID, KEGG Drug ID, and the Drug Name. Each drug in the dataset is associated with either a KEGG Compound ID or a KEGG Drug ID that matches its DrugBank ID.

To calculate the similarity among the drugs, we consider their chemical significance or chemical structure. For this purpose, we utilize the SIMCOMP (SIMilar COMPound) server, which allows us to perform a chemical similarity search. The SIMCOMP score is employed to measure the chemical structure similarity between two drugs [23]. This score is obtained using a graph-based method that identifies the most extensive common elements shared by the chemical structures of the compounds.

To conduct the sequence analysis, we utilize the GenomeNet Tools API[14]. We set a cutoff value of 0.001 for the SIMCOMP score calculation. If the SIMCOMP score for a drug falls below this cutoff or if a drug does not have any available KEGG IDs, we assign a similarity score of 1 to itself and a score of 0 to every remaining pharmaceuticals in the dataset.

Bacteria genomic similarity

To determine the similarity between two bacteria based on their genome sequences, we employ the Genome-Genome Distance Calculator (GGDC) [24]. This method utilizes

[10] https://www.drugbank.ca/releases/latest#open-data
[11] https://www.drugbank.ca/releases/latest#external-links
[12] https://www.genome.jp/kegg/compound/ and https://www.genome.jp/kegg/drug/
[13] https://www.ncbi.nlm.nih.gov/genome/
[14] https://www.genome.jp/tools/gn_tools_api.html

advanced in silico techniques to compare the genomes of different organisms, mimicking the conventional DNA-DNA hybridization (DDH) method with certain improvements and avoiding its limitations. The GGDC tool offers significantly faster performance compared to other techniques available.

We also explored the use of Stretcher, a global alignment tool from the Emboss suite, to determine the similarity between two genome sequences. However, this approach proved to be time-consuming, requiring approximately 10 hours for computing the sequence similarity. In contrast, GGDC delivers the most accurate and efficient results within a considerably shorter timeframe.

For the GGDC analysis, we submit multiple query genome sequences in FASTA format representing various bacteria, along with a single reference genome sequence in FASTA format. These query genomes and the reference genome are treated as inputs for the GGDC web server. The output of the comparison is can be obtained by emailing the first author; it will contain the relevant information derived from the analysis.

Benchmarked algorithms

Matrix completion is a relatively mature area of research. As mentioned before, we are solving the drug-bacteria-association problem as one of matrix completion. From the known entries, some portions are assumed to be unknown and are hidden (10% around in several experimental scenarios). Matrix completion is used to recover these. Once the recovery is done, the recovered entries are compared with the ground-truth (hidden) values.

Let X be the entire drug-bacteria association matrix of size $m \times n$ denoted by $X_{m\times n}$. Here m stands for the total count of drugs and n stands for the total count of bacteria. As mentioned before, the association matrix contains 1 indicating the drug's recognized efficacy against the bacteria. The goal is to estimate X based on its partially observed version Y. Consider M as a masking operator. with 1 indicating the known entries, 0 otherwise. Then,

$$Y = M(X). \tag{1}$$

In the matrix completion paradigm, the assumption is that the underlying matrix to be recovered X, satisfies certain structural properties (low rank and uniformly distributed samples of known entries). Under such conditions, the number of free variables is reduced and one can expect to estimate the underlying matrix [26, 27].

Matrix factorization (MF)

The simplest and most direct approach to complete a partially observed matrix is via matrix factorization. Here, matrix X is expressed as the result of multiplying two matrices $V_{q\times n}$ and $U_{m\times q}$ where q denotes the latent space dimension, so that

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

using the factorized model,

$$X = UV. \tag{3}$$

The factors are learned by finding the solution to the optimization problem outlined below:

$$\min_{U,V} \|Y - M(UV)\|_F^2, \tag{4}$$

with $\|\cdot\|_F$ the Frobenius norm on matrices. There are many variants to the MF approach. The most notable one is non-negative matrix factorization in which the elements of both matrices V and U are constrained to be positive. Other variants include different types of regularizations on U and V [28].

Nuclear norm minimization based matrix completion

Matrix factorization builds upon a bilinear model, and as such, it necessitates a solution of a non-convex minimization task. Furthermore, one needs to have an estimate on the rank of X (i.e., the dimension of the latent space, q). Instead, one might prefer to directly estimate a low-rank matrix, by solving

$$\min_X \text{rank} X \text{ such that } Y{=}M(X) \ . \tag{5}$$

Unfortunately, solving this falls into the category of problems known to be NP-hard, which has the potential to be relaxed by its nearest convex equivalent, leveraging the nuclear norm [29, 30]. It leads to:

$$\min_X ||X||_* \text{ ensuring } Y{=}M(X) \ . \tag{6}$$

We define $\|\cdot\|_*$ as the nuclear norm acting on matrices. The resolution of the above problem is achievable through the application of semi-definite programming, or proximal-based approaches [31, 32].

Deep matrix factorization (DMF)

Fueled by its recent ubiquity in computational science, deep learning has transformed the classic technique of matrix factorization, endowing it with a powerful upgrade known as deep matrix factorization. Instead of factoring X into two matrices, one can factor it into more. Building on the achievements of deep dictionary learning [33], a deep factorization algorithm has been proposed. In this algorithm, X the data matrix is factorized into several-factor matrices enabling the extraction of intricate and potentially hidden structures within the data. For instance, in the case of three factors, we search for,

$$Y = M(U_1 U_2 V), \tag{7}$$

where,

$$X = U_1 U_2 V. \tag{8}$$

Under positivity constraints, the minimization problem is formulated as:

$$\min_{U_1, U_2, V} ||Y - M(U_1 U_2 V)||_F^2 \text{ ensuring } U_1 \geq 0, U_2 \geq 0. \tag{9}$$

Efficient algorithms based on multiplicative updates [34] and alternating least squares [35] can be used for solving the aforesaid optimization problem.

Binary matrix completion (BMC)

Drug-bacteria association problem is binary in nature. The aforementioned methods do not account for this, and return real outputs. AS crude way to binarize the results is to apply the previous algorithms followed by thresholding to binarise the output matrix. A more solution is to go for 1-bit / binary matrix completion [36]. The idea is to solve

$$\min_X \|Y - M(X)\|_F^2 + \lambda ||X||_* \text{ s.t. } X \in \{0,1\}^{n \times m}. \tag{10}$$

An efficient algorithm can be found in [36, 45].

Graph regularised matrix factorization (GRMF)

All the aforementioned algorithms are designed for scenarios where the objective is to fill a sparse matrix without additional metadata about the rows or columns. In the specific context of drug-bacteria association, we have access to such metadata in the form of genomic sequence of the bacteria and chemical structure of the drugs. This information can be employed to regularize the matrix completion problem, as elaborated in [37]. Essentially, the model incorporates a graph Laplacian penalty into the cost function, derived from similarity weights among drugs/bacteria (i.e., rows/columns). The formulation is as follows:

$$\min_{U,V} ||Y - M(UV)||_F^2 + \mu_1 \text{tr}(U^\top L_d U) + \mu_2 \text{tr}(V L_b V^\top), \tag{11}$$

Here $\mu_1 > 0$ and $\mu_2 > 0$ are penalization weights, $\text{tr}(\cdot)$ corresponds to the trace operator, $L_d = D_d - S_d$ and $L_b = D_b - S_b$ represent the graph Laplacians [38] for S_d (row/drug) and S_b (column/bacteria) similarity matrices, respectively. D_d, D_b are diagonal matrices comprising entries $D_d^{ii} = \sum_j S_d^{ij}$ and $D_b^{ii} = \sum_j S_b^{ij}$, i.e. the degree matrices. Such regularization minimizes the distance between latent feature vectors of two neighboring drugs (respectively, bacteria). A resolution technique for the above formulation has been presented in [39].

Graph regularised matrix completion (GRMC)

Graph regularization can also be incorporated into the nuclear norm minimization framework. Again, this approach utilizes graph Laplacian penalties, which incorporate drug-drug and bacteria-bacteria similarity. The minimization problem can be written as:

$$\min_{X} ||Y - M(X)||_F^2 + \lambda ||X||_* + \mu_1 \text{tr}(X^\top L_d X) + \mu_2 \text{tr}(X L_b X^\top), \tag{12}$$

Here $\mu_1 > 0$, $\mu_2 > 0$ and $\lambda > 0$ denote penalization weights. This problem can be solved using a primal-dual splitting technique, such as ADMM (alternating direction method of multipliers), exploiting the problem convexity [40].

Graph regularized deep matrix factorization (GRDMF)

Graph regularisation can also be incorporated into the deep matrix factorization approach [41]. The formulation for the graph-regularized three-factor model is:

$$\min_{U_1,U_2,V} ||Y - M(U_1 U_2 V)||_F^2 + \mu_1 \text{tr}({U_1}^\top L_d U_1) + \mu_2 \text{tr}(V L_b V^\top), \tag{13}$$

Here $\mu_1 > 0$ and $\mu_2 > 0$ are penalization weights. In this context, it is assumed that U_1 corresponds to drugs and V corresponds to bacteria. However, unlike the two-factor model where there was a distinct one-to-one mapping between drugs, bacteria, and the factor matrices, the deeper model may encounter identifiability issues. This is because it is unclear whether U_1 exclusively pertains to drugs or if $U_1 U_2$ is related to drugs, and similarly, whether V exclusively pertains to bacteria or if $U_2 V$ is associated with bacteria.

Graph regularised binary matrix completion (GRBMC)

A prior study has also introduced graph regularization into the binary matrix completion framework [42]. The minimization problem reads as:

$$\min_{X} ||Y - M(X)||_F^2 + \lambda ||X||_* + \mu_1 \text{tr}(X^\top L_d X) + \mu_2 \text{tr}(X L_b X^\top), \text{ s.t } X \in \{0,1\}^{n \times m} \tag{14}$$

with $\lambda > 0$, $\mu_1 > 0$, $\mu_2 > 0$ penalization weights. A proximal-based approach is introduced in [43] to address the above problem. We also refer the reader to [42].

In this work we have concentrated on matrix completion based techniques for drug repurposing. This is driven by our own experience in prior studies, where we found that this class of techniques yields better results than others. However, we will briefly mention other classes of techniques for the sake of completeness. The simplest method involves application of neighbourhood models from collaborative filtering to this class of problems [69]. A slightly more sophisticated approach is to use network diffusion based approaches [70, 71]. Such methods are unfortunately not as accurate as matrix completion based techniques; furthermore current network based techniques are formulated as matrix completion and not network diffusion for the same reason. Current approaches solve the problem using deep learning [72, 73]. Deep learning based techniques happen to be accurate; but they require large volumes of data which is typically absent in a problem such as ours. Hence, we have not considered this class of techniques.

Results

Within this segment we assess the effectiveness of several matrix completion algorithms. The Benchmarked algorithms subsection provides detailed descriptions of all these algorithms. Eight matrix completion algorithms or methods are used, which are divided into three categories listed here.

- Basic Frameworks (Matrix factorization(MF) [28] and Matrix completion(MC) or Nuclear norm minimization [32] and Binary matrix completion [36]).
- Deep Frameworks (Deep matrix factorization) [35]
- Graph Regularized Frameworks (Graph regularized matrix factorization(GRMF) [39], Graph regularized matrix completion(GRMC) [40], Graph regularised deep matrix factorization(GRDMF) [41], Graph regularized binary matrix completion(GRBMC) [42])

Figure 1 presents a diagrammatic illustration of the proposed system's operation. The graph-regularised methods consider all three data sets (drug-bacteria association data, drug similarity, and bacteria similarity) to predict the associations or to reconstruct the data matrix.

Experimental protocols

To provide concrete examples and compare the effectiveness of various algorithms mentioned earlier in retrieving drug-infection associations from our curated dataset, we conducted a series of experimental protocols. These protocols involved implementing 10-fold cross-validation settings on three distinct scenarios.

The 10-fold cross-validations were performed on the training data, and we utilized eight matrix completion algorithms to predict the outcomes. Below are the details of the experimental protocols or scenarios that were employed:

- 10-fold Cross validation setting on scenario 1 (CV1):10 % of drug-bacteria association selected randomly are masked or left out for testing while the remaining ones are considered for training. This allows us to determine the ability of each algorithm to predict associations between existing drugs and bacteria.

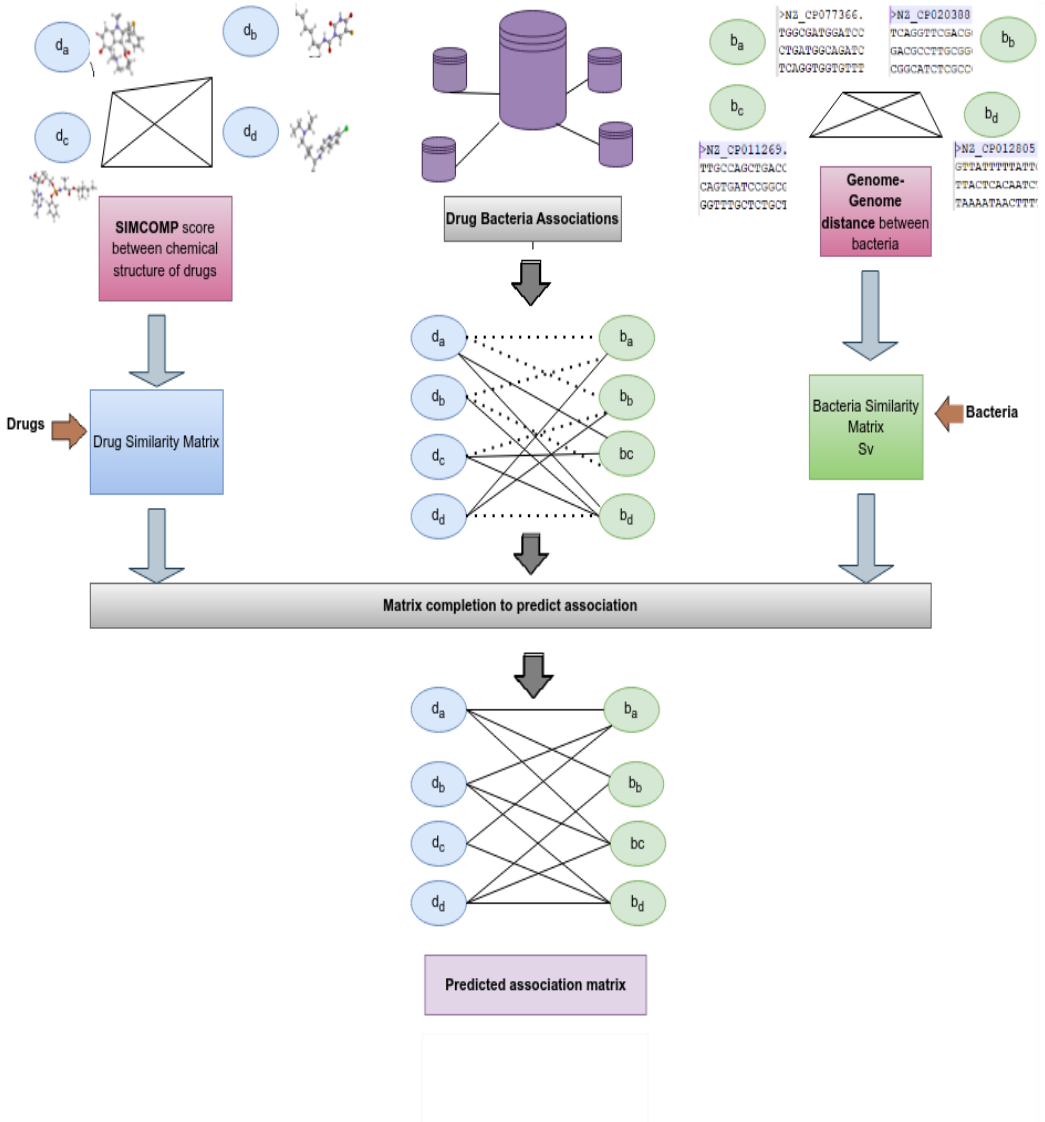

Fig 1. Schematic diagram illustrating the DBA framework.

- 10-fold Cross validation setting on scenario 2 (CV2):10 % of randomly selected bacteria are masked or left out for testing while the remaining ones are considered. This helps to repurpose drugs for new bacterial infections.
- 10-fold Cross validation setting on scenario 3 (CV3):10 % of randomly selected drugs are masked or left out for testing while the remaining ones are considered. This helps find possible applications of a novel antibiotic.

Hyperparameter settings

The hyperparameters, encompassing the step size, regularization parameters, and latent factor dimensions utilized in the algorithms for various scenarios, underwent fine-tuning following the experimental protocols detailed above (see subsection "Experimental protocols" for more information). Optimal values for these hyperparameters were determined through a grid search conducted on the training data.

The hyperparameters obtained from the second cross-validation (CV2) were utilized for predicting drugs against multidrug-resistant bacteria. Similarly, the hyperparameters obtained from the third cross-validation (CV3) were utilized for predicting bacteria for any novel drug.

For each scenario and algorithm, the specific hyperparameters utilized can be found in Tables Table 1, Table 2, and Table 3. These tables illustrate a comprehensive overview of the selected hyperparameter values for each algorithm across different scenarios.

Table 1. Parameters used in several algorithms for CV1

Algorithms	Hyperparameters for CV1
MC	$\delta = 5.05$
MF	$k = 5; \alpha = 0.5;$
BMC	$\sigma = 0.4$
DMF	$\alpha = 2; k_1 = 25; k_2 = 5;$
GRMC	$p = 7; \lambda = 0.03; \mu_1 = 0.02; \mu_2 = 0.005; \nu_1 = 0.4; \nu_2 = 0.3;$
GRMF	$p = 7; \lambda_l = 0.05; \lambda_d = 0.03; \lambda_t = 0.03;$
GRDMF	$\mu_1 = 0.1; \mu_2 = 0.05, \alpha = 1; k_1 = 25; k_2 = 5; p = 2;$
GRBMC	$p = 2; \lambda = 0.8; \mu_1 = 0.1; \mu_2 = 0.1;$

Table 2. Parameters used in several algorithms for CV2

Algorithms	Hyperparameters for CV2
MC	$\delta = 15.15;$
MF	$k = 20; \alpha = 0.05;$
BMC	$\sigma = 0.3;$
DMF	$\alpha = 0.7; k_1 = 30; k_2 = 20;$
GRMC	$p = 2; \lambda = 0.1; \mu_1 = 0.01; \mu_2 = 1; \nu_1 = 0.4; \nu_2 = 0.1;$
GRMF	$p = 8; \lambda_l = 1.5; \lambda_d = 0.1; \lambda_t = 0.5;$
GRDMF	$\mu_1 = 0.06; \mu_2 = 1.6, \alpha = 20; k_1 = 25; k_2 = 20; p = 2;$
GRBMC	$p = 2; \lambda = 0.5; \mu_1 = 3; \mu_2 = 0.02;$

Table 3. Parameters used in several algorithms for CV3

Algorithms	Hyperparameters for CV3
MC	$\delta = 5.05;$
MF	$k = 5; \alpha = 0.5;$
BMC	$\sigma = 0.5;$
DMF	$\alpha = 1.5; k_1 = 20; k_2 = 5;$
GRMC	$p = 2; \lambda = 0.01; \mu_1 = 1; \mu_2 = 0.1; \nu_1 = 0.4; \nu_2 = 0.1$
GRMF	$p = 7; \lambda_l = 0.0413; \lambda_d = 0.02; \lambda_t = 0.03;$
GRDMF	$\mu_1 = 0.4; \mu_2 = 0.7, \alpha = 1; k_1 = 25; k_2 = 5; p = 2;$
GRBMC	$p = 2; \lambda = 0.8; \mu_1 = 0.08; \mu_2 = 1;$

Experimental evaluation:

As explained in the subsection Experimental protocols, various experimental scenarios were employed by randomly concealing a subset of data. These hidden or withheld data served as the test set and were predicted by the mentioned algorithms. The performance of these algorithms was evaluated using two standard metrics: AUC (Area

Under the Receiver Operating Characteristic curve) and AUPR (Area Under the Precision-Recall curve).

Table 4 presents the results of AUC and AUPR for all the algorithms used in CV1. Similarly, Table 5 showcases the AUC and AUPR results for the algorithms utilized in CV2, and Table 6 showcases the AUC and AUPR results for the algorithms employed in CV3.

In this subsection, we evaluate and assess the performance of different algorithms in retrieving the drug-bacteria association matrix. In CV1, 10 % of randomly selected associations were masked, while in CV2 and CV3, 10 % of randomly chosen bacteria and 10 % of complete drug elements were masked, respectively.

To evaluate the models, two widely adopted metrics, namely AUC and AUPR, were utilized. AUC provides a quantitative assessment of the model's capability to separate negative and positive classes, assuming a balanced class distribution. However, since drug-bacteria association problems often exhibit imbalanced classes, AUPR is considered a more suitable metric for evaluation [39, 44].

The experimental results from CV1, CV2, and CV3 are presented in Table 4, Table 5, and Table 6, respectively. These tables display the AUC and AUPR values obtained for all eight algorithms employed in this study. Notably, the graph-regularized matrix factorization techniques outperformed the non-graph factorization techniques in terms of both AUC and AUPR. This outcome was expected, as the graphical versions incorporate additional information, such as the similarity among drugs and the similarity among bacteria, which is not available in the non-graphical versions.

The CV1 scenario focuses on the re-utilization of existing drugs to combat current infections. While this scenario may not hold significant practical importance, it serves as a standard benchmark in drug-disease association studies. Surprisingly, in this scenario, the matrix completion (MC) algorithm performs the worst, despite being a mathematically stronger technique. However, past observations have shown that matrix factorization (MF) tends to outperform nuclear norm minimization (MC) in practical applications [45]. In our case, the binarized version of MC (BMC) achieves better results, likely due to the binary nature of our problem.

Both MF and DMF (deep matrix factorization) demonstrate substantial improvements over MC and BMC. However, MF performs slightly better than DMF, indicating that deeper models may lead to overfitting due to the small size of our dataset. Interestingly, the graph-regularized versions of the algorithms exhibit similar performance in this scenario, demonstrating comparable results to each other.

The CV2 scenario carries substantial practical significance as it simulates a situation where existing drugs can be employed to treat new bacterial infections. This scenario closely mirrors real-world challenges in addressing novel strains of bacterial infections. Expanding on the insights from CV1, we note similar trends in CV2 with some slight deviations. Consistent with the preceding scenario, the deeper versions of the algorithms generally outperform their shallower counterparts, indicating the potential advantages of integrating deeper architectures in predicting drug-bacteria associations for new infections.

In scenario CV3, the emphasis is on identifying which existing bacterial infections can be effectively treated with new antibiotics, holding practical importance, especially in the context of antibiotic development. The experimental results unveil intriguing findings. Surprisingly, the MC algorithm outperforms both MF and BMC, although the reason behind this anomaly remains unclear. As anticipated, the graph-regularized versions consistently yield better results compared to their non-graph-regularized counterparts. Additionally, the deeper versions of the algorithms demonstrate improved performance over the shallower ones. Notably, the GRBMC algorithm emerges as the top performer in this scenario, achieving an AUC of 0.8029 and an AUPR of 0.5825,

indicating its effectiveness in predicting associations between new antibiotics and existing bacterial infections.

Table 4. Performance comparison of association prediction for all techniques under CV1

Metric	MC	MF	BMC	DMF	GRMC	GRMF	GRDMF	GRBMC
AUC	0.5318	0.9239	0.6883	0.9168	0.9313	0.9558	0.9244	0.9501
AUPR	0.2484	0.8270	0.4122	0.7961	0.8108	0.8928	0.8232	0.8775

Table 5. Performance comparison of association prediction for all techniques under CV2.

Metric	MC	MF	BMC	DMF	GRMC	GRMF	GRDMF	GRBMC
AUC	0.4703	0.5138	0.5046	0.6049	0.7340	0.7096	0.7320	0.7301
AUPR	0.2306	0.2233	0.2079	0.3383	0.4315	0.3922	0.4216	0.4320

Table 6. Performance comparison of association prediction for all techniques under CV3

Metric	MC	MF	BMC	DMF	GRMC	GRMF	GRDMF	GRBMC
AUC	0.5206	0.5018	0.4872	0.5837	0.7969	0.7904	0.7115	0.8029
AUPR	0.2521	0.3080	0.2293	0.2856	0.5523	0.5480	0.4344	0.5825

Case studies in drug resistant bacteria

Our initial case study focuses on the Klebsiella pneumoniae U25 isolate [46], a drug-resistant strain discovered in India that exhibits resistance to ciprofloxacin, norfloxacin, and other fluoroquinolones.

Table 7 presents the top-6 recommendations suggested by the various algorithms. All algorithms provided consistent recommendations, with piperacillin/tazobactam being predicted as the top two choices. Previous reports from the last decade have indicated the effectiveness of these drugs, comparable to carbapenems, in treating Klebsiella infections [47, 48]. Additionally, three of the algorithms predicted different carbapenems, which are also known to be effective against Klebsiella infections according to the aforementioned reports. Colistin, a last resort antibiotic for carbapenem-resistant strains, was correctly predicted by all algorithms and would have been a suitable treatment option for U25, considering it was not known to be resistant to colistin [49]. Amikacin, which recent studies have shown to be effective against Klebsiella infections [50, 51], was predicted by all four algorithms.

However, GRMF and GRBMC also predicted vancomycin, for which we could not find literature supporting its use in Klebsiella pneumoniae-induced infections. GRMF specifically predicted ceftriaxone (a cephalosporin), which is the first-line treatment for non-multi-drug-resistant Klebsiella infections. However, there is no clinical trial evidence to support its efficacy against U25 or other multi-drug resistant Klebsiella pneumoniae infections.

GRMC predicts sulbactam (ampicillin/sulbactam), we could not find any literature supporting its use in multi-drug-resistant Klebsiella infections. Rather, there is literature to support growing resistance to this drug [52].

Table 7. Predicting drugs for Klebsiella pneumoniae strain U25

Algorithms Used	Recommended drugs
GRMF	Tazobactam
	Piperacillin
	Vancomycin
	Amikacin
	Colistin
	Ceftriaxone
GRDMF	Piperacillin
	Tazobactam
	Colistin
	Amikacin
	Ertapenem
	Meropenem
GRBMC	Tazobactam
	Piperacillin
	Amikacin
	Colistin
	Meropenem
	Vancomycin
GRMC	Piperacillin
	Tazobactam
	Colistin
	Amikacin
	Meropenem
	Sulbactam

*__Green__ – Effective, **White** – Unknown.

The H041 strain of Neisseria gonorrhoeae, isolated in Japan [53], was found to be resistant to Ceftriaxone and other cephalosporins. Despite this resistance, all algorithms predicted Ceftriaxone as a treatment option, which aligns with the empirical practice of using Ceftriaxone as the first-line treatment. However, the mentioned paper reported that H041 was indeed resistant to this antibiotic, making the prediction incorrect.

On the other hand, all algorithms, except for GRMF, predicted Piperacillin/tazobactam, which is in line with the susceptibility of H041 to this drug combination as indicated in Table 8 of the mentioned paper. Therefore, this prediction is correct. GRDMF specifically predicted Ampicillin among the top 5 (or top 7 in other algorithms), and according to the mentioned paper, H041 was found to be susceptible to this drug, further validating the prediction.

GRMF predicted Penicillin G as a treatment option for H041, but as mentioned in the previous paper, H041 was found to be resistant to this antibiotic. Therefore, this prediction is also incorrect.

All algorithms predicted Vancomycin and Amoxicillin/Clavulanate as potential treatments for H041. However, since these drugs were not tested in the aforementioned paper, it cannot be definitively stated whether they would be effective against H041. While Vancomycin was occasionally used as a drug of choice for treating such infections in the 70s and 80s [54], there has been limited research on its use for treating multi-drug resistant gonorrhea since then. Similarly, Clavulanate has been sparingly used in the

past [55, 56], but there is no recent research available to support its efficacy.

GRDMF predicts gentamicin, which has been found to be effective for uncomplicated gonorrhea in previous studies [57]. In vitro studies indicate that this drug has the potential to treat multi-drug resistant gonorrhea [58–60]. However, further clinical trials are required to confirm its effectiveness in this context.

Table 8. Predicting drugs for Neisseria gonorrhoeae strain H041

Algorithms Used	Recommended drugs
GRMF	Ceftriaxone
	Clavulanate
	Amoxicillin
	Vancomycin
	PenicillinG
	Ampicillin
GRDMF	Ceftriaxone
	Vancomycin
	Piperacillin
	Tazobactam
	Ampicillin
	Gentamicin
GRBMC	Ceftriaxone
	Tazobactam
	Piperacillin
	Vancomycin
	Clavulanate
	Amoxicillin
GRMC	Ceftriaxone
	Tazobactam
	Piperacillin
	Vancomycin
	Clavulanate
	Amoxicillin

*__Green__ – Effective, **Red** – Ineffective, **Blue** – Experimental

The NIRTX011 sequence was obtained from the Chennai Bioproject[15].From Table 9, it can be observed that Amoxycillin/Clavunate is predicted by GRDMF, GRBMC, and GRMC. This drug has been demonstrated to be effective in patients with multi-drug resistant TB [61] and is also included in the WHO guidelines[16]. Similarly, Levofloxacin, which is predicted by GRMC and GRBMC, is recommended as a treatment option for multi-drug resistant TB according to the WHO guidelines. It is also used to treat individuals who have been exposed to such patients.

All the methods predicted Vancomycin. However, we could not find definitive research supporting its usage in clinical settings for multi-drug resistant TB. Nonetheless, in vitro studies indicate its potential effectiveness against multi-drug resistant tuberculosis [62, 63]. Similarly, Ceftriaxone (and Cefazolin, a related cephalosporin) has been found to be effective against multi-drug resistant TB in in vitro studies [64, 65]. With the exception of GRMF, three of the other algorithms predicted

[15] https://www.ncbi.nlm.nih.gov/bioproject/492975
[16] https://apps.who.int/iris/bitstream/handle/10665/250125/9789241549639-eng.pdf

doxycycline. In vitro studies have demonstrated that this drug, either alone [66] or in combination with amikacin [67], has potential for use in multi-drug resistant tuberculosis.

GRMF predicted Piperacillin and Tazobactam, Cefoperazone and Sulbactam. Although known as broad-spectrum antibiotics, we could not find any research supporting their treatment for multi-drug resistant tuberculosis for Piperacillin. In-vitro studies have shown the limited efficacy of sulbactam when administered with meropenem [68]. We could not find any literature on the effectiveness of Tazobactam and Cefoperazone on multi-drug resistant TB.

Table 9. Predicting drugs for Mycobacterium tuberculosis strain NIRTX011

Algorithms Used	Recommended drugs
GRMF	Vancomycin
	Ceftriaxone
	Tazobactam
	Piperacillin
	Cefoperazone
	Sulbactam
GRDMF	Ceftriaxone
	Vancomycin
	Doxycycline
	Clavulanate
	Amoxicillin
	Tazobactam
GRBMC	Ceftriaxone
	Vancomycin
	Amoxicillin
	Clavulanate
	Doxycycline
	Levofloxacin
GRMC	Ceftriaxone
	Vancomycin
	Doxycycline
	Clavulanate
	Amoxicillin
	Levofloxacin

*__Green__ – Effective, **White** – Unknown, **Blue** – Experimental

Discussion

We have compiled a database called the Drug-Bacteria Association Database (DBA), which contains information about bacterial strains and the drugs effective against them. This database also includes similarity information associated with bacteria and drugs (refer to Methods for details). Various matrix completion techniques, including graph-regularized methods, have been applied to this database.

The information on drug-bacteria associations and their associated metadata, such as similarity information, is stored in three matrices: the drug-bacteria association

matrix (Y), the bacteria similarity matrix (Sb), and the drug similarity matrix (Sd). Eight matrix completion methods are explored and evaluated. The matrix completion techniques that do not incorporate similarity information utilize the partially observed association matrix as their input data, aiming to recover the underlying low-rank structure by imputing the missing entries. These methods also incorporate a masking operator that indicates the positions of test and train indices.

Conversely, the techniques incorporating graph regularization not only utilize the association matrix but also leverage the associated metadata, namely the similarity information. The drug similarity information is obtained from the chemical structures of drugs, while the similarity information for bacteria is derived by calculating the genome distance among bacterial strains. This metadata, along with the association matrix, is provided as input to the graph-regularized algorithms. Table 4, Table 5, and Table 6 clearly demonstrates that the graph-regularized methods, which consider the similarity information among bacteria and drugs, exhibit superior performance in association prediction (as indicated by improved AUC and AUPR values) compared to the non-graph-regularized methods.

We have also focused on several multidrug-resistant bacterial strains, namely Klebsiella pneumoniae strain U25, Neisseria gonorrhoeae strain H041, and Mycobacterium tuberculosis strain NIRTX011. Using the graph-regularized method, we have generated drug recommendations (top 6) for these strains, and the outcomes are presented in Table 7, Table 8, and Table 9. It is noteworthy that the predictions made by GRMC, GRDMF, and GRBMC consistently align with each other, while GRMF exhibits inconsistencies. This discrepancy is particularly evident in the case of the NIRTX011 strain. Based on these findings, we advise against using GRMF for antibiotic repurposing and suggest relying on the recommendations from the other graph-regularized methods instead. The data set and code can be freely accessed on GitHub at the following URL: [17].

Conclusion

This study represents the first effort to curate a dedicated dataset specifically tailored for drug-bacteria associations. The primary aim in creating this dataset is to streamline the repurposing of antibiotics for treating bacterial infections in humans. Beyond documenting the associations between antibiotics and bacteria, the dataset includes essential genomic similarity data for bacteria and structural similarity data for drugs.

To tackle this problem, we have primarily resorted to matrix completion techniques. The evaluation encompassed three distinct and significant test scenarios. The first scenario focused on uncovering new applications for existing drugs in the realm of known bacterial infections. The second scenario involved identifying potential targets for newly developed antibiotics. Finally, the third scenario entailed repurposing existing drugs to combat emerging bacterial infections.

While the treatment options for uncomplicated bacterial infections are well-established, our dataset holds potential for training algorithms specifically designed to address the challenges posed by drug-resistant strains of known bacteria. Computational tools for addressing the acute problem of antimicrobial resistance are being proposed in recent times[18] [25, 74]

To demonstrate the practical relevance of our approach, we conducted three in-depth case studies centered around multi-drug-resistant strains of Klebsiella pneumoniae, Neisseria gonorrhea, and Mycobacterium tuberculosis. Notably, an analysis of existing

[17] https://github.com/sayantika21175/Thesis_2023
[18] https://www.sciencedirect.com/science/article/abs/pii/S0010482524000118

medical literature underscored the promising performance of our benchmark algorithms when trained on our meticulously curated dataset.

Acknowledgments

The authors thank Inria Saclay for the funding of this research, through the Inria International Associate Team program COMPASS.

References

1. Duffy D, Kelly E, Trang A, Whellan D, Mills G. Aspirin for cardioprotection and strategies to improve patient adherence. Postgraduate medicine. 2014;126(1):18–28.

2. Lichstein E, Morganroth J, Harrist R, Hubble E. Effect of propranolol on ventricular arrhythmia. The beta-blocker heart attack trial experience. Circulation. 1983;67(6 Pt 2):I5–10.

3. Linde K, Rossnagel K. Propranolol for migraine prophylaxis. Cochrane Database of Systematic Reviews. 2004;(2).

4. Guy RK, DiPaola RS, Romanelli F, Dutch RE. Rapid repurposing of drugs for COVID-19. Science. 2020;368(6493):829–830.

5. Venkatesan P. Repurposing drugs for treatment of COVID-19. The Lancet Respiratory Medicine. 2021;9(7):e63.

6. Kumari P, Pradhan B, Koromina M, Patrinos GP, Steen KV. Discovery of new drug indications for COVID-19: A drug repurposing approach. Plos one. 2022;17(5):e0267095.

7. Habibi M, Taheri G. Topological network based drug repurposing for coronavirus 2019. Plos one. 2021;16(7):e0255270.

8. Zhou Y, Wang F, Tang J, Nussinov R, Cheng F. Artificial intelligence in COVID-19 drug repurposing. The Lancet Digital Health. 2020;2(12):e667–e676.

9. Mongia A, Saha SK, Chouzenoux E, Majumdar A. A computational approach to aid clinicians in selecting anti-viral drugs for COVID-19 trials. Scientific reports. 2021;11(1):9047.

10. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. Molecular systems biology. 2011;7(1):496.

11. Luo H, Wang J, Li M, Luo J, Peng X, Wu FX, et al. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. Bioinformatics. 2016;32(17):2664–2671.

12. Sadeghi SS, Keyvanpour MR. An analytical review of computational drug repurposing. IEEE/ACM transactions on computational biology and bioinformatics. 2019;18(2):472–488.

13. Chen B, Li F, Chen S, Hu R, Chen L. Link prediction based on non-negative matrix factorization. PloS one. 2017;12(8):e0182968.

14. Zhao Z, Gou Z, Du Y, Ma J, Li T, Zhang R. A novel link prediction algorithm based on inductive matrix completion. Expert Systems with Applications. 2022;188:116033.

15. Sadeghi S, Lu J, Ngom A. A network-based drug repurposing method via non-negative matrix factorization. Bioinformatics. 2022;38(5):1369–1377.

16. Hofer U. The cost of antimicrobial resistance. Nature Reviews Microbiology. 2019;17(1):3–3.

17. Dadgostar P. Antimicrobial resistance: implications and costs. Infection and drug resistance. 2019; p. 3903–3910.

18. Shi JY, Yiu SM. SRP: A concise non-parametric similarity-rank-based model for predicting drug-target interactions. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2015. p. 1636–1641.

19. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research. 2018;46(D1):D1074–D1082.

20. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic acids research. 2006;34(suppl_1):D354–D357.

21. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research. 2006;34(suppl_1):D668–D672.

22. NCBI. https://www.ncbi.nlm.nih.gov/;.

23. Hattori M, Tanaka N, Kanehisa M, Goto S. SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. Nucleic acids research. 2010;38(suppl_2):W652–W656.

24. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Research. 2022;50(D1):D801–D807.

25. Singh V, Shrivastava S, Singh SK, Kumar A, Saxena S. Multi-scale temporal convolutional networks and continual learning based in silico discovery of alternative antibiotics to combat multi-drug resistance. Expert Systems with Applications. 2023;215:119295.

26. Sun R, Luo ZQ. Guaranteed matrix completion via non-convex factorization. IEEE Transactions on Information Theory. 2016;62(11):6535–6579.

27. Candes EJ, Plan Y. Matrix completion with noise. Proceedings of the IEEE. 2010;98(6):925–936.

28. Wang YX, Zhang YJ. Nonnegative matrix factorization: A comprehensive review. IEEE Transactions on knowledge and data engineering. 2012;25(6):1336–1353.

29. Candes E, Recht B. Exact matrix completion via convex optimization. Communications of the ACM. 2012;55(6):111–119

30. Candès EJ, Tao T. The power of convex relaxation: Near-optimal matrix completion. IEEE Transactions on Information Theory. 2010;56(5):2053–2080.

31. Cai JF, Candès EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM Journal on optimization. 2010;20(4):1956–1982.

32. Mongia A, Sengupta D, Majumdar A. McImpute: matrix completion based imputation for single cell RNA-seq data. Frontiers in genetics. 2019;10:9.

33. Tariyal S, Majumdar A, Singh R, Vatsa M. Deep dictionary learning. IEEE Access. 2016;4:10096–10109.

34. Trigeorgis G, Bousmalis K, Zafeiriou S, Schuller BW. A deep matrix factorization method for learning attribute representations. IEEE transactions on pattern analysis and machine intelligence. 2016;39(3):417–429.

35. Mongia A, Sengupta D, Majumdar A. deepmc: Deep matrix completion for imputation of single-cell rna-seq data. Journal of Computational Biology. 2020;27(7):1011–1019.

36. Davenport MA, Plan Y, Van Den Berg E, Wootters M. 1-bit matrix completion. Information and Inference: A Journal of the IMA. 2014;3(3):189–223.

37. Cai D, He X, Han J, Huang TS. Graph regularized nonnegative matrix factorization for data representation. IEEE transactions on pattern analysis and machine intelligence. 2010;33(8):1548–1560.

38. Chung FR. Spectral graph theory. vol. 92. American Mathematical Soc.; 1997.

39. Ezzat A, Zhao P, Wu M, Li XL, Kwoh CK. Drug-target interaction prediction with graph regularized matrix factorization. IEEE/ACM transactions on computational biology and bioinformatics. 2016;14(3):646–656.

40. Mongia A, Majumdar A. Drug-target interaction prediction using multi graph regularized nuclear norm minimization. Plos one. 2020;15(1):e0226484.

41. Mongia A, Majumdar A. Deep matrix completion on graphs: Application in drug target interaction prediction. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2020. p. 1324–1328.

42. Mongia A, Chouzenoux E, Majumdar A. Computational prediction of Drug-Disease association based on Graph-regularized one bit Matrix completion. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2022;19(6):3332–3339.

43. Pustelnik N, Chaux C, Pesquet JC. Parallel proximal algorithm for image restoration using hybrid regularization. IEEE transactions on Image Processing. 2011;20(9):2450–2462.

44. Burez J, Van den Poel D. Handling class imbalance in customer churn prediction. Expert Systems with Applications. 2009;36(3):4626–4636.

45. Gu Q, Zhou J, Ding C. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In: Proceedings of the 2010 SIAM international conference on data mining. SIAM; 2010. p. 199–210.

46. Rafiq Z, Sam N, Vaidyanathan R. Whole genome sequence of Klebsiella pneumoniae U25, a hypermucoviscous, multidrug resistant, biofilm producing isolate from India. Memorias do Instituto Oswaldo Cruz. 2016;111:144–146.

47. Zha L, Li X, Ren Z, Zhang D, Zou Y, Pan L, et al. Pragmatic Comparison of Piperacillin/Tazobactam versus Carbapenems in Treating Patients with Nosocomial Pneumonia Caused by Extended-Spectrum β-Lactamase-Producing Klebsiella Pneumoniae. Antibiotics. 2022;11(10):1384.

48. Kang CI, Park SY, Chung DR, Peck KR, Song JH. Piperacillin–tazobactam as an initial empirical therapy of bacteremia caused by extended-spectrum β-lactamase-producing Escherichia coli and Klebsiella pneumoniae. Journal of Infection. 2012;64(5):533–534.

49. Karabinis A, Paramythiotou E, Mylona-Petropoulou D, Kalogeromitros A, Katsarelis N, Kontopidou F, et al. Colistin for Klebsiella pneumoniae—associated sepsis. Clinical Infectious Diseases. 2004;38(1):e7–e9.

50. Rodrigues D, Baldissera GS, Mathos D, Sartori A, Zavascki AP, Rigatto MH. Amikacin for the treatment of carbapenem-resistant Klebsiella pneumoniae infections: clinical efficacy and toxicity. Brazilian Journal of Microbiology. 2021;52(4):1913–1919.

51. Kuti JL, Wang Q, Chen H, Li H, Wang H, Nicolau DP. Defining the potency of amikacin against Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa, and Acinetobacter baumannii derived from Chinese hospitals using CLSI and inhalation-based breakpoints. Infection and Drug Resistance. 2018;11:783.

52. Liu A, Dai J, Shen R, Zhong F, Sheng X, Huang H. Correlation between Drug Resistance of Klebsiella Pneumonia and Antimicrobial Drug Usage. Evidence-Based Complementary and Alternative Medicine. 2022;2022.

53. Ohnishi M, Golparian D, Shimuta K, Saika T, Hoshina S, Iwasaku K, et al. Is Neisseria gonorrhoeae initiating a future era of untreatable gonorrhea?: detailed characterization of the first strain with high-level resistance to ceftriaxone. Antimicrobial agents and chemotherapy. 2011;55(7):3538–3545.

54. Brorson JE, Holmberg I, Nygren B, Seeberg S. Vancomycin-sensitive strains of Neisseria gonorrhoeae. A problem for the diagnostic laboratory. British Journal of Venereal Diseases. 1973;49(5):452.

55. MUNDAY PE, BINGHAM JS, ISON CA, ERDMAN YJ, HARRIS JR, EASMON CS. Treatment of gonorrhea with clavulanate-potentiated amoxicillin (Augmentin®). Sexually transmitted diseases. 1985; p. 163–165.

56. Lim K, Rajan V, Giam Y, Lui E, Sng E, Yeo K. Two dose Augmentin treatment of acute gonorrhoea in men. Sexually Transmitted Infections. 1984;60(3):161–163.

57. Hathorn E, Dhasmana D, Duley L, Ross JD. The effectiveness of gentamicin in the treatment of Neisseria gonorrhoeae: a systematic review. Systematic Reviews. 2014;3:1–9.

58. Li X, Le W, Lou X, Wang B, Genco CA, Rice PA, et al. In vitro efficacy of gentamicin alone and in combination with ceftriaxone, ertapenem, and azithromycin against multidrug-resistant Neisseria gonorrhoeae. Microbiology Spectrum. 2021;9(2):e00181–21.

59. Bala M, Singh V, Bhargava A, Kakran M, Joshi NC, Bhatnagar R. Gentamicin susceptibility among a sample of multidrug-resistant Neisseria gonorrhoeae isolates in India. Antimicrobial Agents and Chemotherapy. 2016;60(12):7518–7521.

60. Liu JW, Xu WQ, Zhu XY, Dai XQ, Chen SC, Han Y, et al. Gentamicin susceptibility of Neisseria gonorrhoeae isolates from 7 provinces in China. Infection and Drug Resistance. 2019; p. 2471–2476.

61. MIShra G, Caminero JA. First successful use of low dose amoxicillin-clavulanic acid in management of drug resistant tuberculosis. J Clin Diagn Res. 2018;12:OD8–OD10.

62. Soetaert K, Rens C, Wang XM, De Bruyn J, Lanéelle MA, Laval F, et al. Increased vancomycin susceptibility in mycobacteria: a new approach to identify synergistic activity against multidrug-resistant mycobacteria. Antimicrobial Agents and Chemotherapy. 2015;59(8):5057–5060.

63. Srivastava S, Chapagain M, van Zyl J, Deshpande D, Gumbo T. Potency of vancomycin against Mycobacterium tuberculosis in the hollow fiber system model. Journal of Global Antimicrobial Resistance. 2021;24:403–410.

64. Srivastava S, Gumbo T, Thomas T. Repurposing cefazolin-avibactam for the treatment of drug resistant Mycobacterium tuberculosis. Frontiers in Pharmacology. 2021;12:776969.

65. Nguyen DC, Redmond SN, Dousa KM, Bethel C, Taracila MA, Li Q, et al. 1432. Exploring Cell Wall Targets to Overcome Mycobacterium tuberculosis (Mtb): Ceftriaxone (CRO) Inhibits LdtMt2, a Major Peptidoglycan (PG) Synthase. In: Open Forum Infectious Diseases. vol. 9. Oxford University Press US; 2022. p. ofac492–1261.

66. Zhang C, Ouyang Q, Zhou X, Huang Y, Zeng Y, Deng L, et al. In vitro activity of tetracycline analogs against multidrug-resistant and extensive drug resistance clinical isolates of Mycobacterium tuberculosis. Tuberculosis. 2023;140:102336.

67. Gonzalo X, Casali N, Broda A, Pardieu C, Drobniewski F. Combination of amikacin and doxycycline against multidrug-resistant and extensively drug-resistant tuberculosis. International journal of antimicrobial agents. 2015;45(4):406–412.

68. Zhang D, Wang Y, Lu J, Pang Y. In vitro activity of β-lactams in combination with β-lactamase inhibitors against multidrug-resistant Mycobacterium tuberculosis isolates. Antimicrobial Agents and Chemotherapy. 2016;60(1):393–399.

69. Zhang J, Li C, Lin Y, Shao Y, Li S. Computational drug repositioning using collaborative filtering via multi-source fusion. Expert Systems with Applications. 2017;84:281–289.

70. Su Y, Wu J, Li X, Li J, Zhao X, Pan B, et al. DTSEA: A network-based drug target set enrichment analysis method for drug repurposing against COVID-19. Computers in Biology and Medicine. 2023;159:106969.

71. Zhang P, Che C, Jin B, Yuan J, Li R, Zhu Y. NCH-DDA: Neighborhood contrastive learning heterogeneous network for drug–disease association prediction. Expert Systems with Applications. 2024;238:121855.

72. Mei X, Cai X, Yang L, Wang N. Relation-aware heterogeneous graph transformer based drug repurposing. Expert Systems with Applications. 2022;190:116165.

73. Doshi S, Chepuri SP. A computational approach to drug repurposing using graph neural networks. Computers in Biology and Medicine. 2022;150:105992.

74. Tarín-Pelló A, Suay-García B, Forés-Martos J, Falcó A, Pérez-Gracia MT. Computer-aided drug repurposing to tackle antibiotic resistance based on topological data analysis. Computers in Biology and Medicine. 2023;166:107496.

75. 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;29(5):441–452.