JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

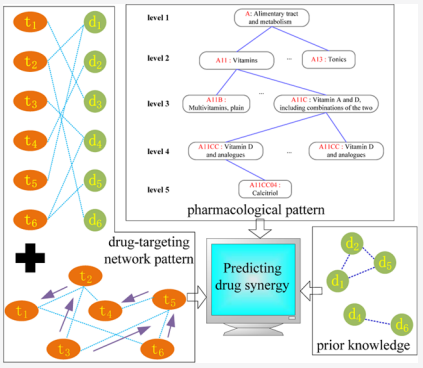
pubs.acs.org/jcim Article

Incorporating Multisource Knowledge To Predict Drug Synergy Based on Graph Co-regularization

Pingjian Ding, Cong Shen, Zihan Lai, Cheng Liang, Guanghui Li, and Jiawei Luo*

Cite This: J. Chem. Inf. Model. 2020, 60, 37−46 Read Online

ABSTRACT: Drug combinations may reduce toxicity and increase therapeutic efficacy, offering a promising strategy to conquer multiple complex diseases. However, due to large-scale combinatorial space, it remains challenging to identify effective combinations. Although many computational methods have focused on predicting drug synergy to reduce combinatorial space, they fail to effectively consider multiple sources of important knowledge. Thus, it is necessary to propose a computational method that can exploit useful information to predict drug synergy. Here, we developed a computational method to predict drug synergy based on graph co-regularization, named DSGCR. By incorporating drug−target network patterns, pharmacological patterns, and prior knowledge of drug combinations, DSGCR performs predictions of synergistic drug combinations. Compared to several existing methods, DSGCR achieves superior performance in predicting drug synergy in terms of various metrics via cross-validation. Additionally, we analyzed the importance of various sources of drug knowledge concerning three DSGCR's scenarios. Finally, the potential of DSGCR to score drug synergy was confirmed by three predicted synergistic drug combinations.



INTRODUCTION

Two or more drugs that are simultaneously or sequentially administered to patients is normally referred to as combined drug therapy for treatment regimens. Adopting combination therapy is an available and favorable way to conquer complicated diseases, such as cancer[1] and acquired immune deficiency syndrome (AIDS),[2] which can effectively improve outcomes by escalating the effect and reducing off-target toxicity by minimizing doses.[3−5] Particularly, the efficacious single-agent doses are significantly higher than the drug concentrations used in the combination of ruxolitinib and BEZ235 by the Bartalucci group. Thus, the toxicity affecting normal hematopoiesis can be reduced by dose reduction.[6] Classically, there are three types of relationships between drug pairs: an additive relationship, which indicates that the sum of the effects of each drug separately is equal to the combined effect of the drug pair; synergistic relationship, which indicates that the sum of the effects of each drug separately is less than the combined effect of the drug pair; and the antagonistic relationship, which indicates that the sum of the effects of each drug separately is greater than the combined effect of the drug pair.[4,7]

Over the past few years, some researchers have developed several platforms for high-throughput experimentation to systematically and rapidly classify antagonistic, additive, and synergistic relationships between drug pairs.[8−10] These platforms and the traditional experiments were combined to produce precision data to effectively screen synergistic drug combinations, contributing to the exploration of the potential mechanism of drug synergy and discovery of synergistic agents. The known synergistic drug combinations were collected to create the Drug Combination Database (DCDB), which facilitates the summarization of patterns of drug synergy and provides a basis to identify such synergistic drug combinations by computational methods.[11]

However, effective drug combinations are difficult to evaluate by biological experiments because of the extremely large combination space. Thus far, more than 200 cancer drugs have been approved by the Food and Drug Administration (FDA), generating at least 19 900 combinations by combining two drugs.[12] The number of combinations to be tested will be in the millions because thousands of chemical compounds are used in clinical trials. Thus, the number of potential combinations will increase exponentially while combining three or more drugs. Hence, we need to use in silico methods to reduce the screening space. Based on whether the model depends on prior knowledge, these existing computational models are mainly divided into two categories: (1) a priori knowledge-free and (2) a priori knowledge-dependent. A priori knowledge-free category can execute enrichment scores based on gene expression profile to predict drug synergy, which makes a significant contribution to screen synergistic drug

Received: September 15, 2019
Published: December 31, 2019

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

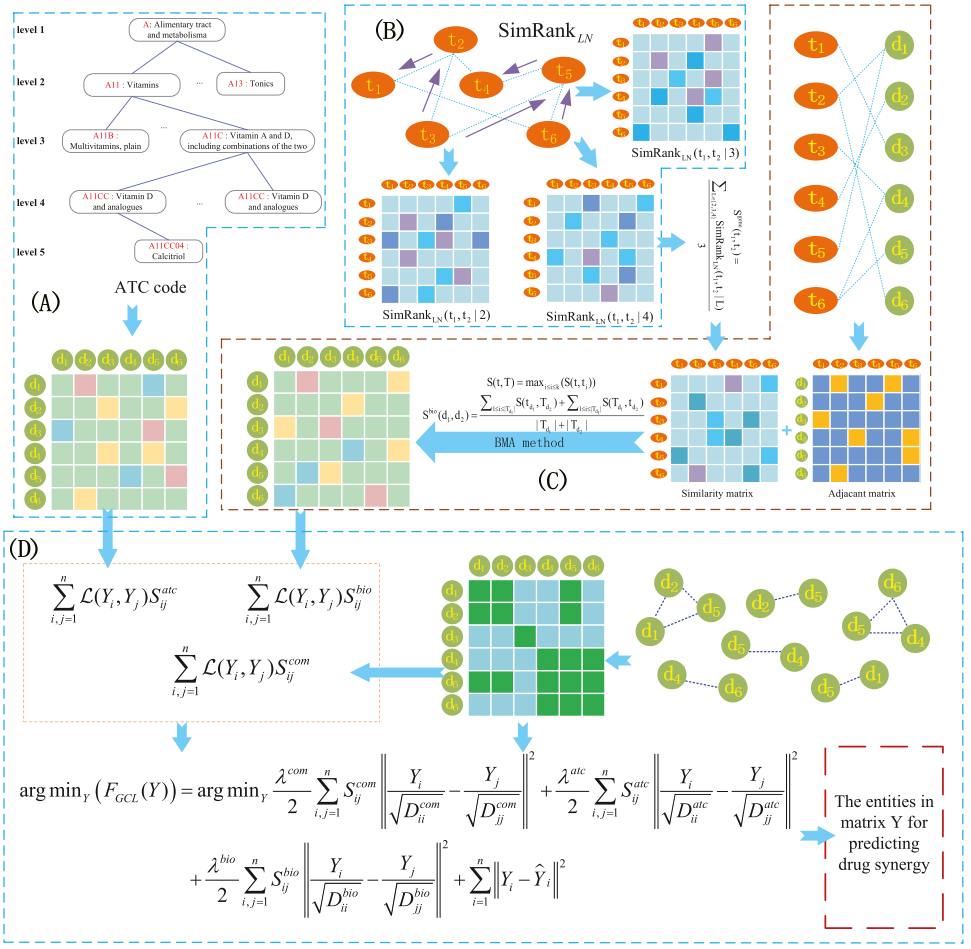


Figure 1. DSGCR workflow to predict drug synergy based on graph co-regularization. (A) The anatomical therapeutic similarity of drug pairs is calculated based on the ATC code. (B) The protein network topology similarity is calculated by using *SimRank*$_{LN}$. (C) The biological similarity of drug pairs is calculated by integrating drug−target protein interactions and network protein topology similarity based on the BMA algorithm. (D) The drug pairs are prioritized by incorporating multisource knowledge based on graph co-regularization.

combinations without known drug combinations.[13,14] It should be noted that Yang et al.[15] evaluated the synergistic score of drug pairs by integrating the gene expression changes after drug treatment and analysis of drug response curves, which significantly improves prediction accuracy by considering a specific order of treatment. However, as mentioned before, the a priori knowledge is overlooked. Due to novel drug combinations being screened based on the feature patterns enriched in the known synergistic drug combinations, a priori knowledge-dependent model which utilizes a known synergistic drug combination is proposed in refs 16−22. Thereinto, some computational methods[18,19] first extracted the feature vector of the known synergistic drug combinations and then utilized traditional machine learning to predict drug combinations. Moreover, some computational methods[20−22] directly scored drug combinations based on the network, rather than extracting feature vectors from the network. The most important problem in network-based method is how to construct a reliable network by using data from various sources and develop a network-based learning method for predicting drug synergy. Heterogeneous information network is a useful tool to automatically incorporate multiple-source information for a given task, which leads to a gain in performance.[21,22] Chen et al.[21] incorporated drug structure similarity, drug biological similarity, and the known drug combinations to build a heterogeneous information network and combined the least-squares term and graph regularization term to construct an objective function in different subnetworks. Various scores in different subnetworks were combined to finally predict the synergistic drug combination. Furthermore, to effectively incorporate biological, chemical, pharmacological, and network knowledge to score drug synergy, Ding et al.[22] proposed a method to combine synergy scores from the feature-based and network-based models with a novel ensemble prediction framework, named EPSDC. However, selecting negative samples from neutral samples would affect the prediction results of EPSDC. Hence, for the existing methods, it is difficult to effectively screen synergistic drug combinations by integrating various types of information that contributes to distinguishing synergistic drug combinations from drug pairs.

In general, the drug−target network pattern and pharmacological pattern are the important sources to investigate drug synergy, which has been validated by many studies.[23] Thus, we propose a computational method to predict drug synergy based on graph co-regularization, termed DSGCR, which could effectively incorporate drug pharmacological data and the drug−target network. In DSGCR, the Jaccard coefficient was employed to measure drug anatomical therapeutic similarity based on the Anatomical Therapeutic Chemical coding system.

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

Moreover, motivated by SimRank,[24] we developed a novel measurement, $SimRank_{LN}$, to calculate the protein network topology similarity based on the protein−protein interactions network, and then calculated the drug biological similarity with the best match average (BMA)[25] method based on the protein similarity and drug−target protein interactions. Finally, we respectively defined the graph regularization term on the drug anatomical therapeutic similarity network, the drug biological similarity network, and known drug combination network and then collectively considered various graph regularization terms to develop a novel objective function for predicting drug synergy. In the experiment, we applied the proposed method to the Drug Combinations Database (DCDB, version 2.0).[11] The comparison between DSGCR and the existing methods of predicting drug synergy was conducted in the fivefold cross-validation and leave-one-out cross-validation. Additionally, the importance of various drug similarities that imply the pharmacological and drug−target network patterns was analyzed. Finally, we presented the top-ranked synergistic drug combinations screened by DSGCR to illustrate the ability to predict drug synergy.

MATERIALS AND METHODS

In this research, we integrated various graph regularization terms, including graph regularizer for drug−target network, graph regularizer for pharmacological information, and graph regularizer for the known synergistic drug combinations, to construct an objective function to predict drug synergy. The schematic representation of DSGCR is provided in Figure 1. In DSGCR, we first utilized the Jaccard coefficient to calculate the anatomical therapeutic similarities of drug pairs based on the ATC coding system. Furthermore, a novel computational method named $SimRank_{LN}$ was proposed to measure protein network topology similarity based on the protein−protein interactions network. Next, we calculated the biological similarity of drug pairs with the BMA algorithm[25] by integrating drug target proteins and protein network topology similarities calculated by $SimRank_{LN}$. Finally, drug synergy was learned based on an objective function integrating multiple graph regularization terms defined on the known synergistic drug combinations, pharmacological knowledge, and drug−target network knowledge. Table 1 provides and summarizes the notations defined in DSGCR. A binary matrix $\hat{Y}$ with 1 for entries of known association and 0 otherwise denotes the synergistic drug combinations network, and a score matrix Y denotes the synergy score of drug pairs calculated by DSGCR.

Collecting Gold Standard Drug Combinations. The known synergistic drug combinations are collected and organized to facilitate conducting in-depth analyses of summarizing patterns of coordinated drug actions, which is beneficial to predict drug synergy with a computational model.[5] In this study, we downloaded synergistic drug combinations supported by experiments from the Drug Combination Database, DCDB 2.0 (http://www.cls.zju.edu.cn/dcdb/).[11] These drug combinations are collected from PubMed and the US FDA OrangeBook.[26] In total, the current version of DCDB includes 904 individual drugs and 1363 approved or investigational drug combinations.

Anatomical Therapeutic Similarity Analysis. Previous studies have shown that individual agents in combinations tend to belong to the same anatomical and therapeutic category in the ATC system.[23] The first three levels in the ATC codes denote the main anatomical group, main therapeutic group, and therapeutic/pharmacological subgroup. Hence, we calculated the anatomical therapeutic similarity of drug pairs based on the first three levels of the ATC code,[27] which is commonly used to facilitate the prediction of new drug combinations.[28] In this study, the ATC codes of single drugs were collected from the DrugBank Database (www.drugbank.ca).[29] Assuming that $ATC_k(d_1)$ and $ATC_k(d_2)$ represent the ATC code at the k-th level of drugs d_1 and d_2, respectively, the drug ATC similarity $S_k(d_1,d_2)$ at the k-th level between drugs d_1 and d_2 can be computed based on the Jaccard coefficient as follows

$$S_k(d_1, d_2) = \frac{|ATC_k(d_1) \cap ATC_k(d_2)|}{|ATC_k(d_1) \cup ATC_k(d_2)|}, \quad k = 1, 2, 3 \tag{1}$$

where $|.|$ represents the number of codes. The average value of ATC similarities at the first three levels is then used to define the anatomical therapeutic similarity. Thus, the anatomical therapeutic similarity between drugs d_1 and d_2 can be defined as follows

$$S^{atc}(d_1, d_2) = \frac{\sum_{k=1}^{3} S_k(d_1, d_2)}{3} \tag{2}$$

It is important to note that multiple ATC codes can be used to represent a drug. For example, three different ATC codes: D03BA02, M09AB02, and D03BA52, denote collagenase clostridium histolyticum, and three different ATC codes: D04AA01, R06AC06, and R01AC06, denote thonzylamine. The drug ATC similarity between collagenase clostridium histolyticum and thonzylamine at the first level is then:

$$S_1(\text{collagenase clostridium histolyticum, thonzylamine}) = \frac{|\{D\}|}{|\{D, M, R\}|} = \frac{1}{3}$$

Protein Network Topology Similarity Analysis. The Human Protein Reference Database (HPRD) provides the experimentally validated protein−protein interaction network in *Homo sapiens* and can be downloaded from http://www.hprd.org/.[30] Additionally, most target proteins of synergistic

Table 1. Notations

notation	definition
n	number of drugs
$S^{com} \in \mathbb{R}^{n*n}_{\geq 0}$	binary graph representing the known synergistic drug combinations
$S^{atc} \in \mathbb{R}^{n*n}_{\geq 0}$	graph representing the drug anatomical therapeutic similarity
$S^{bio} \in \mathbb{R}^{n*n}_{\geq 0}$	graph representing the drug biological similarity
$S^{com} \in \mathbb{R}^{n*n}_{\geq 0}$	normalized drug combination network $S^{com} = D^{com^{-1/2}} S^{com} D^{com^{-1/2}}$
$\overline{S}^{atc} \in \mathbb{R}^{n*n}_{\geq 0}$	normalized drug anatomical therapeutic similarity network $S^{atc} = D^{atc^{-1/2}} S^{atc} D^{atc^{-1/2}}$
$\overline{S}^{bio} \in \mathbb{R}^{n*n}_{\geq 0}$	normalized drug biological similarity network $S^{bio} = D^{bio^{-1/2}} S^{bio} D^{bio^{-1/2}}$
$L^{com} = D^{com} - S^{com}$	graph Laplacian matrix of S^{com}
$L^{atc}=D^{atc}-S^{atc}$	graph Laplacian matrix of S^{atc}
$L^{bio} = D^{bio} - S^{bio}$	graph Laplacian matrix of S^{bio}
$\overline{L}^{com} = I - \overline{S}^{com}$	normalized graph Laplacian matrix of S^{com}
$\overline{L}^{atc} = I - \overline{S}^{atc}$	normalized graph Laplacian matrix of S^{atc}
$\overline{L}^{bio} = I - \overline{S}^{bio}$	normalized graph Laplacian matrix of S^{bio}
$\hat{Y} \in \mathbb{R}^{n*n}_{\geq 0}$	known synergistic drug combinations for training
$Y \in \mathbb{R}^{n*n}_{\geq 0}$	potential drug combinations to learn

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

agents can be connected by a path length of 2 to 4 through analyzing the drug−target network.[31] Thus, these proteins connected by a 2-to-4 length path should achieve higher topological similarity. Furthermore, the protein network topology similarity can commonly contribute to the reconstruction of protein−protein interaction networks,[32] identification of protein complexes,[33,34] etc. Motivated by SimRank[24] and HeteSim,[35−37] we developed a new SimRank for the specific length to measure protein network topology similarity, named $SimRank_{LN}$. Particularly, we first introduced the $SimRank_L$ relatedness $SimRank_L(t_1,t_2|L)$ between protein t_1 and protein t_2 for the path of length L as follows

$$SimRank_L(t_1, t_2|L) = \frac{1}{|N(t_1)||N(t_2)|} * \sum_{i=1}^{|N(t_1)|} \sum_{j=1}^{|N(t_2)|} SimRank_L(t_1, t_2|L-2) \quad (3)$$

where $N(t_1)$ and $N(t_2)$ are the neighbors of t_1 and t_2, respectively, and |.| represents the number of nodes. Considering that $SimRank_L$ does not have a self-maximum property, i.e., $s(t_i, t_j) \in [0,1]$, and $s(t_i, t_i) = 1$, which is a good and important property of similarity measurement,[38] the $SimRank_L$ relatedness is normalized to be the $SimRank_{LN}$ relatedness, making the relevance measure more reasonable. Details of the $SimRank_{LN}$ are shown in the Supporting Information. Finally, we define the protein network topology similarity $S^{pro}(t_1,t_2)$ between proteins t_1 and t_2 by using the average value of $SimRank_{LN}$ relatedness based on a 2-to-4 path length

$$S^{pro}(t_1, t_2) = \frac{\sum_{L\in\{2,3,4\}} SimRank_{LN}(t_1, t_2|L)}{3} \quad (4)$$

Biological Similarity of Drug Pair Analysis. Considering the assumption that proteins with high topological similarity are associated with individual agents in drug combinations,[23] the above protein network topology similarity and known drug target protein are integrated to measure the biological similarity of the drug pair. Thus, the high-quality physical drug−target interactions used in this study are collected from DrugBank.[29] Defining the similarity between a protein and a group of proteins is the first step to calculate the biological similarity of drug pair. Given a protein t and a group of proteins $T = \{t_1,t_2,...,t_k\}$, the similarity between protein t and protein group T is calculated as follows

$$S^G(t, T) = \max_{1\le i\le k}(S^{pro}(t, t_i)) \quad (5)$$

The process of calculating the biological similarity between drug d_1 and drug d_2 is selected to clearly describe the computational method for the biological similarity of the drug pair. Assuming that T_{d_1} denotes a group of proteins related to drug d_1 and T_{d_2} denotes a group of proteins related to drug d_2, we herein introduce the biological similarity $S^{bio}(d_1,d_2)$ between drug d_1 and drug d_2 as follows

$$S^{bio}(d_1, d_2) = \frac{\sum_{1\le i\le|T_{d1}|} S^G(t_{d1_i}, T_{d2}) + \sum_{1\le j\le|T_{d2}|} S^G(T_{d1}, t_{d2_j})}{|T_{d1}| + |T_{d2}|} \quad (6)$$

Moreover, the above-described method is called the BMA (best match average) method, which is also used to calculate the similarity of various biological entities.[25,39,40]

Graph Co-Regularization. Many studies have illustrated that different types of relationships between various biological entities could be successfully predicted with a computational method that utilizes graph regularization as a base model.[41−43] Thus, we first introduced the graph regularization in the Supporting Information in detail. Moreover, to simultaneously learn drug synergy from the drug−target network pattern and pharmacological pattern, we introduced graph co-regularization by effectively extending the graph regularization to predict synergistic drug combinations.[44] Assuming that the training synergistic drug combination is $\hat{Y}$, the known drug combination matrix is S^{com}, equaling $\hat{Y}$, the drug anatomical therapeutic similarity matrix is S^{atc}, and the drug biological similarity matrix is S^{bio}, the graph co-regularization simultaneously preserves the geometric structure with respect to the drug combinations network, drug anatomical therapeutic similarity network, and drug biological similarity network by the following objective function

$$F_{GC}(Y) = \frac{1}{2}\left(\lambda^{com}\sum_{i,j=1}^{n}\mathcal{L}(Y_i, Y_j)S_{ij}^{com} + \lambda^{atc}\sum_{i,j=1}^{n}\mathcal{L}(Y_i, Y_j)S_{ij}^{atc} + \lambda^{bio}\sum_{i,j=1}^{n}\mathcal{L}(Y_i, Y_j)S_{ij}^{bio}\right) \quad (7)$$

where $\lambda^{com} \in (0,1)$, $\lambda^{atc} \in (0,1)$, and $\lambda^{bio} \in (0,1)$ are three regularization hyperparameters, implying the importance of the drug combinations network, drug anatomical therapeutic similarity network, and drug biological similarity network, respectively. Thus, in the objective function, the cost terms $\sum_{i,j=1}^{n}\mathcal{L}(Y_i, Y_j)S_{ij}^{com}$, $\sum_{i,j=1}^{n}\mathcal{L}(Y_i, Y_j)S_{ij}^{atc}$, and $\sum_{i,j=1}^{n}\mathcal{L}(Y_i, Y_j)S_{ij}^{bio}$ are used to, respectively, apply synergistic drug combination network information, drug anatomical therapeutic similarity network information, and drug biological similarity network information to predict novel drug combinations. Similar to graph regularization, the loss function $\mathcal{L}$ can be the Euclidian distance. Thus, the objective function of the prediction framework termed SDCLR (synergistic drug combinations based on Laplacian regularization) can be formulated as follows

$$F_{GCE}(Y) = \frac{\lambda^{com}}{2}\sum_{i,j=1}^{n}S_{ij}^{com}\left\|Y_i - Y_j\right\|^2 + \frac{\lambda^{atc}}{2}\sum_{i,j=1}^{n}S_{ij}^{atc}\left\|Y_i - Y_j\right\|^2 + \frac{\lambda^{bio}}{2}\sum_{i,j=1}^{n}S_{ij}^{bio}\left\|Y_i - Y_j\right\|^2 + \sum_{i=1}^{n}\left\|Y_i - \hat{Y}_i\right\|^2 \quad (8)$$

The details to obtain the analytical solution of the objective function $F_{GCE}(Y)$ are described in the Supporting Information. Moreover, a normalization technique in graph-based learning was adopted to suppress hubs and has been validated by many studies.[45] Thus, to reduce the effect of drugs that have similarity with many drugs, the prediction framework termed DSGCR can be formulated as follows

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

$$\arg\min_{Y}(F_{GCL}(Y)) = \arg\min_{Y} \frac{\lambda^{com}}{2}\sum_{i,j=1}^{n} S_{ij}^{com}\left\|\frac{Y_i}{\sqrt{D_{ii}^{com}}} - \frac{Y_j}{\sqrt{D_{jj}^{com}}}\right\|^2 + \frac{\lambda^{atc}}{2}\sum_{i,j=1}^{n} S_{ij}^{atc}\left\|\frac{Y_i}{\sqrt{D_{ii}^{atc}}} - \frac{Y_j}{\sqrt{D_{jj}^{atc}}}\right\|^2 + \frac{\lambda^{bio}}{2}\sum_{i,j=1}^{n} S_{ij}^{bio}\left\|\frac{Y_i}{\sqrt{D_{ii}^{bio}}} - \frac{Y_j}{\sqrt{D_{jj}^{bio}}}\right\|^2 + \sum_{i=1}^{n}\left\|Y_i - \hat{Y}_i\right\|^2 \quad (9)$$

We then calculate the derivation of $F_{GCL}(Y)$ with respect to Y to solve the above optimization problem as follows

$$\frac{\partial(F_{GCL}(Y))}{\partial Y} = \lambda^{com}Y\overline{L}^{com} + \lambda^{atc}Y\overline{L}^{atc} + \lambda^{bio}Y\overline{L}^{bio} + Y - \hat{Y} = 0 \quad (10)$$

Furthermore, the Supporting Information provides details about analytically solving the objective function $F_{GCL}(Y)$. Thus, the analytical solution of the objective function can be obtained after some algebraic transformations and is displayed as follows

$$Y = \hat{Y}*(\lambda^{com}\overline{L}^{com} + \lambda^{atc}\overline{L}^{atc} + \lambda^{bio}\overline{L}^{bio} + I)^{-1} \quad (11)$$

where I is the identity matrix of size $n * n$. Moreover, the Supporting Information provides a process, proving that the matrix $\lambda^{com}\overline{L}^{com} + \lambda^{atc}\overline{L}^{atc} + \lambda^{bio}\overline{L}^{bio} + I$ is a positive definite and invertible. Y represents the synergy scores of drug pairs. Because the matrix Y may not be symmetric, we consider $Y_{ij} + Y_{ji}$ as the synergy score between the i-th drug and j-th drug.

Performance Evaluation Metrics. The performance of predicting drug synergy was estimated using the precision, recall, AUC (the area under the receiver operating characteristic (ROC) curve),[46] AUPR (the area under the precision recall (PR) curve),[47] and P-value calculated by paired t-test.[48]

The overall performance to predict drug synergy could be estimated by AUC, driven by the true-positive rates (TPRs) and the false-positive rates (FPRs) at different ranking cutoffs. Therefore, TPR denotes the ratio of identified positive samples accounting for all of the positive samples, while FPR denotes the proportion of negative samples predicted incorrectly among the total negative samples. Given the threshold set $\Theta = \{\theta_1,\theta_2,...,\theta_m\}$, AUC can be calculated as follows

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\theta_i)d(\mathrm{FPR}(\theta_i)) \quad (12)$$

Because biologists tend to adopt wet-lab experiments to further validate the top-rankings, there is a tremendous need to ensure more positive samples in the top-ranked candidates. Meanwhile, compared to the ROC curve, the PR curve can heavily punish negative samples with a high rank. Thus, we also employed AUPR to estimate the performance to predict drug synergy. The PR curve can be drawn with precisions and recalls at different ranking cutoffs. Therefore, given true-positive TP, true-negative TN, false-positive FP, and false-negative FN, precisions and recalls can be defined as follows

$$\mathrm{precision}(\theta_i) = \frac{\mathrm{TP}(\theta_i)}{\mathrm{TP}(\theta_i) + \mathrm{FP}(\theta_i)},\ i = 1, 2, \ldots m \quad (13)$$

$$\mathrm{recall}(\theta_i) = \frac{\mathrm{TP}(\theta_i)}{\mathrm{TP}(\theta_i) + \mathrm{FN}(\theta_i)},\ i = 1, 2, \ldots m \quad (14)$$

Thus, AUPR can be calculated as follows

$$\mathrm{AUPR} = \int_0^1 \mathrm{precision}(\theta_i)d(\mathrm{recall}(\theta_i)) \quad (15)$$

Moreover, a statistical method, the paired t-test, is applied to claim the differences between the paired observations. We can suggest that the results of one method is significantly different from that of another method, while the P-value is less than the significance level (0.05). Furthermore, it should be noted that P-values are two-sided in this study.[48]

Software Package. We have uploaded the R software package through GitHub to https://github.com/KDDing/DSGCR, containing all codes used to run all proposed methods. Furthermore, the package can be used to execute fivefold cross-validation and leave-one-out cross-validation, as well as select hyperparameters for reproducing the results. The readme file details how to run the program and interpret the generated results.

EXPERIMENTS AND RESULTS

In the experiments, both the fivefold cross-validation and leave-one-out cross-validation are applied to evaluate the performance in predicting drug synergy based on multisource drug knowledge. Additionally, the effect of normalization is estimated by comparing DSGCR with SDCLR. Moreover, we demonstrate how the drug−target network pattern and pharmacological pattern can benefit the predictions by analyzing the impact of two types of drug similarity. Finally, we demonstrate the ability of DSGCR to score drug synergy by showing several examples.

Data Preparation. The known synergistic drug combinations were downloaded from the Drug Combination Database (DCDB 2.0, http://www.cls.zju.edu.cn/dcdb/)[11] that stores information on drug combinations for facilitating systems-oriented new drug discovery. We downloaded the human protein−protein interactions network from the Human Protein Reference Database (HPRD, http://www.hprd.org/)[30] that collects and organizes proteomic information pertaining to human proteins. The ATC code of the drug and drug−protein interactions were collected from DrugBank 5.0 (www.drugbank.ca)[29] that is a database of comprehensive molecular information about drugs, their mechanisms, interactions, and targets. In the experiments, the same preprocessing method as in ref 22 was employed to eliminate single drugs for a fair comparison. In fact, single drugs in study could be represented as $SD_{st} \cap SD_{ATC}$, where SD_{st} denotes the set of single drugs with chemical structure information and SD_{ATC} denotes the set of single drugs with ATC code information in DCDB dataset. After eliminating single drugs, there remain 139 single drugs with both chemical structure information and ATC code information. However, a recent study reported that chemical structure information cannot be applied to distinguish synergistic drug combinations from random drug pairs.[16] Thus, DSGCR scores drug synergy without considering the chemical structure. After eliminating protein without related drug, 173 approval pairwise synergistic drug combinations and 449 drug−target interactions between 139 single drugs and 218 target proteins were identified in the benchmark dataset. The drug−target network and protein−protein interaction network based on the benchmark database are provided in Figures S2 and S3, respectively. The benchmark dataset is available through GitHub at https://github.com/KDDing/DSGCR.

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

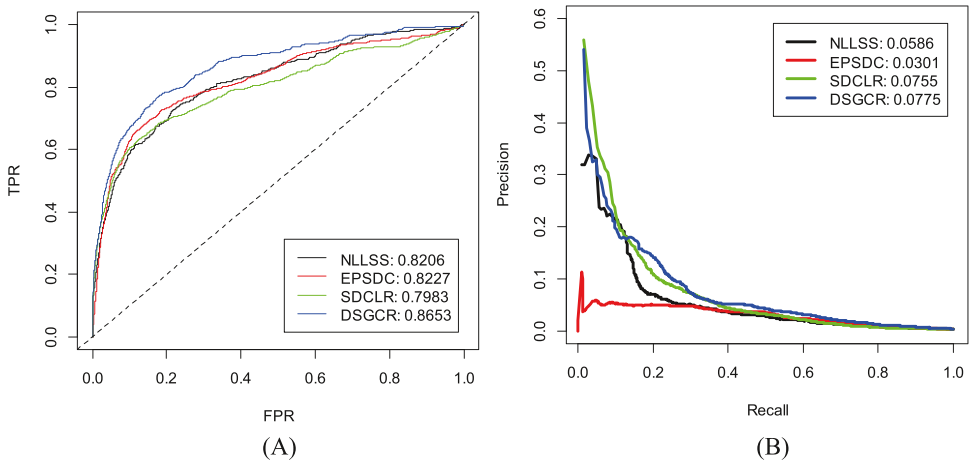


Figure 2. Performance of different methods to predict drug synergy with fivefold cross-validation. (A) ROC curves of various methods with fivefold cross-validation. (B) PR curves of various methods with fivefold cross-validation.

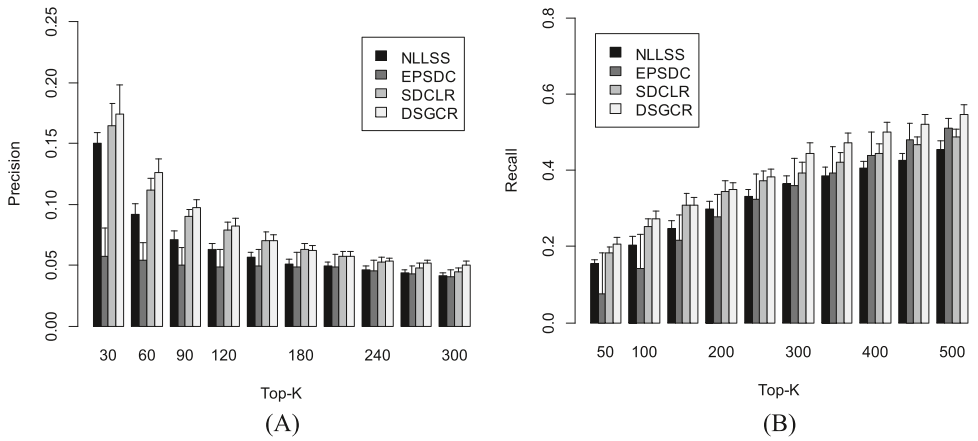


Figure 3. Evaluation of different methods to predict drug synergy at top-k rankings. (A) Average precision values of various methods with fivefold cross-validation. (B) Average recall values of various methods with fivefold cross-validation.

Comparison with Other Methods. In this paper, fivefold cross-validation (FFCV) experiments were conducted to identify optimal hyperparameters. The grid search method was used to determine all parameter combinations. We determined the optimal combination for λ^{com}, λ^{atc}, and λ^{bio} from the set {0.01,0.1,0.3,0.5,0.7,0.9}. Subsequently, we set λ^{com} = 0.3, λ^{atc} = 0.1, and λ^{bio} = 0.01. To ensure a fair comparison, we set the parameters to their default values based on the authors' recommendations (η_A = 0.3, η_P = 0.3 for NLLSS, and maximum length = 4 for EPSDC).

To systematically estimate the prediction ability of DSGCR on the benchmark dataset, DSGCR was compared to other computational methods (such as NLLSS,[21] EPSDC,[22] and SDCLR.) via fivefold cross-validation (FFCV) and leave-one-out cross-validation (LOOCV) experiments. Specifically, in NLLSS,[21] a heterogeneous information network was first constructed by combining the known drug combinations and drug similarity; then, the Laplacian regularized least-squares algorithm was applied to the constructed network to predict drug synergy. Moreover, for a fair comparison, the chemical similarity is replaced with the anatomical therapeutic similarity in NLLSS to ensure that all of the methods perform predictions on the same dataset. Moreover, the performance of NLLSS may be improved by replacing the anatomical therapeutic similarity because drug structural features cannot distinguish synergistic combinations from random ones in the context of cancer therapy, a finding that has been reported previously.[31] However, most of the individual agents in synergistic drug combinations belong to the same anatomical and therapeutic group reported in recent studies.[16] In EPSDC, the natural properties and drug-related relationships were integrated with ensemble rules to identify synergistic drug combinations. Compared to DSGCR, SDCLR makes predictions without normalization. In each run of FFCV, we equally divide all of the known pairwise drug combinations into five parts. Each part is selected as the positive samples in the testing set in order, and the remaining four parts are used as the positive samples in the training set. In LOOCV, each known drug combination is selected as a positive sample in the testing set in turn, and the remaining combinations are utilized as the positive samples in the training set.

To estimate the performance of DSGCR, we conducted a comparison between DSGCR and two existing methods, NLLSS and EPSDC, to predict drug synergy, and we also used a graph co-regularization-based method without normalization, SDCLR. To evaluate the performance of the learning methods with probability estimations, AUC and AUPR are the outstanding metrics.[49,50] Thus, we evaluated the prediction performance of DSGCR and other methods with AUC and AUPR. We performed 10 runs of FFCV to plot ROC curves

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

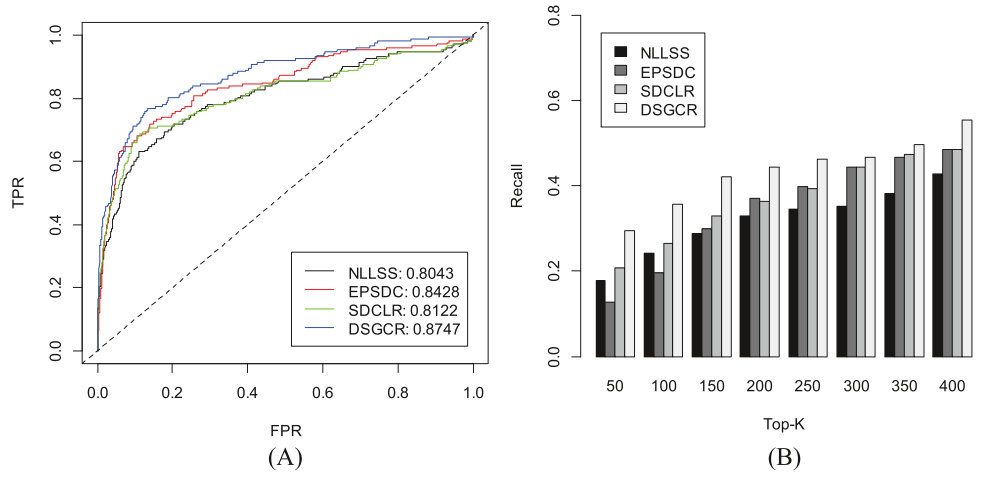


Figure 4. Performance of different methods in predicting drug synergy with leave-one-out cross-validation. (A) ROC curves of various methods with leave-one-out cross-validation. (B) Average recall values of various methods with leave-one-out cross-validation.

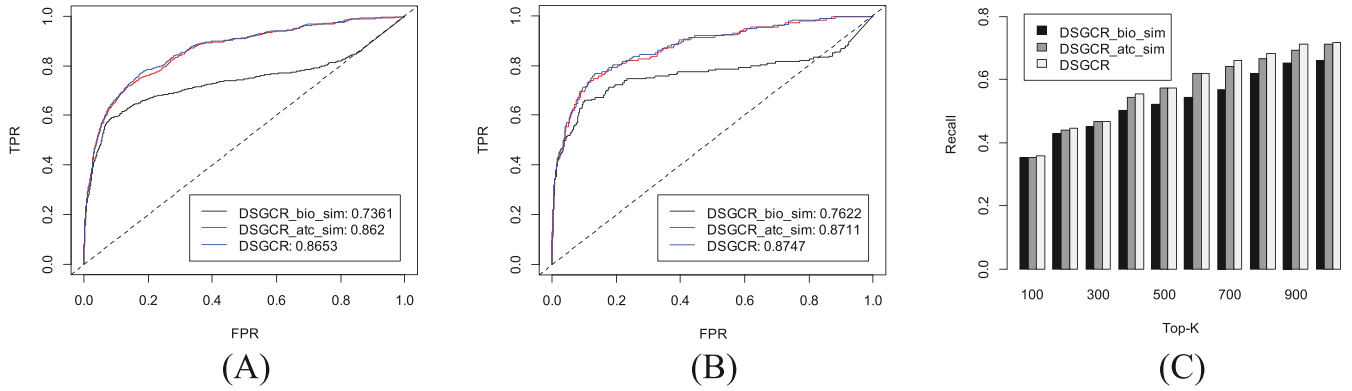


Figure 5. (A) ROC curves of DSGCR trained with both or either one of two similarities via FFCV. (B) ROC curves of DSGCR trained with both or either one of two similarities via LOOCV. (C) Average recall values of DSGCR trained with both or either one of two similarities via LOOCV.

and PR curves and obtain the corresponding AUCs and AUPRs of different methods (Figure 2). As shown in Figure 2A, DSGCR obtained the highest AUC of 86.53%. DSGCR outperforms NLLSS by 4.47%, EPSDC by 4.26%, and SDCLR by 6.7%. The AUCs of DSGCR and other methods with different runs were compared using paired *t*-test via fivefold cross-validation. *P*-values (*P*-value = 1.85E − 12 comparing DSGCR with NLLSS, *P*-value = 1.55E − 09 comparing DSGCR with EPSDC, and *P*-value = 0.72E − 11 comparing SDCLR with DSGCR) were less than 0.05, suggesting that the differences between AUCs were statistically significant. As shown in Figure 2B, the highest AUPR of 7.75% was produced by DSGCR. Its AUPR was 1.89, 4.74, and 0.2% better than those of NLLSS, EPSDC, and SDCLR, respectively. Furthermore, DSGCR yields a better performance than SDCLR, illustrating the effectiveness of normalization. Similarly, the AUPRs of DSGCR and other methods with different runs via fivefold cross-validation were compared using paired *t*-test. *P*-values (*P*-value = 0.2E − 03 comparing DSGCR with NLLSS, *P*-value = 5.8E − 09 comparing DSGCR with EPSDC, and *P*-value = 0.21E − 03 comparing SDCLR with DSGCR) were less than 0.05, suggesting that the differences between AUPRs were statistically significant.

The precision within the top-k ranking list implies the reliability of potential synergistic drug combinations. As shown in Figure 3A, the overall DSGCR outperformed the other methods from the top 30 to top 300 in terms of precision. The exact precision values of various methods are shown in Table S1. Furthermore, we applied the paired *t*-test based on different top-rankings to show that DSGCR achieved significant performance in terms of precision compared to NLLSS and EPSDC (*P*-value = 0.42E − 03 comparing DSGCR with NLLSS, *P*-value = 0.15E − 01 comparing DSGCR with EPSDC, and *P*-value = 0.17E − 01 comparing SDCLR with DSGCR). The higher the recall on the top-k rankings is, the more positive samples in the test set are successfully identified. As shown in Figure 3B, the overall DSGCR performed better than the other methods from the top 50 to top 500 in terms of recall. The exact recall values of various methods are also shown in Table S1. Moreover, paired *t*-tests based on different top-rankings reported that the performance of DSGCR was significantly better than the prediction results of other methods in terms of the obtained recalls (*P*-value = 5.42E − 02 comparing DSGCR with NLLSS, *P*-value = 3.31E − 05 comparing DSGCR with EPSDC, and *P*-value = 0.19E − 02 comparing SDCLR with DSGCR). Generally, DSGCR is always the best method in terms of precision and recall.

Additionally, we implemented leave-one-out cross-validation (LOOCV) to estimate the performance of various methods because the training set in LOOCV is more similar to the real data. Because the positive/negative sample rate (1/9419) in the testing set is extremely low in LOOCV, the AUC and recall are considered as metrics, ignoring AUPR and precision. As

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

shown in Figure 4A, NLLSS, EPSDC, SDCLR, and DSGCR achieved AUCs of 80.43, 84.28, 81.22, and 87.47%, respectively. DSGCR performed better than NLLSS by 7.04%, EPSDC by 3.19%, and SDCLR by 6.25% in terms of the AUC. At the top 50, the average recalls of NLLSS, EPSDC, SDCLR, and DSGCR were 17.92, 12.72, 20.81, and 29.48%, respectively. At the top 400, these four methods achieved average recall values of 42.77, 48.55, 48.55, and 55.5%, respectively. Additionally, we implemented paired *t*-test to assess the statistical significance. Table S1 lists *P*-values for the leave-one-out cross-validation, which further reported that DSGCR obtained better performance in terms of the AUC and recall at the significance level of 0.05. Generally, DSGCR is always the best method in terms of recall via LOOCV from the top 50 to top 400.

Importance of Various Drug Similarities. In the experiments, we constructed three DSGCR scenarios, DSGCR_bio_sim, DSGCR_atc_sim, and DSCGR, to evaluate the importance of drug anatomical therapeutic similarity and drug biological similarity. DSGCR_bio_sim is trained without drug anatomical therapeutic similarity; DSGCR_atc_sim is trained without drug biological similarity; and DSGCR is trained with both drug anatomical therapeutic similarity and drug biological similarity. As shown in Figure 5, DSGCR with both similarities performed better than DSGCR_bio_sim and DSGCR_atc_sim via FFCV and LOOCV. DSGCR exceeded DSGCR_bio_sim with an AUC of 12.92% and DSGCR_cli_sim with an AUC of 0.33% (Figure 5A). Additionally, as shown in Figure 5B, the average AUCs of DSGCR_bio_sim, DSGCR_cli_sim, and DSGCR were 76.22, 87.11, and 87.47% via LOOCV, respectively. DSGCR's AUC was 11.25 and 0.36% higher than those of DSGCR_bio_sim and DSGCR_cli_sim, respectively. Figure 5C provides the average recalls of three DSGCR scenarios from top 100 to top 1000 via LOOCV. Additionally, for AUCs with different runs in fivefold cross-validation, as well as AUCs and recalls with different ranks in leave-one-out cross-validation, paired *t*-tests were also carried out to measure the statistical significance. The *P*-values are provided in Table S2, which claimed that DSGCR combining two similarities achieved higher AUCs and recalls at the significance level of 0.05. Therefore, we could say that incorporating the drug anatomical therapeutic similarity could greatly improve the prediction performance in terms of AUC and recall, and DSGCR_bio_sim could improve the prediction of drug synergy to a certain extent in this study.

Predicted Synergistic Drug Combinations. After estimating the prediction performance of DSGCR by cross-validation, we further illustrated DSGCR's ability to predict drug synergy based on the novel synergistic drug combinations predicted by DSGCR. The prediction model of DSGCR was trained on all of the known synergistic drug combinations. Next, we obtained the potential candidate combinations with DSGCR, and the top three drug combinations prioritized by DSGCR are listed in Table 2. Thus, the drug targets described in Table 2 were obtained from DrugBank. These inferred synergistic drug combinations were searched in PubMed to obtain related information from the literature. Bryant et al.[51] reported that both gemcitabine and 5-fluorouracil can be used to manage pancreatic ductal adenocarcinoma by indirectly blocking replicative forks. Gutierrez-Delgado et al.[52] reported that the oxaliplatin and cyclophosphamide combination is well tolerated and has an efficacy comparable to other neoadjuvant chemotherapy regimens. Ketter et al.[53] suggested that carbamazepine and valproic acid used together may have synergistic anticonvulsant effects. Moreover, we have provided the full rankings of potential synergistic drug combinations that are available in https://github.com/KDDing/DSGCR.

Table 2. Top Three Novel Combinations Inferred by DSGCR

rank	drug name	DrugBank ID	ATC code	targets
1	Gemcitabine	DB00441	L01BC05	DNA; PRM1; TYMS; CMPK1
	Fluorouracil	DB00544	L01BC52; L01BC02	DNA; RNA; TYMS
2	Oxaliplatin	DB00526	L01XA03	DNA
	Cyclophosphamide	DB00531	L01AA01	DNA; NR1I2
3	Carbamazepine	DB00564	N03AF01	SCN5A; CHRNA4; CHRNB2; NR1I2
	Valproic acid	DB00313	N03AG01	ACADSB; HDAC9; OGDH; ALDH5A1; HDAC2; PPARA; PPARD; PPARG

CONCLUSIONS

More durable clinical responses may be caused by synergistic drug combinations, which can overcome drug resistance to monotherapies. However, due to the large-scale combinatorial space,[54] it remains a challenge to identify drug synergy. According to various mechanism of action patterns potentially related to drug synergy, the potential synergistic drug combinations could be quickly screened based on machine learning. Therefore, we have proposed a novel graph co-regularization-based method to predict drug synergy, named DSGCR. This method requires the known synergistic drug combinations, pharmacological pattern and drug−target network pattern. Specifically, the proposed protein network topology similarity measurement $SimRank_{LN}$ considered the characteristic of synergistic agents that can meet by the 2-to-4 length path in the protein−protein interaction network. In the experiments, for predicting drug synergy, DSGCR outperformed other methods by a wide margin in cross-validation. Additionally, we investigated the importance of various mechanisms of action patterns potentially related to drug synergy. Finally, the predicted synergistic drug combinations further illustrated the powerful ability to identify synergistic drug combinations from the neutral samples. Overall, our findings suggest that DSGCR can be a useful tool to score drug synergy. However, this computational model cannot explore the dose of synergistic drug agents. Furthermore, the synergy score matrix may not be symmetric, that is, DSGCR generates a nonintuitive result. Hence, the future work needs to ensure that the synergistic score matrix is symmetric to generate intuitive synergy scores in future work. Moreover, although DSGCR can predict drug synergy by incorporating multisource information, the performance improvement obtained by combining the biological knowledge of drug is limited. The major reason may be that the proposed method depends on the assumption that synergistic effects from drugs act on the similar pathways, but many synergistic effects are from drugs with different modes of action. Additionally, the cause may be that drug−target interactions at the genetic level, rather than at the binding site level, are used in the study. Hence, it is urgent to design various in silico methods to consider the diversity of mechanisms of synergism in promoting complementary

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46

actions[7] to predict synergistic drug combinations. Moreover, although historical validation was carried out in the experiment, we would like to try wet-lab experiment for validating top-rank drug pair that is not validated yet in the future work.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00793.

SimRank$_{LN}$ calculation; graph regularization; solving the objective function in SDCLR and DSGCR; drug−target protein network; protein−protein interactions network; precisions and recalls of various methods with fivefold cross-validation; and *P*-values obtained by paired *t*-test (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jiawei Luo − *Hunan University, Changsha, China*; Email: luojiawei@hnu.edu.cn

Other Authors

Pingjian Ding − *University of South China, Hengyang, China*; orcid.org/0000-0002-2613-2496

Cong Shen − *Hunan University, Changsha, China*

Zihan Lai − *Hunan University, Changsha, China*

Cheng Liang − *Shandong Normal University, Jinan, China*

Guanghui Li − *East China Jiaotong University, Nanchang, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.9b00793

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 61873089, 61862025, 61602283, and 61572180), Shandong Provincial Natural Science Foundation (No. ZR2016FB10), and Jiangxi Provincial Natural Science Foundation of China (No. 20181BAB211016).

■ REFERENCES

(1) Humphrey, R. W.; Brockway-Lunardi, L. M.; Bonk, D. T.; Dohoney, K. M.; Doroshow, J. H.; Meech, S. J.; Ratain, M. J.; Topalian, S. L.; Pardoll, D. M. Opportunities and challenges in the development of experimental drug combinations for cancer. *J. Natl. Cancer Inst.* **2011**, *103*, 1222−1226.

(2) Larder, B. A.; Kemp, S. D.; Harrigan, P. R. Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science* **1995**, *269*, 696−699.

(3) Meyer, C. T.; Wooten, D. J.; Paudel, B. B.; Bauer, J.; Hardeman, K. N.; Westover, D.; Lovly, C. M.; Harris, L. A.; Tyson, D. R.; Quaranta, V. Quantifying drug combination synergy along potency and efficacy axes. *Cell Syst.* **2019**, *8*, 97−108.

(4) Chou, T.-C. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacol. Rev.* **2006**, *58*, 621−681.

(5) Ding, P.; Luo, J.; Liang, C.; Xiao, Q.; Cao, B.; Li, G. Discovering synergistic drug combination from a computational perspective. *Curr. Top. Med. Chem.* **2018**, *18*, 965−974.

(6) McLornan, D.; Harrison, C. Combination therapies in Myeloproliferative Neoplasms: why do we need them and how to identify potential winners. *J. Cell. Mol. Med.* **2013**, *17*, 1410.

(7) Jia, J.; Zhu, F.; Ma, X.; Cao, Z. W.; Li, Y. X.; Chen, Y. Z. Mechanisms of drug combinations: interaction and network perspectives. *Nat. Rev. Drug Discovery* **2009**, *8*, 111.

(8) Cokol, M.; Chua, H. N.; Tasan, M.; et al. Systematic exploration of synergistic drug pairs. *Mol. Syst. Biol.* **2011**, *7*, 544.

(9) Griner, L. A. M.; Guha, R.; Shinn, P.; Young, R.; Keller, J.; Liu, D.; Goldlust, I.; Yasgar, A.; et al. High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 2349−2354.

(10) Tan, X.; Hu, L.; Luquette, L. J., III; Gao, G.; Liu, Y.; Qu, H.; Xi, R.; Lu, Z. J.; Park, P. J.; Elledge, S. J. Systematic identification of synergistic drug pairs targeting HIV. *Nat. Biotechnol.* **2012**, *30*, 1125.

(11) Liu, Y.; Wei, Q.; Yu, G.; Gai, W.; Li, Y.; Chen, X. DCDB 2.0: a major update of the drug combination database. *Database* **2014**, *2014*, No. bau124.

(12) Jeon, M.; Kim, S.; Park, S.; Lee, H.; Kang, J. In silico drug combination discovery for personalized cancer therapy. *BMC Syst. Biol.* **2018**, *12*, No. 16.

(13) Lee, J.-H.; Kim, D. G.; Bae, T. J.; Rho, K.; Kim, J.-T.; Lee, J.-J.; Jang, Y.; Kim, B. C.; Park, K. M.; Kim, S. CDA: combinatorial drug discovery using transcriptional response modules. *PLoS One* **2012**, *7*, No. e42573.

(14) Bansal, M.; Yang, J.; Karan, C.; Menden, M.; Costello, J.; Tang, H.; Xiao, G.; Li, Y.; Allen, J.; Zhong, R.; et al. A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* **2014**, *32*, 1213.

(15) Yang, J.; Tang, H.; Li, Y.; Zhong, R.; Wang, T.; Wong, S.; Xiao, G.; Xie, Y. DIGRE: drug-induced genomic residual effect model for successful prediction of multidrug effects. *CPT: Pharmacometrics Syst. Pharmacol.* **2015**, *4*, 91−97.

(16) Zhao, X.-M.; Iskar, M.; Zeller, G.; Kuhn, M.; Van Noort, V.; Bork, P. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput. Biol.* **2011**, *7*, No. e1002323.

(17) Xu, K.-J.; Song, J.; Zhao, X.-M. The drug cocktail network. *BMC Syst. Biol.* **2012**, *6*, No. S5.

(18) Zou, J.; Ji, P.; Zhao, Y.-L.; Li, L.-L.; Wei, Y.-Q.; Chen, Y.-Z.; Yang, S.-Y. Neighbor communities in drug combination networks characterize synergistic effect. *Mol. BioSyst.* **2012**, *8*, 3185−3196.

(19) Xu, Q.; Xiong, Y.; Dai, H.; Kumari, K. M.; Xu, Q.; Ou, H.-Y.; Wei, D.-Q. PDC-SGB: prediction of effective drug combinations using a stochastic gradient boosting algorithm. *J. Theor. Biol.* **2017**, *417*, 1−7.

(20) Huang, L.; Li, F.; Sheng, J.; Xia, X.; Ma, J.; Zhan, M.; Wong, S. T. DrugComboRanker: drug combination discovery based on target network analysis. *Bioinformatics* **2014**, *30*, i228−i236.

(21) Chen, X.; Ren, B.; Chen, M.; Wang, Q.; Zhang, L.; Yan, G. NLLSS: predicting synergistic drug combinations based on semi-supervised learning. *PLoS Comput. Biol.* **2016**, *12*, No. e1004975.

(22) Ding, P.; Yin, R.; Luo, J.; Kwoh, C. K. Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological and network knowledge. *IEEE J. Biomed. Health Inf.* **2019**, *23*, 1336−1345.

(23) Sheng, Z.; Sun, Y.; Yin, Z.; Tang, K.; Cao, Z. Advances in computational approaches in identifying synergistic drug combinations. *Briefings Bioinf.* **2017**, *19*, 1172−1182.

(24) Jeh, G.; Widom, J. In *SimRank: A Measure of Structural-Context Similarity*. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002; pp 538−543.

(25) Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644−1650.

(26) Hare, D.; Foster, T. The orange book: the food and drug administration's advice on therapeutic equivalence. *Am. Pharm.* **1990**, *30*, 35−37.
(27) Cheng, X.; Zhao, S.-G.; Xiao, X.; Chou, K.-C. iATC-mISF: a multi-label classifier for predicting the classes of anatomical therapeutic chemicals. *Bioinformatics* **2016**, *33*, 341−346.
(28) Cheng, F.; Kovács, I. A.; Barabási, A.-L. Network-based prediction of drug combinations. *Nat. Commun.* **2019**, *10*, No. 1197.
(29) Wishart, D.; Feunang, Y.; Guo, A.; Lo, E.; Marcu, A.; Grant, J.; Sajed, T.; Johnson, D.; et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074−D1082.
(30) Keshava Prasad, T. S.; Goel, R.; Kandasamy, K.; Keerthikumar, S.; Kumar, S.; Mathivanan, S.; Telikicherla, D.; Raju, R.; Shafreen, B.; Venugopal, A.; et al. Human protein reference database-2009 update. *Nucleic Acids Res.* **2009**, *37*, D767−D772.
(31) Sun, Y.; Sheng, Z.; Ma, C.; Tang, K.; Zhu, R.; Wu, Z.; Shen, R.; Feng, J.; et al. Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. *Nat. Commun.* **2015**, *6*, No. 8481.
(32) Lei, C.; Ruan, J. A novel link prediction algorithm for reconstructing protein-protein interaction networks by topological similarity. *Bioinformatics* **2013**, *29*, 355−364.
(33) Asur, S.; Ucar, D.; Parthasarathy, S. An ensemble framework for clustering protein-protein interaction networks. *Bioinformatics* **2007**, *23*, i29−i40.
(34) Cao, B.; Luo, J.; Liang, C.; Wang, S.; Ding, P. Pce-fr: A novel method for identifying overlapping protein complexes in weighted protein-protein interaction networks using pseudo-clique extension based on fuzzy relation. *IEEE Trans. Nanobiosci.* **2016**, *15*, 728−738.
(35) Shi, C.; Kong, X.; Huang, Y.; Philip, S. Y.; Wu, B. Hetesim: A general framework for relevance measure in heterogeneous networks. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 2479−2492.
(36) Qu, J.; Chen, X.; Sun, Y.; Zhao, Y.; Cai, S.; Ming, Z.; You, Z.; Li, J. In silico prediction of small molecule miRNA associations based on the HeteSim algorithm. *Mol. Ther. − Nucleic Acids* **2019**, *14*, 274−286.
(37) Zeng, X.; Liao, Y.; Liu, Y.; Zou, Q. Prediction and validation of disease genes using HeteSim Scores. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2017**, *14*, 687−695.
(38) Sun, Y.; Han, J.; Yan, X.; Yu, P. S.; Wu, T. In *Pathsim: Meta Path-Based Top-k Similarity Search in Heterogeneous Information Networks*. Proceedings of the VLDB Endowment, 2011; pp 992−1003.
(39) Wang, J. Z.; Du, Z.; Payattakool, R.; Yu, P. S.; Chen, C.-F. A new method to measure the semantic similarity of GO terms. *Bioinformatics* **2007**, *23*, 1274−1281.
(40) Zeng, X.; Liu, L.; Lv, L.; Zou, Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* **2018**, *34*, 2425−2432.
(41) Luo, J.; Ding, P.; Liang, C.; Chen, X. Semi-supervised prediction of human miRNA-disease association based on graph regularization framework in heterogeneous networks. *Neurocomputing* **2018**, *294*, 29−38.
(42) Wang, F.; Huang, Z.-A.; Chen, X.; Zhu, Z.; Wen, Z.; Zhao, J.; Yan, G.-Y. LRLSHMDA: laplacian regularized least squares for human microbe-disease association prediction. *Sci. Rep.* **2017**, *7*, No. 7601.
(43) Xiao, Q.; Luo, J.; Liang, C.; Cai, J.; Ding, P. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* **2018**, *34*, 239−248.
(44) Long, M.; Wang, J.; Ding, G.; Shen, D.; Yang, Q. Transfer learning with graph co-regularization. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 1805−1818.
(45) Ji, M.; Sun, Y.; Danilevsky, M.; Han, J.; Gao, J. Graph Regularized Transductive Classification on Heterogeneous Information Networks. In *Machine Learning and Knowledge Discovery in Databases*, 2010; pp 570−586.
(46) Hajian-Tilaki, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J. Intern. Med.* **2013**, *4*, 627.
(47) Davis, J.; Goadrich, M. In *The Relationship Between Precision-Recall and ROC Curves*. Proceedings of the 23rd International Conference on Machine Learning, 2006; pp 233−240.
(48) Hsu, H.; Lachenbruch, P. A. Paired *t* Test. In *Wiley Encyclopedia of Clinical Trials*, 2007; pp 1−3.
(49) Ling, C. X.; Huang, J.; Zhang, H. In *AUC: A Better Measure than Accuracy in Comparing Learning Algorithms*. Conference of the Canadian Society for Computational Studies of Intelligence, 2003; pp 329−341.
(50) Wang, C.-C.; Chen, X.; Qu, J.; Sun, Y.-Z.; Li, J.-Q. RFSMMA: a new computational model to identify and prioritize potential small molecule-miRNA associations. *J. Chem. Inf. Model.* **2019**, *59*, 1668−1679.
(51) Bryant, V. L.; Elias, R. M.; Mccarthy, S. M.; Yeatman, T. J.; Alexandrow, M. G. Suppression of reserve MCM complexes chemosensitizes to gemcitabine and 5-fluorouracil. *Mol. Cancer Res.* **2015**, *13*, 1296−1305.
(52) Gutierrez-Delgado, F.; Lopez-Mariscal, A.; Maldonado-Hernandez, H.; Luna-Benitez, I.; Salazar-Macias, F.; Aceves-Escarcega, A.; Delgadillo-Hernandez, J. Oxaliplatin and cyclophosphamide as neoadjuvant chemotherapy (NACT) followed by surgery for patients with locally advanced cervical cancer (LACC). A preliminary report. *J. Clin. Oncol.* **2005**, *23*, No. 5173.
(53) Ketter, T. A.; Pazzaglia, P. J.; Post, R. M. Synergy of carbamazepine and valproic acid in affective illness: case report. *J. Clin. Psychopharmacol.* **1992**, *12*, 276−281.
(54) Iorio, F.; Knijnenburg, T.; Vis, D.; Bignell, G.; Menden, M.; Schubert, M.; Aben, N.; et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **2016**, *166*, 740−754.

https://dx.doi.org/10.1021/acs.jcim.9b00793
J. Chem. Inf. Model. 2020, 60, 37−46