Journal of Biomedical Informatics 68 (2017) 167–183

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Heter-LP: A heterogeneous label propagation algorithm and its application in drug repositioning

Maryam Lotfi Shahreza [a], Nasser Ghadiri [a,*], Seyed Rasoul Mousavi [b,c], Jaleh Varshosaz [d], James R. Green [e]

[a] Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran
[b] Department of Computer Engineering, Amirkabir University of Technology, Tehran 15916-34311, Iran
[c] School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[d] Department of Pharmaceutics, School of Pharmacy and Pharmaceutical Science, Isfahan University of Medical Sciences, Isfahan, Iran
[e] Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada

ARTICLE INFO

Article history:
Received 6 November 2016
Revised 9 February 2017
Accepted 10 March 2017
Available online 11 March 2017

Keywords:
Semi-supervised learning
Heterogeneous networks
Label propagation
Drug-disease associations
Drug–target interactions
Disease-target interactions

ABSTRACT

Drug repositioning offers an effective solution to drug discovery, saving both time and resources by finding new indications for existing drugs. Typically, a drug takes effect via its protein targets in the cell. As a result, it is necessary for drug development studies to conduct an investigation into the interrelationships of drugs, protein targets, and diseases. Although previous studies have made a strong case for the effectiveness of integrative network-based methods for predicting these interrelationships, little progress has been achieved in this regard within drug repositioning research. Moreover, the interactions of new drugs and targets (lacking any known targets and drugs, respectively) cannot be accurately predicted by most established methods.

In this paper, we propose a novel semi-supervised heterogeneous label propagation algorithm named Heter-LP, which applies both local and global network features for data integration. To predict drug-target, disease-target, and drug-disease associations, we use information about drugs, diseases, and targets as collected from multiple sources at different levels. Our algorithm integrates these various types of data into a heterogeneous network and implements a label propagation algorithm to find new interactions. Statistical analyses of 10-fold cross-validation results and experimental analyses support the effectiveness of the proposed algorithm.

1. Introduction

The process of finding additional indications for existing drugs is known as Drug Repositioning (DR). DR is a way to save time and costs when compared to the *de novo* drug development process. Computational methods can guide lab experimental design by narrowing the scope of candidate targets to accelerate drug discovery and can provide supporting evidence for experimental results.

Recent research shows that simultaneous use of the three concepts of diseases, drugs, and targets together leads to better results for drug repositioning [1]. Specifically, the application of network-based approaches in the fields of genomics, transcriptomics, proteomics, and systems biology have the potential to improve drug development. These improvements will help decrease the time between lead development and drug marketability [2].

The use of heterogeneous networks in DR is motivated by the fact that drugs tend to take effect via interaction with one or more protein targets within a cell. Therefore, it is necessary to consider drugs, protein targets, and diseases simultaneously to investigate their interrelationships.

The research objective of this paper is to provide a computational framework to facilitate drug repositioning tasks based on a heterogeneous network, generated by integrating drug, disease, and target information in different levels.

The contribution of this paper is twofold. (1) *Construction of a heterogeneous network*: This network is composed of six networks: the drug similarity network, disease similarity network, target similarity network, known drug-disease associations, known drug–target interactions and known disease-target interactions. (2) *Algorithmic prediction of different potential interactions*: In this research we develop a heterogeneous label propagation algorithm to predict potential drug–target interactions,

* Corresponding author.
E-mail addresses: maryam.lotfi@ec.iut.ac.ir (M. Lotfi Shahreza), nghadiri@cc.iut.ac.ir, nghadiri@gmail.com (N. Ghadiri), srm@aut.ac.ir (S.R. Mousavi), varshosaz@pharm.mui.ac.ir (J. Varshosaz), jrgreen@sce.carleton.ca (J.R. Green).

http://dx.doi.org/10.1016/j.jbi.2017.03.006
1532-0464/

drug-disease associations, and disease-target interactions by integrating multi-source information. The reason for this choice is that heterogeneous network label propagation is an effective and efficient technique to utilize both local and global features in a network for semi-supervised learning [3]. We use Heter-LP to develop a new drug repositioning method, by performing label propagation to integrate different levels of biological information and apply an optimization algorithm to find new drug-target interactions. The evaluation is based on a 10-fold cross-validation experiment design and we analyze the results using performance metrics such as the Area Under the Receiver Operating Characteristic Curve (AUC) and Area Under the Precision-Recall curve (AUPR).

The goals of this research are to identify putative candidates for drug repositioning, to further improve the prediction accuracy of drug-target interactions, and to discover useful drug-disease and disease-target associations. In this application, the inputs to our algorithm are three similarity matrices and three association matrices which are generated using three different levels of information (molecular originated profiles, molecular activity information, and phenotypic properties). The primary output will be three matrices representing drug-target interactions, drug-disease associations, and disease-target associations. The final output is a ranked list of candidates for drug repositioning. Secondary outputs include new similarity matrices for drugs, diseases, and targets which would have application in, for instance, clustering.

Unlike existing methods, our proposed method can predict interactions of *new* drugs (where a drug has no known target) and *new* targets (where a target has no known drug). Moreover, in our approach, there is no requirement for negative training exemplars. The other benefit of the proposed method is its ability to predict both trivial and non-trivial interactions. We believe that meaningful and efficient integration of information is achieved due to our use of an appropriate structural network model and suitable label propagation algorithm. Moreover, the pre-phase projection phase enriches the algorithm. The statistical and experimental analysis will demonstrate these claims.

The paper is organized as follows. In following sub-section the related work is represented. A complete description of our proposed method and material is provided in Section 2. Proof of the convergence of the method and the regularization framework is also presented in this section. The performance evaluation and a brief analysis of computational complexity of the algorithm is presented in Section 3. Finally, Section 4 provides a summary of the research.

1.1. Related work

Gene expression patterns change systematically in response to disease processes. Transcriptome data provide a snapshot of such whole-genome dynamics and can provide insights into the mechanism of action of drugs [4]. Differential gene expression analysis is an effective way to identify genes that lead to disease. In this regard, some drug repositioning methods have been developed based on gene expression analysis, such as [5–11]. In spite of observed good performance, there are limitations associated with such methods which compare gene expression signatures. First, the set of drugs and diseases included in current databases of gene expressions are limited, so these methods are restricted to the subset of known diseases (such as particular kinds of cancers). So it seems that we cannot rely solely on such limited data and we need more data sources to complement them. Second, results coming from cell lines cannot always be extrapolated to *in vivo* tissues.

Protein-protein interaction (PPI) is often used as the basis for drug target identification because the PPI network provides the context in which the target protein operates. This approach assumes that the proteins targeted by similar drugs tend to be functionally associated and be close in the PPI network. Here, *similar* drugs refers to drugs with "similar therapeutic effects" [4]. Among the most important drug repositioning methods based on PPI are [12,13]. Despite demonstrated success in repositioning drugs using PPI networks, there are also some limitations. First, the required PPI data are noisy and incomplete, and the extracted networks are incomplete and biased [14]. Second, as in the case in gene regulatory networks, there is no simple mapping between a simulated network and a living organism's actual response [14].

By using a metabolic network, several important physiological properties of a cell could be extrapolated. So metabolic networks can also be used to predict drug targets. Two drug repositioning methods based on metabolic networks are described in [15,16]. Most of these methods are based on Flux Balance Analysis (FBA) which is sensitive to the selected metabolites, their defined objective function, and the constraints of the objective function [17]. In other words, the strength of predictions depends on the appropriate definition of biomass reactions.

The identification of novel drug-target interactions (DTI) is the basis for drug discovery and design, and accurate prediction of drug targets is a key to effective drug repositioning. Many drugs are non-specific and show reactivity to additional targets besides the primary targets. Although these off-target effects often lead to unwanted side effects (discussed below), these one-drug-multiple-target data can also be leveraged by DR methods. Motivated by this, researchers have developed many methods based on drug-target interactions, including [18–32]. Some of these methods are based strictly on DTI, while others are extended to leverage additional data such as protein-protein similarity and drug-drug similarity,[1] often leading to increased performance. For example, when a given drug has several known targets, additional candidate targets can be ranked by calculating the similarity between candidate targets and known targets. However, in the case where a drug has no known target, (i.e. a new drug), computing target similarity is not possible. Hence, drug similarity must be used instead. In this case, potential targets of this new drug are selected based on target information for similar drugs for which target data is available. Semi-supervised learning methods can address the problem of predicting interactions for new drugs (and new targets).

Further drug repositioning methods, which leverage drug-drug similarity, include [33,34]. One of the main limitations of drug repositioning based on chemical structure similarity of drugs is that many structures and chemical properties of known drug compounds are inaccurate. Furthermore, many physiological effects of a drug cannot be predicted by structural properties alone [35].

Computational assessment of similarities in molecular profiles is another approach for relating drugs to disease states for the purpose of repositioning [35,36]. The role of molecular profile could be described as a signature of molecular activity after exposing a drug to a biological system. It may contain different measures such as a change in transcriptional activity. The similarity of these profiles could be used to establish useful relationships between drugs and diseases. There have been important methods of this type in the literature, e.g., [37–39].

Many drugs induce some unintended effects in the living organism besides the primary desired effects, which constitute a drug's overall effect profile. Those wanted or unwanted behavioral or physiological changes in response to drug treatment can be measured as a drug's indication and side effects, respectively [14,40]. We know that side effects are generated when a drug binds to

[1] There are many metrics to measure the similarity between two drugs. For example, some metrics are based on similarity of biological effects while others are based on similarity of drug chemical structure, etc. In this paper, specific types of similarities will be clarified explicitly whenever needed; otherwise, the term "drug similarity" can refer to any type of similarity.

off-targets, which perturb unexpected metabolic or signaling pathways. These off-targets may, in fact, lead to the identification of novel therapeutic targets [4,41]. So pharmacological information associated with drugs provides an alternative way to predict drug targets; this approach has proven to be complementary with the commonly used molecular information. Ye et al. [41] tried to reposition drugs by statistical analysis of drug side effects. The PREDICT [38] algorithm also used drug side effects to rank drug-disease associations.

There are some limitations to use side-effects. First, the scarcity of drug adverse reaction information limits the application of this kind of approach. In fact, side-effect-similarity approaches need a well-defined side effect profile for a drug, while current disease and drug phenotype data are noisy and far from complete. Second, there is no side effect profile available for newly approved drugs [14,4,35]. Third, all drugs with similar effects do not necessarily possess similar targets. Fourth, there is no simple mapping between phenotype and mechanism of action. The living organism's genetic map, medication history, and other traits could affect the phenotypic outcomes of a drug. So, we could not conclude that a similar phenotype corresponds to the same mode of action [14,4]. Finally, it should be noted that side effect information could be confused by a patient's medication history, genotype, and other hidden factors [14].

In drug repositioning, it is assumed that if molecular pathophysiology of therapeutic effects of two drugs has sufficient commonalities, they are interchangeable [35]. So to reposition drugs, we require computational strategies for finding molecular relationships between distinct disease pathologies. This approach has been leveraged for DR in [40].

Previous studies indicate that integrative analysis, where multiple lines of evidence are considered simultaneously, is a practical approach to finding the most probable candidates for drug repositioning. Recently, some novel integrative methods have been proposed, include PreDr [42], Yamanishi et al. [43], TL_HGBI [44], SLAMS [1], NRWRH [45].

While integrative methods have been shown to outperform other approaches, these methods still face some common limitations outlined below:

(1) Most existing similarity-based prediction algorithms use only immediate similarities and don't consider transitivity of similarity. To address this deficiency, Zhang et al. [46] proposed a label propagation approach that considers higher-order similarity which could be useful in drug repositioning research and help us in this regard.
(2) As mentioned above, in many methods negative drug-target interactions are selected randomly without experimental confirmation [31].
(3) Interactions with new drugs (drugs without any known target) and new targets (targets without any known drug) cannot be predicted by some methods [47]. Semi-supervised learning methods could be useful in addressing this problem.
(4) Most existing methods are based on one or two kinds of data (like chemical structure similarity of drugs or sequence similarity of protein targets). On the other hand, existing training samples of established methods are very few when compared with all available unlabeled data.

Our proposed method is a semi-supervised method without requiring negative training samples, and capable of utilizing information from unlabelled samples. Furthermore, it is applicable in the case of the new entity problem. Lastly, integration of different data types is used to improve the prediction accuracy.

2. Materials and method

We design a novel algorithm to predict drug repositioning by associating known drugs with new diseases, different diseases with new targets, and drugs with new targets.

This section will introduce the Heter-LP method, using drug repositioning as the illustrative application example. Section 2.1 presents the formal notations and settings used in the problem. Section 2.2 covers data collection. Section 2.3 is about the projection step of the algorithm. Section 2.4 explains the label propagation algorithm. Finally, in Section 2.5, pseudo-code of the algorithm is presented.

The proposed data model consists of six parts, three homogeneous (1, 2, 3) and three heterogeneous (4, 5, 6) sub-networks: (1) Drug similarity network, (2) Disease similarity network, (3) Target similarity network, (4) Known drug-disease associations, (5) Known drug–target interactions, (6) Known disease-target associations.

A comprehensive description of using data for each part is presented in Section 2.2. Our aim is to optimally integrate these different data sources and provide a ranked list of putative novel associations between drugs, diseases, and targets. Fig. 1 is a schematic view of our heterogeneous network model and used datasets.

2.1. Notations and problem settings

There are three types of nodes in the proposed heterogeneous network: drugs, diseases, and targets. There are six different kinds of edges, each of which represents one type of similarity or association: drug similarity, disease similarity, target similarity, known drug-disease association, known drug–target interaction and known disease-target interaction.

Therefore, we have a heterogeneous graph $G = (V, E)$ with three homo-sub-networks and three hetero-sub-networks (Fig. 1). The homo-sub-networks are defined as $G_i = (V_i, E_i)$ where $i = 1, 2, 3$ for drugs, diseases, and targets, respectively. The hetero-sub-networks are: $G_{i,j} = (V_i \bigcup V_j, E_{i,j})$ for $i, j = 1, 2, 3$, and $i < j$.

Each E_i is the set of edges between vertices in the vertex set V_i of homo-sub-network G_i. On the other hand, each $E_{i,j} \subseteq V_i \times V_j$ is the set of edges connecting a vertex in V_i to a vertex in V_j. So in $G, V = \{V_1 \cup V_2 \cup V_3\}$ and $E = \{E_1 \cup E_2 \cup E_3 \cup E_{1,2} \cup E_{1,3} \cup, E_{2,3}\}$.

We represent the inputs of homo-subnetworks by one $n_i \times n_i$ affinity matrix[2] A_i, where n_i is the number of vertices in corresponding homo-subnetwork and $A_i(k, k') \geq 0$ is the similarity between entity k and k'. For example, the input drug network is represented by an $|V_1| \times |V_1|$ element square symmetric affinity matrix A_1, where $A_1(k, k') \geq 0$ is the similarity between drugs k and k'. For each hetero-subnetwork, there is a relationship matrix $A_{i,j}$ with $|V_i|$ rows and $|V_j|$ columns. Each entry $A_{i,j}(k, k') \in [0,1]$ reflects the strength of the relationship between entity k and entity k′, respectively. For example, the input drug-target network is represented by a $|V_1| \times |V_3|$, binary matrix, $A_{1,3}$. All A_i and $A_{i,j}$ matrices must be normalized (once at initialization) to ensure convergence of the updates. The normalized matrices are denoted by S_i and $S_{i,j}$.

In Fig. 2, we try to clarify the process using the workflow. A brief description of each part and the data preparation methods are represented in next subsections.

2.2. Data preparation

In this section, the characteristics of the datasets that are used in this research will be studied. In fact this section is a description of part (A) in Fig. 2. We used a gold standard dataset provided by

[2] All homogenous sub-networks are affinities symmetric.

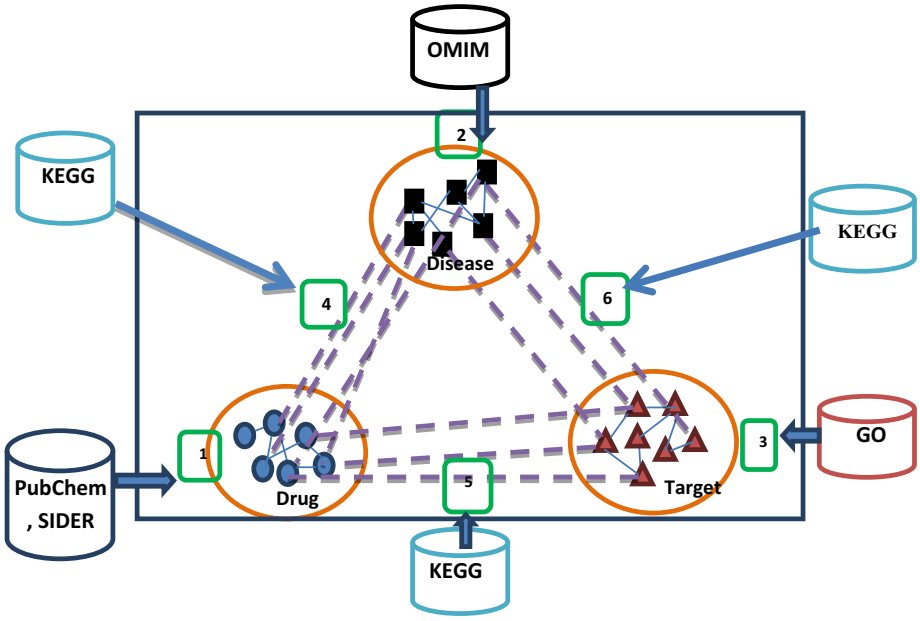


Fig. 1. Heterogeneous network model and used datasets. Parts (1), (2) and (3) are homo-sub-networks and parts (4), (5) and (6) are hetero-sub-networks. Each cylinder in the outside part of the box represents the corresponding dataset used for each part. Different shapes are used to represent different concepts: Blue circles for drugs, Black square for each disease, Orange triangle for each target, the thin line for related similarity (weighted) and the dashed line for the existence of association (not weighted). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Yamanishi (2008) [22] to compare prediction accuracy of the proposed method with previous methods more accurately. We also gathered several independent datasets to provide a more realistic experimental analysis. Further details are provided in the next subsections.

2.2.1. Gold standard dataset

Yamanishi et al. [22] provide a dataset containing drugs, protein targets, and their interactions which are categorized by four groups of protein targets (Enzyme, GPCR, Ion Channel, and Nuclear Receptor). The primary resources of these data are KEGG,[3] BRITE,[4] BRENDA,[5] SuperTarget,[6] and DrugBank.[7] The similarity-by structure of drugs is computed by SIMCOMP on chemical substructures, and similarity of targets is calculated by a normalized version of Smith-Waterman score. Some of the most important characteristics of these data are represented in supplementary materials.

Although this dataset provides a useful benchmark for comparison of different methods, it has some limitations. In addition to being outdated, there is no information about associated diseases of drugs and targets in this dataset. Since the concept of disease is a fundamental element in our model, we had to incorporate some extra data. For this purpose, we provided a list of related diseases for each mentioned group of gold standard datasets based on data provided by Li et al. [49] and DisGeNet [50], for drugs and targets respectively, and provided disease similarity from OMIM [51]. For example, we extracted all diseases related to drugs from the Enzymes files of [22], from drug-disease lists of [49] in the form of a separate matrix, and in the same way for other groups. Code written for this purpose is available upon request.

2.2.2. Independent test datasets

In addition to the gold standard dataset, we build an updated heterogeneous dataset gathered from a number of datasets representing the three key concepts used by the model (drug, disease, and target) and their interrelationships. A brief description of each dataset is provided below.

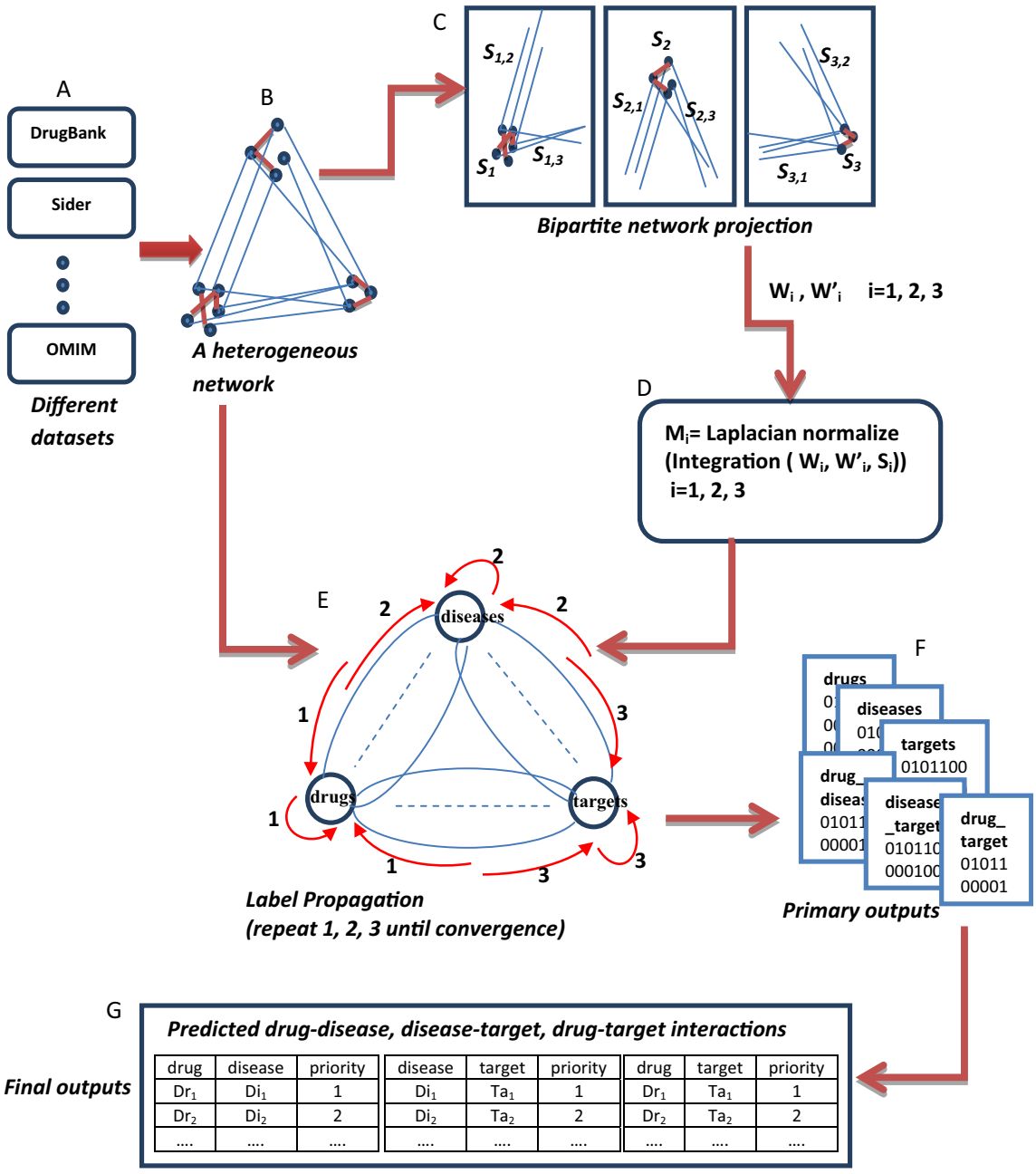
Homogeneous sub-graphs

G1. ***Chemical substructures:*** It is believed that drugs with similar chemical structure carry out similar therapeutic functions, thus, are likely to treat common diseases [42].

In this research, the pairwise similarity of two drugs is calculated based on the 2*D* chemical fingerprint descriptor of each chemical structure in PubChem. That is, each drug d is represented by a binary fingerprint $h(d)$ in which each bit represents a predefined chemical structure fragment. Each $h(d)$ is an 881-dimensional chemical substructure vector defined in PubChem. These data were obtained from [52],[8] which contains 888 drugs and 881 substructures, and the description of the 881 chemical substructures is available at PubChem's website. The pairwise chemical similarity between two drugs d_1 and d_2 is computed as the Tanimoto coefficient of their fingerprints using the "proxy" R package[9] and gathered as a chemical substructure similarity matrix ($g1$).

Side effect similarities: Many drugs have adverse effects in addition to their indication, referred to as side effects. It has been shown that more accurate drug-target prediction can be achieved by integrating these side effects with other information, such as chemical similarities [4]. SIDER [53] is the primary resource for side effect data. We found 888 drugs and 1385 side effects in this online dataset. These data also represented drug-side-effect relationships using a fingerprint matrix (like one described in chemical structure similarity). The similarity of side effects between two drugs d_1 and d_2 is computed as the Tanimoto coefficient of their fingerprints using the "proxy" R package and gathered as a side effect similarity matrix ($g2$). This distance is integrated with the chemical structure similarity matrix to produce the G_1 sub-network. The combination of two drug-drug similarity matrices g_1, g_2 is described by Eq. (1) below. Note that not all 888 drugs with structural fingerprint data had side effect information and vice versa.

[3] http://www.kegg.jp.
[4] http://www.genome.jp/kegg/brite.html.
[5] http://www.brenda-enzymes.org/.
[6] http://insilico.charite.de/supertarget/.
[7] http://www.drugbank.ca.
[8] http://cbio.ensmp.fr/yyamanishi/side-effect.
[9] https://cran.r-project.org/web/packages/proxy/index.html.



drug	disease	priority	disease	target	priority	drug	target	priority
Dr_1	Di_1	1	Di_1	Ta_1	1	Dr_1	Ta_1	1
Dr_2	Di_2	2	Di_2	Ta_2	2	Dr_2	Ta_2	2
....	....	....	....	....	....	....	....	....

Fig. 2. The overall process workflow for Heter-LP algorithm. (A) A schematic view of the datasets, explained in Section 2.2 (B) The constructed network explained in Section 2.1 and Fig. 1; (C) Our projection phase (equivalent to pseudo code lines 2–7 of Algorithm 1) explained in Section 2.3.; (D) Generate a Laplacian normalization view of projection output (equivalent to pseudo code lines 8–10 of Algorithm 1). (E) The label propagation phase of the algorithm (equivalent to pseudo code lines 11–13 of Algorithm 1) explained in Section 2.4. Arrows labeled with 1 correspond to line 11 and so on. (F) The six primary output matrices. Three of them are related to homo-sub-networks and three are related to hetro-sub-networks. Explanation is available in the end of Section 2.5. (G) Final outputs are three sorted lists of predicted interactions. Explanation is available in the end of Section 2.5.

$$integration\ (g_1, g_2) = g$$

$$g(i,j) = \begin{cases} (g_1(i,j) + g_2(i,j))/2 & if\ entry(i,j)\ exists\ in\ both\ g_1\ and\ g_2 \\ 2g_1(i,j)/3 & if\ entry\ (i,j)\ does\ not\ existing_2 \\ 2g_2(i,j)/3 & if\ entry\ (i,j)\ does\ not\ existing_1 \end{cases} \quad (1)$$

It should be noted that Eq. (1) performs a simple weighted averaging function. At the first attempt for combining the two matrices, we only calculated an average for each entry in all cases as $g(i,\ j) = \frac{(g_1(i,j)+g_2(i,j))}{2}$, and set $g_1(i,j) = 0$ if entry (i,j) does not exist in g_1, and the same for g_2. It was observed that this calculation caused an undesirable reduction of the resulting value in the second and third cases of Eq. (1). Therefore, in these two cases we assigned an extra weight (2/3) to decrease the effect of the missing input in the averaging process. An additional advantage of this weighted averaging is that we could differentiate the effectiveness of truly zero value entries from missing/imputed values.

G2. ***Semantic similarity of disease phenotypes:*** Text mining techniques were utilized to classify human phenotypes contained in the Online Mendelian Inheritance in Man (OMIM) database [51]. The phenotype similarity data are accessible through the website at http://www.cmbi.ru.nl/MimMiner/. A matrix of pairwise semantic similarity between diseases is available from the mentioned website. This similarity is calculated based on the number of

co-occurring MeSH (Medical Subject Headings) terms in the specialist descriptions of each disease pair from the OMIM database. In other words, each row (or column) is the phenotype similarity profile for a single disease. The number of diseases in this matrix is 4784.

G3. *Gene Ontology-based Sematic Similarity Measures:* A drug target is a human protein, whose activity is modified by a drug, resulting in a desirable therapeutic effect [1]. Semantic similarity measures can be used to estimate the functional similarities among Gene Ontology terms and gene products. The pairwise semantic similarity protein pairs (drug targets) was computed using the "GOSemSim" package in R [54]. We have 1537 protein targets in this research.

Heterogeneous sub-graphs

Wu et al. [39] provided a useful list of relationships between drugs, diseases, and protein targets extracted from KEGG. We divided these data into three separate lists of drug-disease, disease-target, and drug-target relationships and then, by using the "reshape2" [55] R package, converted them into corresponding fingerprint matrices. So we have a drug-disease matrix, $G_{1,2}$, containing 584 drugs and 203 diseases with 1041 relationships between them; a drug-disease matrix, $G_{1,3}$, containing 3592 drugs and 1504 targets with 11610 relationships between them; and a disease-target matrix, $G_{2,3}$, containing 1087 diseases and 2255 targets with 3296 relationships. The number of each type of entity in different sub-networks is shown in supplementary materials.

In the present work, drugs and diseases are referred to by their names, while targets are referred to by their KEGG IDs. Conversion between IDs is accomplished using available online tools like Synergizer.[10]

2.3. Projection

Bipartite networks have received considerable attention by research community in different scientific areas. The vertex set in a bipartite network consists of two nonempty disjoint vertex sets X and Y such that each edge has one endpoint in X and another in Y. Let $A_{n\times m}$ be the adjacency matrix representing bipartite network $G = (X \cup Y, E)$. To find a direct association between vertices in a particular set of vertices, say X, we use "one-mode projection", by which we mean to obtain a network containing only X vertices from the original bipartite network G. In this newly generated network, two vertices are connected if they have at least one common neighbor in G. The output of the projection may be weighted or unweighted; the weighted type is usually preferred. To improve the projection, the similarity matrix of the vertices in X (S_x) and the similarity matrix of the vertices in Y (S_y) can also be used.

We will use a weighted one-mode projection technique based on the method represented by Zhou et.al [56]. The weight of edge (i, j) (in the resulting network) is calculated by Eq. (2):

$$w(i,j) = \frac{s_x(i,j)}{K(x_i)^{1-\lambda}K(x_j)^{\lambda}} \sum_{l=1}^{m} \frac{a(i,l) * a(j,l)}{K(y_l)} \tag{2}$$

where

- W is the projected matrix of A onto X and $w(i,j)$ is the entry of row i and column j in W.
- S_x is similarity matrix of vertex set X and $s_x(i,j)$ is the entry of row i and column j in S_x.
- $K(x_i)$ and $K(y_i)$ are the degrees of vertices x_i and y_i, respectively, in G.
- $0 < \lambda < 1$ is diffusion parameter of the projection, which in an input to the proposed algorithm.
- $a(i, l)$ represents the weight of edge (x_i, y_l) in G.

[10] http://llama.mshri.on.ca/synergizer/translate/.

There is no edge from i to j if, and only if, w(i, j) = 0. As shown in Fig. 2 part (C), the primitive heterogeneous network includes three bipartite sub-networks $G_{i,j} = (V_i \cup V_j, E_{i,j})$, $1 \leqslant i \leqslant j \leqslant 3$. We apply our one-mode projection technique twice for each of these subnetworks separately. For example, in one run we use drug-target interactions matrix as A, drugs as X and targets as Y, and in the other we use target-drug interactions as A, targets as X and drugs as Y. This process is repeated for the other two sub-networks (corresponding to drug-disease and disease-target interactions). So we will have six different projection matrices. Each matrix is used as a topological similarity matrix in the next section to improve the label propagation accuracy. We call them topological similarity matrices because of their ability to show different relationships by weighted edges between vertices of the same type directly.

2.4. Label propagation

As noted in Section 1, we develop a heterogeneous label propagation algorithm to predict different types of the potential interactions in the network. In the naïve Label Propagation (LP) algorithm, there is an undirected weighted network with n nodes, some of which are labeled, and the goal is to estimate the labels of the unlabeled nodes. In each iteration, we have only one labeled node, and we attempt to predict the labels of others. In drug-drug homogeneous matrix the predicted labels would indicate similarities of drugs, and in drug-target heterogeneous matrix the predicted labels would indicate the existence of an interaction between corresponding drug and target.

The label propagation algorithm is closely related to the Random Walk (RW) algorithm. There are two major differences between LP and RW: (i) LP fixes the labeled points, and (ii) the solution of LP is an equilibrium state, whereas RW is dynamic [30].

In this study, we require an algorithm to propagate label information across a heterogeneous network of drugs, targets, and diseases. A heterogeneous network is a network that consists of several sub-networks of different types of vertices and edges. For instance, in a drug-target hetero-network, there are two kinds of nodes and three kinds of edges. Edges between two drugs and between two targets are weighted as an explanation of their similarity. The edges between a drug and a target are un-weighted edges which are here binary, indicating the presence or absence of a relationship between the drug and the target. Most graph-based label propagation algorithms propagate label information only on a homogenous network, which are not suitable for spreading label information across heterogeneous networks. In this regard, Hwang et al. [57] proposed a heterogeneous label propagation algorithm, named MINProp. This method sequentially propagates the label on each sub-network. Another heterogeneous label propagation algorithm is LPMIHN [30], with the primary purpose of inferring potential drug–target interactions using a heterogeneous network. The LPMIHN algorithm propagates labels on each homogenous sub-network separately. Then the interactions between the two homogenous sub-networks are used only as extra information in the form of a similarity matrix.

In Heter-LP we apply label propagation on each sub-network using the existing information derived from the other sub-networks (part E of Fig. 2). The process repeats until convergence.

In brief, the inputs to the proposed algorithm are the similarity matrices and interaction matrices. In each iteration, the aim is to find the relationships between each pair of entities using these inputs. We initially set the label of a particular entity to one and all others to zero. This label information is propagated through the entire network to determine the relationships between the investigated entity and all others as newly assigned labels emerge. These new labels are saved in three vectors and, before the next iteration, are saved in specific matrices. Finally, these output

matrices are sorted from largest to smallest value of achieved label and determine the most important relationships as the top scoring elements in each matrix.

2.5. The proposed algorithm

To clarify details of the proposed method, the pseudo code of our Heter-LP algorithm is presented below as Algorithm 1.

Algorithm 1. Heter-LP

Input

1) σ: total convergence threshold
2) σ': homogenous convergence threshold
3) α: diffusion parameter of label propagation
4) λ: diffusion parameter of projection
5) y_1, y_2, y_3: vectors of initial label values
6) S_1, S_2, S_3: homo-subnetwork matrices
7) $S_{1,2}, S_{1,3}, S_{2,3}$: hetero-subnetwork matrices
8) drugs list (n_1 is the number of total drugs)
9) diseases list (n_2 is the number of total diseases)
10) targets list (n_3 is the number of total targets)

Output

1) F_1, F_2, F_3: homo-subnetwork matrices of final label values
2) $F_{1,2}, F_{1,3}, F_{2,3}$: hetero-subnetwork matrices of final label values

Algorithm

1. $F_k = 0$, $F_{k,k'} = 0$ for all k, $k' = 1, 2, 3, k<k$
   //Projection
2. $W1_{n1*n1}$ = projection of $S_{1,2}$ on S_1
3. $W'1_{n1*n1}$ = projection of $S_{1,3}$ on S_1
4. $W2_{n2*n2}$ = projection of $S_{1,2}$ on S_2
5. $W'2_{n2*n2}$ = projection of $S_{2,3}$ on S_2
6. $W3_{n3*n3}$ = projection of $S_{1,3}$ on S_3
7. $W'3_{n3*n3}$ = projection of $S_{2,3}$ on S_3
   //Integration of similarity matrix with projected matrices
8. M_1 = `NormalizeSumOf`(S_1, W_1, W'_1)
9. M_2 = `NormalizeSumOf`(S_2, W_2, W'_2)
10. M_3 = `NormalizeSumOf`(S_3, W_3, W'_3)
    // label propagation
11. **for** $i = 1 \ldots y_1.length$
    11.1) $y_1[i] = 1$, $y_1[j] = 0$ **for all** $j \neq i$
    11.2) $y_2 = y_3 = 0$
    11.3) $f_1 = y_1$, $f_2 = y_2$, $f_3 = y_3$ // vectors of final label values
    11.4) `LabelPropagation`(f_1, f_2, f_3)
    11.5) update $F_1, F_{1,2}, F_{1,3}$
12. **for** $i = 1 \ldots y_2.length$
    12.1) $y_2[i] = 1$, $y_2[j] = 0$ **for all** $j \neq i$
    12.2) $y_1 = y_3 = 0$
    12.3) $f_1 = y_1, f_2 = y_2, f_3 = y_3$
    12.4) `LabelPropagation`(f_1, f_2, f_3)
    12.5) update $F_2, F_{2,1}, F_{2,3}$
13. **for** $i = 1 \ldots y_3.length$
    13.1) $y_3[i] = 1$, $y_3[j] = 0$ **for all** $j \neq i$
    13.2) $y_1 = y_2 = 0$
    13.3) $f_1 = y_1$, $f_2 = y_2$, $f_3 = y_3$ // vectors of final label values
    13.4) `LabelPropagation`(f_1, f_2, f_3)
    13.5) update $F_3, F_{3,1}, F_{3,2}$
14. $F_{1,2}$ = mean ($F_{1,2}$, transpose($F_{2,1}$))
15. $F_{1,3}$ = mean ($F_{1,3}$, transpose($F_{3,1}$))
16. $F_{2,3}$ = mean ($F_{2,3}$, transpose($F_{3,2}$))
17. **return** $F_1, F_2, F_3, F_{1,2}, F_{1,3}, F_{2,3}$

`NormalizeSumOf`(S, W, W')

1. $d = \{0\}$ //a vector with S.numberOfRows length
2. **for** $i = 1 \ldots S.$***numberOfRows***
   2.1. **for** $j = 1 \ldots S.$***numberOfColumns***
      2.1.1. $M[i, j] = S[i, j] + W[i, j] + W'[i, j]$
      2.1.2. $d[i] = d[i] + M[i, j]$
   2.2. **if** $(d[i] = 0)$ $d[i] = 1$
3. **for** $i = 1 \ldots M.$***numberOfRows***
   3.1. **for** $j = 1 \ldots M.$***numberOfColumns***
      3.1.1. **if** $(i = j)$ $M[i, j] = 1$
      3.1.2. **else if** $(M[i,j]! = 0)$ $M[i, j] = \frac{M[i,j]}{\sqrt{d[i]d[j]}}$
4. **return** (M)

`LabelPropagation`(f_1, f_2, f_3)

1. **repeat** (steps 2–10)
   //drug
2. $f_1old = f_1$
3. $y'_1 = (1 - \alpha_1)f_1 + \alpha_1(S_{1,2} * f_2 + S_{1,3} * f_3)$ //f_1 is equal to y_1
4. $f_1 = (1 - \alpha_1)y'_1 + \alpha_1 * M_1 * f_1$
   //disease
5. $f_2old = f_2$
6. $y'_2 = (1 - \alpha_2)f_2 + \alpha_2((S_{1,2})^T * f_1 + S_{2,3} * f_3)$ //f_2 is equal to y_2
7. $f_2 = (1 - \alpha_2)y'_2 + \alpha_2 * M_2 * f_2$
   //target
8. $f3old = f3$
9. $y'_3 = (1 - \alpha_3)f_3 + \alpha_3((S_{1,3})^T * f_1 + (S_{2,3})^T * f_2)$ //f_3 is equal to y_3
10. $f_3 = (1 - \alpha_3)y'_3 + \alpha_3 * M_3 * f_3$
11. **while** $(||f1 - f_1old||>\sigma$ *or* $||f_2\text{-}f_2old||>\sigma$ *or* $||f_3\text{-}f_3old||>\sigma)$

In Algorithm 1, first, we initialize all labels to zero. During the projection phase, we project interactions onto similarity matrices by Eq. (2). We will have six projected matrices which will be integrated with corresponding similarity matrices in lines 8–10. In label propagation phase we have three iterative loops (lines 11–13). In each loop, we set one of the original labels to one and all others to zero. The label propagation function is applied, and output matrices are updated. The primary output consists of nine matrices; two of which correspond to drug-target interactions, two others correspond to drug-disease associations, and another two correspond to disease-target associations. Three remaining matrices present drug-drug, disease-disease and target-target relationships as separate matrices. We obtain six matrices from these nine matrices by combining matrices related to similar concepts (for example two drug-target interaction matrices are merged to produce one matrix and similarly for others). Final interaction prediction is achieved by sorting the rows of these matrices.

Algorithm 1 was implemented in C# using Visual Studio 2013 (the source code is available upon request from the corresponding author).

2.6. Key points in Heter-LP

The main idea of our projection phase is inspired by [58,59]. DT-Hybrid algorithm [20] which is one of the best approaches for prediction of drug-target interactions, is also a recommendation method based on projection. DT-Hybrid provides a similarity matrix for a set of vertices (like Y) by using the similarity between

other sets of vertices of the corresponding bipartite network (like X) across their relationships. However, here we focus on relationships between two sets and try to extract a topological similarity matrix by it. In this way we reduce the required computational costs in comparison to DT-Hybrid. Moreover, we use the similarity matrix of set X at label propagation phase. In this way, the similarity effects of vertices are not only through direct links but also we provide a kind of transitivity.

In label propagation phase, although there are some similarities with MINProp algorithm [57] there are several notable advantages. First, we use integrated data from primary input data and output of the projection phase as input for label propagation section. Second, there is one less iterative loop here which reduce the computational complexity of the algorithm impressively. It is noticeable that other heterogeneous label propagation algorithms like LPMIHN [30] also could not reduce the iterative loops in this way.

Moreover, Heter-LP has some more general advantages which include:

- No need to any inadvisable preprocessing of data. In other heterogeneous label propagation algorithms, it is assumed that there is coincidence between one homogenous set of vertices with equivalent vertices in heterogeneous networks. To provide it, they had to remove some informative data in the preprocessing phase. Heter-LP does not need to such a preprocessing because of no need to such a coincidence.
- No need to know the negative samples. In this concept, negative samples mean interactions which could not exist biologically. Most of other drug repositioning methods need to know negative samples and try to provide them randomly (there is no category for negative samples in this field).
- Heter-LP can predict interactions of *new* drugs (where a drug has no known target) and *new* targets (where a target has no known drug). This property is considered only in a few other methods.
- The other benefit of the proposed method is its ability to predict both trivial and non-trivial interactions. Trivial means the interactions which are predictable at first glance by everyone because of the existing similarities. Non-trivial interactions are the interactions that more evidence is required to find them.

3. Results and discussion

3.1. Convergence augment

An important part of the proposed method is the "`LabelPropagation`(f_1, f_2, f_3)" function. This function is based on an iterative algorithm whose convergence is here demonstrated. In fact, we will show that the sequences $\{f_1(t)\}$, $\{f_2(t)\}$, *and* $\{f_3(t)\}$ will ultimately converge and their corresponding answers are:

$$\begin{aligned} f_1^* &= (I-\alpha M_1)^{-1}[(1-\alpha)^2 y_1 + (1-\alpha)^3\alpha S_{1,2}y_2 + (1-\alpha)^3\alpha S_{1,3}y_3] \\ f_2^* &= (I-\alpha M_2)^{-1}[(1-\alpha)^2 y_2 + (1-\alpha)^3\alpha S_{2,1}y_2 + (1-\alpha)^3\alpha S_{2,3}y_3] \\ f_3^* &= (I-\alpha M_3)^{-1}[(1-\alpha)^2 y_3 + (1-\alpha)^3\alpha S_{3,1}y_1 + (1-\alpha)^3\alpha S_{3,2}y_2] \end{aligned} \tag{3}$$

Without loss of generality, we consider the same α for all subnetworks and rewrite $(1-\alpha)$ as β. So our first iterative equations will be as below:

$$\begin{aligned} f_1(t) &= \beta(\beta y_1 + \alpha S_{1,2}f_2(t-1) + \alpha S_{1,3}f_3(t-1) + \alpha M_1 f_1(t-1) \\ f_2(t) &= \beta(\beta y_2 + \alpha S_{2,1}f_1(t) + \alpha S_{2,3}f_3(t-1) + \alpha M_2 f_2(t-1) \\ f_3(t) &= \beta(\beta y_3 + \alpha S_{3,1}f_1(t) + \alpha S_{3,2}f_2(t) + \alpha M_3 f_3(t-1) \end{aligned} \tag{4}$$

By substitution of the above, we find:

$$\begin{aligned} f_1(t) &= \left[\beta^2\sum_{i=0}^{t-1}(\alpha M_1)^i y_1 + (\alpha M_1)^t y_1\right] \\ &\quad + \left[\beta^3\alpha\sum_{i=0}^{t-2}(\alpha M_1)^i S_{1,2}y_2 + \beta\alpha(\alpha M_1)^{t-1}S_{1,2}y_2\right] \\ &\quad + \left[\beta^3\alpha\sum_{i=0}^{t-2}(\alpha M_1)^i S_{1,3}y_3 + \beta\alpha(\alpha M_1)^{t-1}S_{1,3}y_3\right] + P_1 \\ f_2(t) &= \beta^2\sum_{i=0}^{t}(\alpha M_2)^i y_2 + \beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_2)^i S_{2,1}y_1 + \beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_2)^i S_{2,3}y_3 + P_2 \\ f_3(t) &= \beta^2\sum_{i=0}^{t}(\alpha M_3)^i y_3 + \beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_3)^i S_{3,1}y_1 + \beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_3)^i S_{3,2}y_2 + P_3 \end{aligned} \tag{5}$$

where each P_i is a summation of different t power of S_i and $S_{i,j}$ and they will converge to zero as iterations progress.

The final results will be achieved by $\lim_{t\to\infty} f_j(t)$ which are:

$$\begin{aligned} \mathrm{f}_1^* &= \lim_{t\to\infty} f_1(t) \\ &= \lim_{t\to\infty}\left[\beta^2\sum_{i=0}^{t-1}(\alpha M_1)^i y_1 + (\alpha M_1)^t y_1\right] \\ &\quad + \lim_{t\to\infty}\left[\beta^3\alpha\sum_{i=0}^{t-2}(\alpha M_1)^i S_{1,2}y_2 + \beta\alpha(\alpha M_1)^{t-1}S_{1,2}y_2\right] \\ &\quad + \lim_{t\to\infty}\left[\beta^3\alpha\sum_{i=0}^{t-2}(\alpha M_1)^i S_{1,3}y_3 + \beta\alpha(\alpha M_1)^{t-1}S_{1,3}y_3\right] + \lim_{t\to\infty} P_1 \\ &\cong \lim_{t\to\infty}[\beta^2\sum_{i=0}^{t}(\alpha M_1)^i y_1] + \lim_{t\to\infty}\left[\beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_1)^i S_{1,2}y_2\right] \\ &\quad + \lim_{t\to\infty}\left[\beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_1)^i S_{1,3}y_3\right] + \lim_{t\to\infty} P_1 \end{aligned} \tag{6}$$

We know that:

$$\begin{aligned} &\lim_{t\to\infty}[\beta^2\sum_{i=0}^{t}(\alpha M_1)^i y_1] = \beta^2(I-\alpha M_1)^{-1}y_1 \\ &\lim_{t\to\infty}[\beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_1)^i S_{1,2}y_2] = \beta^3\alpha(I-\alpha M_1)^{-1}S_{1,2}y_2 \\ &\lim_{t\to\infty}[\beta^3\alpha\sum_{i=0}^{t-1}(\alpha M_1)^i S_{1,3}y_3] = \beta^3\alpha(I-\alpha M_1)^{-1}S_{1,3}y_3 \\ &\lim_{t\to\infty} P_1 = 0 \end{aligned} \tag{7}$$

So the final equation for f_1^* will be:

$$\begin{aligned} f_1^* &= \lim_{t\to\infty} f_1(t) \\ &= (1-\alpha)^2(I-\alpha M_1)^{-1}y_1 + (1-\alpha)^3\alpha(I-\alpha M_1)^{-1}S_{1,2}y_2 \\ &\quad + (1-\alpha)^3\alpha(I-\alpha M_1)^{-1}S_{1,3}y_3 \\ &= (I-\alpha M_1)^{-1}[(1-\alpha)^2 y_1 + (1-\alpha)^3\alpha S_{1,2}y_2 \\ &\quad + (1-\alpha)^3\alpha S_{1,3}y_3] \end{aligned} \tag{8}$$

In the same way for f_2^* and f_3^*, we can write:

$$\begin{aligned} f_2^* &= \lim_{t\to\infty} f_2(t) = (1-\alpha)^2(I-\alpha M_2)^{-1}y_2 \\ &\quad + (1-\alpha)^3\alpha(I-\alpha M_2)^{-1}S_{2,1}y_1 + (1-\alpha)^3\alpha(I-\alpha M_2)^{-1}S_{2,3}y_3 \\ &= (I-\alpha M_2)^{-1}[(1-\alpha)^2 y_2 + (1-\alpha)^3\alpha S_{2,1}y_2 + (1-\alpha)^3\alpha S_{2,3}y_3] \\ f_3^* &= \lim_{t\to\infty} f_3(t) = (1-\alpha)^2(I-\alpha M_3)^{-1}y_3 + (1-\alpha)^3\alpha(I-\alpha M_3)^{-1}S_{3,1}y_1 \\ &\quad + (1-\alpha)^3\alpha(I-\alpha M_1)^{-1}S_{3,2}y_2 = (I-\alpha M_3)^{-1}[(1-\alpha)^2 y_3 \\ &\quad + (1-\alpha)^3\alpha S_{3,1}y_1 + (1-\alpha)^3\alpha S_{3,2}y_2] \end{aligned} \tag{9}$$

We express the resulting equation in closed-form as Eq. (10) below:

$$f_i^* = (I - \alpha M_i)^{-1}[(-\alpha)^2 y_i + (1-\alpha)^3 \alpha \sum_{j \neq i} S_{i,j} y_j] \text{ for } i,j = 1,2,3 \quad (10)$$

We have three homogeneous sub-networks. It can be easily verified that if we set $i, j = 1, 2, \ldots, k$, then Eq. (10) will be correct for a heterogeneous network with k different homogenous sub-networks.

Now we must choose a value for the constant parameter α. It seems that α cannot be arbitrarily large. If we let $\alpha \to 0$, then all final labels will be equal to the initial ones. If we increase α from zero, we will come to a point at $(I - \alpha M_i)^{-1}$ which will diverge and cause the divergence of f_i. This will happen when the determinant of $(I - \alpha M_i)$ passes through zero [60]. We can rewrite this condition as $det(M_i - \alpha^{-1} I) = 0$. This will happen when the roots of α^{-1} are equal to the eigenvalues of M_i. When α^{-1} becomes equal to the largest eigenvalue (k_i) of M_i, the determinant first crosses zero. Therefore, we must choose a value of α less than $1/k_i$.

The final labels can be calculated directly from Eq. (10). However, this would require the inversion of $(I - \alpha M_i)$. In order to invert this $n * n$ matrix, a time proportional to $O(n^3)$ is required by matrix multiplication algorithms, which can at best be reduced to $O(n^{2.373})$ through algorithmic optimization.[11] Therefore, we instead prefer to use the direct equations represented in Eq. (4). By repeating the process several times, the results will converge to the correct values.

We measure the required runtime for each step (projection and label propagation) separately. In the projection phase, we need to know the degrees of all the vertices of the interaction matrix. The required time is $O(2(|V_i||V_j|)$. The expected time to project one interaction matrix on one similarity matrix is $O(|V_i|^2|V_j|)$ which is repeated six times. So the total runtime for projection phase is $O(6(2(|V_i||V_j| + |V_i|^2|V_j|))$.

For label propagation on each homogeneous sub-network, the required time is equal to $O\left(t\left(|V_i| + \sum_{j \neq i} |V_i||V_j| + |V_i|^2\right)\right)$ and the total runtime of label propagation phase will be:

$$O\left(t \sum_{i=1}^{3} \left(|V_i| + \sum_{j \neq i} |V_i||V_j| + |V_i|^2\right)\right)$$

The total time for calculation is proportional to t, where t is the number of iterations required to reach convergence. The value of t depends on the input data structures and the parameter α. Therefore, we cannot determine the necessary time independently on the runtime parameters. However, in our experiments, α was always smaller than and equal to 0.3 and the value of t was always smaller than 10.

3.2. Regularization framework

Here we develop a regularization framework for the proposed label propagation algorithm. First, an objective function is determined. Then we will show that this function is strictly convex and will therefore have a globally optimal solution. We describe this solution and find it equal to the results of the previous section. In this way, it is proved that our proposed method will find an optimal solution.

The objective function: A cost function is defined to propagate a label in a heterogeneous network $G\ (V, E)$, as Eq. (11) below:

$$\Omega(f) = \sum_{i=1}^{3} \left(f_i^T \Delta^{(i)} f_i + \mu_i f_i - y_i^2 + \mu_i \sum_{i=1}^{2} \sum_{j=2}^{3} [f_i^T f_j^T] \sum^{(i,j)} \begin{bmatrix} f_i \\ f_j \end{bmatrix} \right) \quad (11)$$

where

- $f \in R^{|v|}$ is label vector, y_i is the initial label vector
- $\Delta^{(i)} = I - M_i$
- $\sum^{(i,j)} = I - \begin{bmatrix} 0 & S_{ij} \\ (S_{ij})^T & 0 \end{bmatrix}$ is the normalized graph laplacian of S^{ij}
- and ||.|| indicates a vector norm.
- $\Omega(f)$ consists of three cost terms:
- $f_i^T \Delta^{(i)} f_i$ is a smoothness term on the homogenous sub-network $G^{(i)} = (V^{(i)}, E^{(i)})$ which causes the similarity of labels of the connected nodes in the network $G^{(i)}$.
- $f_i - y_i^2$ is a fitting term which tends to keep the new label values near to the initial ones.
- $[f_i^T f_j^T] \sum^{(i,j)} \begin{bmatrix} f_i \\ f_j \end{bmatrix}$ is a smoothness term on the heterogeneous sub-network $G^{(i,j)} = (V^{(i)} U V^j, E^{(i,j)})$ which ensures the similarity of labels of connected nodes in the network $G^{(i,j)}$.

Proposition 1. *$\Omega(f)$ is strictly convex.*

Proof. All of Δ^i and $\sum^{(i,j)}$ terms are graph laplacian. This means that they are positive semi-definite and cause the convexity of the first and third cost terms of $\Omega(f)$. The second cost term is also convex. All of them are multiplied by some positive constants. So $\Omega(f)$ is a nonnegative-weighted sum of some convex function and is therefore itself a convex function [61]. On the other hand, the Hessian matrix of $\Omega(f)$ is a summation of Δ^i, $\sum^{(i,j)}$ and I. The Δ^i and $\sum^{(i,j)}$ terms are positive semi-definite and I is positive definite, so the Hessian matrix of $\Omega(f)$ is positive definite.[12] Therefore, $\Omega(f)$ is strictly convex. □

Proposition 2. *The optimal solution of $\Omega(f)$ is*:

$$f_i^* = (I - \alpha 1_i M_i)^{-1}[\alpha 2_i y_i + \alpha 3_i \sum_{j \neq i} S_{i,j} f_j] \quad (12)$$

Proof. As proven in Proposition 1, $\Omega(f)$ is strictly convex, so it can be solved by alternating optimization [61]. For each f_i, all f_j terms for $j \in \{n | n \in \mathbb{N}, 1 \leqslant n \leqslant 3, n \neq i\}$ are considered constant. We can then differentiate $\Omega(f)$ with respect to f_i and set it equal to zero to compute the solution:

$$\begin{aligned} &d\left(\frac{\Omega(f)}{f_i}\right) = 0 \\ &f_i^* - M_i f_i^* + 2\mu_i (f_i^* - y_i) + 2 f_i^* - \sum_{j \neq i} \mu_i S^{i,j} f_j = 0 \end{aligned} \quad (13)$$

[11] https://en.wikipedia.org/wiki/Computational_complexity_of_mathematical_operations.

[12] If $\nabla^2 f(x) \succ 0$ for all x∈ domain(f), then f is strictly convex 61 Boyd, S., and Vandenberghe, L.: 'Convex Optimization' (Cambridge University Press, 2004. 2004).

We will have:

$$((3+2\mu_i)I_i - M_i)f_i^* - 2\mu_i y_i - \sum_{j\neq i}\mu_i S^{i,j} f_j = 0$$
$$f_i^* = \left(I - \frac{M_i}{3+2\mu_i}\right)^{-1}\left(\frac{2\mu_i}{3+2\mu_i}y_i + \frac{\mu_i}{3+2\mu_i}\sum_{j\neq i}S^{i,j}f_j\right) \quad (14)$$

We will set:

$$\alpha 1_i = \frac{1}{3+2\mu_i}, \alpha 2_i = \frac{2\mu_i}{3+2\mu_i}, \alpha 3_i = \frac{\mu_i}{3+2\mu_i} \quad (15)$$

So the closed form of the solution will be:

$$f_i^* = (I_i - \alpha 1_i M_i)^{-1}[\alpha 2_i y_i + \alpha 3_i \sum_{j\neq i} S^{(i,j)} f_j] \quad \square \quad (16)$$

Proposition 3. *The proposed algorithm will minimize the objective function* $\Omega(f)$.

Proof. We should show that the result of Proposition 2 is equal with our answer in Eq. (10).

We could rewrite the iterative Eq. (3) as:

$$f_i(t) = (1-\alpha)^2 y_i + \alpha M_i f_i(t-1) + \sum_{j\neq i}(1-\alpha)^2 \alpha S^{i,j} f_j(t-1)$$
$$i,j = 1,2,3 \quad (17)$$

This is the equation solved by our proposed method and the result is:

$$f_i^* = (I - \alpha M_i)^{-1}\left[(1-\alpha)^2 y_i + (1-\alpha)^3 \alpha \sum_{j\neq i} S^{i,j} y_j\right] \quad (18)$$

If we set:

$$\alpha 1_i = \alpha, \alpha 2_i = (1-\alpha)^2, \alpha 3_i = (1-\alpha)^3\alpha$$

Both Eqs. (16) and (18) will be equal. □

So we have an optimization problem with the objective function represented by Eq. (11) which is strictly convex and, therefore, the proposed method can find its global minimum.

In next section, we report on the results of our experiments to evaluate the performance of the proposed algorithm using the datasets described in Section 2.

3.3. Statistical analysis

We considered three different scores as indicators for prediction accuracy:

1. Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve
2. Area Under the Precision-Recall (AUPR) curve
3. BestAccuracy

ROC is the plot of true positive rate (TPR) as a function of false positive rate (FPR) evaluated at various decision thresholds, where the true positives are correctly predicted true interactions and the false positives are predicted interactions not present in the gold standard set of interactions.

Precision is defined as the fraction of true drug targets identified among the candidate proteins ranked above the particular decision threshold. Recall is the fraction of true drug targets identified from among the total number of true drug targets in the gold standard set of interactions.

Although AUC represents the overall performance of the algorithm, previous studies have demonstrated that precision-recall curves more accurately assess a method's performance in the face of skewed node degree distributions in scale-free biological networks [6]. Precision-recall (PR) curves are also more informative when significant class imbalance exists. Furthermore, a curve dominates in ROC space if and only if it also dominates in PR space [33].

Accuracy measures the difference between a measurement with the actual value. It could be defined as Eq. (19):

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (19)$$

Here, TP = True Positives, FP = False Positives, TN = True Negatives, and FN = False Negatives.

The highest achieved accuracy repeated experiments with the same parameter values over the same data is defined as "BestAccuracy".

3.3.1. Results based on gold standard dataset

We performed 10-fold cross-validation (10-CV) to analyze the performance of the proposed Heter-LP method. In a 10-CV experiment, we split the dataset into ten subsets of equal size; each subset is then taken in turn as a test set, and the training is performed on the remaining nine sets.

It should be noted that the 10-CV testing is done on one set of interactions while keeping all others constant. We have six different datasets as input each time. We performed the experiment six times separately for each dataset. As an illustrative example, suppose that we want to evaluate the performance of the method on dataset E of the gold standard data. We have six matrices for dataset E as follows:

- Drug-drug similarity with 445 drugs.
- Target-target similarity with 664 targets.
- Disease-disease similarity with 53 diseases.
- Drug-target association with 295480 interactions.
- Drug-disease association with 23585 interactions.
- Disease-target association with 35192 interactions.

First, we applied 10-CV on the drug-drug similarity matrix by setting 445/10 entries of this matrix to zero randomly. All other entries of this matrix and all other input matrices were unchanged. The evaluation was run using these inputs and the corresponding performance metrics (AUC, AUPR, BestACC) were measured based on the changed entries. This process is iterated ten times for the drug-drug similarity matrix. Then the original drug-drug similarity matrix was used, and the same process was performed for another matrix from the six input matrices. The average values of calculated measures were used for evaluation of the method for dataset E. A similar process was performed for each of the four individual datasets of the gold standard data (E, GPCR, IC and NR). All processes were done programmatically using custom C# software.

Table 1 shows the results of the proposed method during 10-CV over the gold standard datasets of Yamanishi 2008 [22] with augmented information as described in Section 2.2.1. One can observe that the results of drug-target prediction are the strongest. This indicates that Heter-LP can predict drug-target interactions more accurately than the other two interactions. We repeated this test using some other methods and observed the same trend. This result appears to be due to incompleteness of the input data. As explained before, the available golden standard dataset doesn't contain any information about diseases and our attempt for the addition of such information was insufficient to create a suitable complete dataset. In other words, the disease similarity matrix and the primitive interaction matrices used here are not sufficiently informative.

Table 1
Results of Heter-LP method during 10-CV testing on the gold standard dataset.

Interaction	Dataset: E			Dataset: GPCR			Dataset: IC			Dataset: NR		
	AUC	AUPR	BestAcc	AUC	AUPR	BestAcc	AUC	AUPR	BestAcc	AUC	AUPR	BestAcc
Drug-Disease	0.8292	0.8475	0.7231	0.8606	0.8697	0.7468	0.8400	0.8539	0.7340	0.8640	0.8625	0.7582
Drug-Target	0.9918	0.7967	0.9917	0.9928	0.8575	0.9873	0.9878	0.7684	0.9856	0.9823	0.8965	0.9789
Disease-Target	0.9147	0.8020	0.9449	0.8529	0.7381	0.8850	0.9163	0.8096	0.9454	0.9383	0.9459	0.9444

Table 2
Results of 10-fold cross validation on E, GPCR, IC and NR datasets.

Dataset	Method	Drug-disease			Drug-target			Disease-target		
		AUC	AUPR	BestACC	AUC	AUPR	BestACC	AUC	AUPR	BestACC
Dataset E	DT-Hybrid (R)	0.716	0.736	0.695	0.947	**0.824**	**0.997**	0.637	0.641	0.612
	Heter-LP	**0.771**	**0.814**	**0.705**	**0.953**	0.715	0.991	**0.850**	**0.756**	**0.886**
	MINProp	0.478	0.475	0.504	0.511	0.012	0.990	0.520	0.512	0.522
Dataset GPCR	DT-Hybrid (R)	0.696	0.713	0.667	0.937	**0.829**	**0.988**	0.669	0.683	0.637
	Heter-LP	**0.793**	**0.832**	**0.723**	**0.967**	0.796	0.986	**0.796**	**0.698**	**0.840**
	MINProp	0.500	0.248	0.504	0.498	0.024	0.970	0.501	0.533	0.503
Dataset IC	DT-Hybrid (R)	0.674	0.707	0.642	0.923	**0.764**	0.983	0.648	0.655	0.628
	Heter-LP	**0.778**	**0.818**	**0.712**	**0.948**	0.711	**0.984**	**0.857**	**0.771**	**0.899**
	MINProp	0.498	0.249	0.500	0.499	0.017	0.965	0.501	0.249	0.500
Dataset NR	DT-Hybrid (R)	0.689	0.699	0.661	**0.948**	**0.903**	**0.988**	0.757	0.772	0.726
	Heter-LP	**0.805**	**0.829**	**0.735**	**0.948**	0.838	0.975	**0.881**	**0.907**	**0.898**
	MINProp	0.499	0.247	0.506	0.523	0.068	0.936	0.493	0.473	0.514

3.3.2. Comparison with state of the art methods on gold standard data

To provide the opportunity to compare the performance of different methods, two-column charts are represented in Figs. 3 and 4. Represented performances are self-reported by the corresponding papers (more explanation is in supplementary material). The corresponding numerical values are available in supplementary materials. We found that "DT-Hybrid" performs best. So we downloaded this package and evaluated it for two reasons. Firstly, to

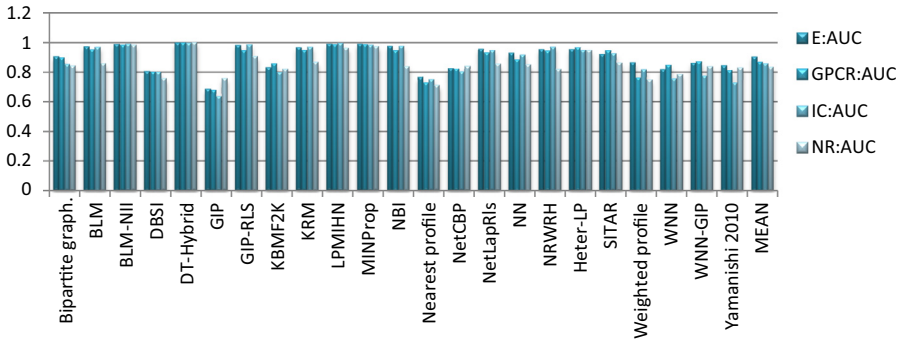


Fig. 3. Self-reported AUC of various methods in gold standard datasets.

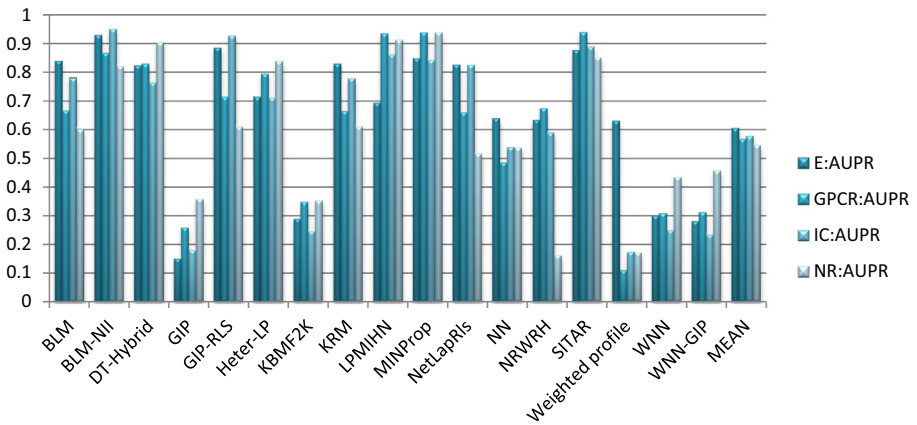


Fig. 4. Self-reported AUPR of some different methods in gold standard datasets.

analyze two other types of interactions mentioned before (disease-target, drug-disease) and secondly, to find the accurate performance of this package with 10-CV.

The mean values of the AUC and AUPR are provided in last columns of Figs. 3 and 4, respectively. Both of the AUC and AUPR values of the Heter-LP algorithm in drug-target prediction are higher than the mean values of the represented methods. One can observe that the comparisons in Figs. 3 and 4 are based only on drug-target prediction and no other relations. Although, as mentioned before, the disease relations are also important for perfect repositioning of drugs; unfortunately, no comparison method is available for evaluating the presented methods from this aspect.

3.3.3. Comparison of 10-fold CV results

The three methods (DT-Hybrid, Heter-LP, and MINProp) were compared across three types of predictions (Drug-disease, Drug-target, and Disease-target) using three different performance metrics (AUC, AUPR, BestACC). These results are presented in Table 2 for the four datasets: E, GPCR, IC and NR. Heter-LP was the top performing method in 29 of the 36 evaluations (see results in bold). This provides strong evidence that Heter-LP outperforms both DT-Hybrid and MINProp, particularly for Drug-disease and Disease-target predictions, where Heter-LP dominated.

To facilitate the comparison, we provide Table 3 that demonstrates the average of AUC, AUPR and BestACC of each method across the four gold standard datasets (E, GPCR, IC and NR). Table 3 shows that Heter-LP outperforms DT-Hybrid and MINProp in all cases but two. A "$*$" appears in cases where the best method for that case is deemed to be significantly ($p < 0.05$) superior when compared with both other methods according to Table 4.

We also performed a paired sample *t*-test on the results to provide an informative comparison between the three methods represented in Table 2. In this regard, we used the "*t*.test" function of R and the p-value are presented in Tables 4 and 5. According to [62], the paired sample *t*-test is a robust and reliable method to compare the accuracies of different approaches statistically, particularly when one expects the performance of two methods to be correlated for different samples. As is commonly done, we here assume that if the p-value is lower than or equal to 0.05, the differences of two paired samples are statistically significant. In Table 4 comparison is performed between any two pairs of methods over the four datasets (E, GPCR, IC and NR). In all cases the accuracy parameters of "DT-Hybrid and MINPROP" and "Heter-LP and MINPROP" are significantly different. Differences of "DT-Hybrid and Heter-LP" are not significant in some cases, so we performed a more detailed comparison of these two methods (Table 5). A *t*-test is applied to the 10-CV results for two methods (for each dataset, for each performance metric, and for each prediction type). Note that since the 10 folds in our 10-CV were selected identically for all methods (i.e. that the same targets/drugs/diseases were used in each fold when evaluating each method) then this is a paired sample *t*-test. Table 5 shows that in all cases where Heter-LP outperforms DT-Hybrid (according to Table 2), the differences are significant. Only in two cases does DT-Hybrid significantly outperform Heter-LP: AUPR and BestACC of drug-target prediction on datasets E and GPCR.

3.3.4. Prediction of a deleted interaction

A standard approach for validating a method is to remove some of the desirable entries from the input, then to run the process to recover the missing results (as one does during a cross-validation experiment). In this regard, we performed two distinct examinations. We chose an arbitrary drug (or target) and, as Test1, we deleted one of its interactions from the input dataset and, as Test2,

Table 3
Average of AUC and AUPR for DT-Hybrid and Heter-LP on gold standard datasets.

Interaction	Method	AUC average	AUPR average	BestACC average
Drug-disease	DT-Hybrid	0.6935	0.7139	0.6663
	Heter-LP	**0.7865***	**0.8234***	**0.7188***
	MINProp	0.4938	0.3048	0.5035
Drug-target	DT-Hybrid	0.9389	**0.8304***	**0.9890**
	Heter-LP	**0.9541**	0.6901	0.9840
	MINProp	0.5018	0.0303	0.9653
Disease-target	DT-Hybrid	0.6778	0.6878	0.6508
	Heter-LP	**0.8460***	**0.7827***	**0.8808***
	MINProp	0.5038	0.4418	0.5098

Table 4
P-value comparison of DT-Hybrid, Heter-LP and MINPROP based on four datasets for each interaction.

Comparing methods	Drug-disease			Drug-target			Disease-target		
	AUC	AUPR	BestACC	AUC	AUPR	BestACC	AUC	AUPR	BestACC
DT-Hybrid (R)& Heter-LP	5.95e−3	2.27e−3	3.74e−2	1.26e−1	2.73e−2	1.97e−1	6.47e−3	3.92e−2	2.83e−3
DT-Hybrid (R)& MINPROP	6.44e−4	3.69e−3	5.61e−4	1.52e−6	2.82e−05	9.32e−2	1.20e−2	3.29e−2	1.18e−2
Heter-LP& MINPROP	1.31e−5	3.25e−3	4.51e−5	1.87e−5	5.14e−5	9.59e−2	4.17e−4	2.57e−2	1.04e−4

Table 5
P-value for DT-Hybrid and Heter-LP based on 10-fold CV.

Dataset	Drug-disease			Drug-target			Disease−target		
	AUC	AUPR	BestACC	AUC	AUPR	BestACC	AUC	AUPR	BestACC
Dataset E	3.11e−11	3.87e−09	2.17e−3	1.16e−05	5.56e−12	2.09e−12	2.2e−16	2.2e−16	2.2e−16
Dataset GPCR	2.94e−13	1.24e−10	9.97e−09	2.21e−06	2.23e−3	2.77e−3	5.75e−15	1.83e−09	1.05e−14
Dataset IC	2.13e−15	1.41e−14	6.80e−10	9.30e−05	7.05e−1	2.54e−05	2.2e−16	2.2e−16	2.32e−16
Dataset NR	5.13e−16	6.81e−13	1.04e−10	6.18e−4	4.78e−1	3.29e−05	4.11e−09	7.33e−12	2.92e−11

Table 6
Top 20 predicted targets of Drug: D00232 by proposed method.

No.	Heter-LP			DT-Hybrid		
	Target ID	Official Symbol	Official Full Name	Target ID	Official Symbol	Official Full Name
1	**hsa:1128**	CHRM1	cholinergic receptor muscarinic 1	**hsa:1128**	CHRM1	cholinergic receptor muscarinic 1
2	**hsa:1131**	CHRM3	cholinergic receptor muscarinic 3	**hsa:1131**	CHRM3	cholinergic receptor muscarinic 3
3	**hsa:1129**	CHRM2	cholinergic receptor muscarinic 2	**hsa:1129**	CHRM2	cholinergic receptor muscarinic 2
4	hsa:11255	HRH3	histamine receptor H3	hsa:1132	CHRM4	cholinergic receptor muscarinic 4
5	hsa:154	ADRB2	adrenoceptor beta 2	hsa:3269	HRH1	histamine receptor H1
6	hsa:3269	HRH1	histamine receptor H1	hsa:1133	CHRM5	cholinergic receptor muscarinic 5
7	hsa:153	ADRB1	adrenoceptor beta 1	hsa:3360	HTR4	5-hydroxytryptamine receptor 4
8	hsa:1813	DRD2	dopamine receptor D2	hsa:8843	HCAR3	hydroxycarboxylic acid receptor 3
9	hsa:148	ADRA1A	adrenoceptor alpha 1A	hsa:1813	DRD2	dopamine receptor D2
10	hsa:4988	OPRM1	opioid receptor mu 1	hsa:3356	HTR2A	5-hydroxytryptamine receptor 2A
11	hsa:185	AGTR1	angiotensin II receptor type 1	hsa:148	ADRA1A	adrenoceptor alpha 1A
12	hsa:150	ADRA2A	adrenoceptor alpha 2A	hsa:3358	HTR2C	5-hydroxytryptamine receptor 2C
13	hsa:3577	CXCR1	C-X-C motif chemokine receptor 1	hsa:1812	DRD1	dopamine receptor D1
14	hsa:3274	HRH2	histamine receptor H2	hsa:1815	DRD4	dopamine receptor D4
15	hsa:152	ADRA2C	adrenoceptor alpha 2C	hsa:146	ADRA1D	adrenoceptor alpha 1D
16	hsa:147	ADRA1B	adrenoceptor alpha 1B	hsa:147	ADRA1B	adrenoceptor alpha 1B
17	hsa:3360	HTR4	5-hydroxytryptamine receptor 4	hsa:59340	HRH4	histamine receptor H4
18	hsa:146	ADRA1D	adrenoceptor alpha 1D	hsa:11255	HRH3	histamine receptor H3
19	hsa:1814	DRD3	dopamine receptor D3	hsa:150	ADRA2A	adrenoceptor alpha 2A
20	hsa:155	ADRB3	adrenoceptor beta 3	hsa:151	ADRA2B	adrenoceptor alpha 2B

we deleted all of its interactions with targets (or drugs) from the dataset. Test1 is explained here and Test2 will be described in the next section.

By deleting a single interaction, we investigate the ability of the method in predicting new interactions for known drugs or targets (i.e. nodes that have some known interaction with others). We perform different analysis tasks in this regard, and here we illustrate a case study. D00232 is a drug with three interactions with targets: hsa:1128, hsa:1129 and hsa:1131. We deleted its interaction with hsa:1129 from the input network and investigated the results. Both Heter-LP and DT-Hybrid correctly predict this removed interaction, as can be seen in Table 6 (3rd highest ranked prediction). This list also contains several high-scoring novel drug-target predictions that merit experimental validation.

3.3.5. Prediction of pseudo-new drugs

As mentioned above, our method can predict the interactions between new targets and new drugs correctly. This feature offers a great advantage that most of the existing methods do not provide. In this regard, we perform Test2 to create a pseudo-new drug and compare the results of our proposed method with DT-Hybrid.

In Test2, we left out all interactions of a given drug (or given target) each time by setting to zero the corresponding entries of drug-target association matrix. In this way, we create a pseudo-new drug (or target) with no known interactions in the input matrices. None of the other entries was changed. At the next step, each method was run using these simulated inputs and the results were investigated especially for the pseudo-new drug (or target).

In the previous section, we explained that drug D00232 has three targets in the gold standard drug-target interaction. We first deleted all of these interactions. In this way, D00232 is like a new drug in that it no longer has any known interactions with any target in the input network. Both Heter-LP and DT-Hybrid were then applied to the censored dataset. The top 20 targets predicted by Heter-LP are presented in Table 7. All of the desired targets (which we deleted from the input data) are predicted successfully by our method (see bold entries in Table 7). Recall that these were recovered from among 989 possible targets.

Table 7
Top 20 Heter-LP predicted targets of Drug: D00232.

No.	Heter-LP			DT-Hybrid
	Predicted target ID	Official Symbol	Official Full Name	–
1	hsa:154	ADRB2	adrenoceptor beta 2	–
2	hsa:3269	HRH1	histamine receptor H1	–
3	hsa:153	ADRB1	adrenoceptor beta 1	–
4	**hsa:1128**	CHRM1	cholinergic receptor muscarinic 1	–
5	hsa:1813	DRD2	dopamine receptor D2	–
6	hsa:148	ADRA1A	adrenoceptor alpha 1A	–
7	hsa:4988	OPRM1	opioid receptor mu 1	–
8	hsa:185	AGTR1	angiotensin II receptor type 1	–
9	**hsa:1129**	CHRM2	cholinergic receptor muscarinic 2	–
10	hsa:3577	CXCR1	C-X-C motif chemokine receptor 1	–
11	hsa:150	ADRA2A	adrenoceptor alpha 2A	–
12	hsa:3274	HRH2	histamine receptor H2	–
13	hsa:1814	DRD3	dopamine receptor D3	–
14	hsa:146	ADRA1D	adrenoceptor alpha 1D	–
15	hsa:3360	HTR4	5-hydroxytryptamine receptor 4	–
16	**hsa:1131**	CHRM3	cholinergic receptor muscarinic 3	–
17	hsa:3356	HTR2A	5-hydroxytryptamine receptor 2A	–
18	hsa:147	ADRA1B	adrenoceptor alpha 1B	–
19	hsa:155	ADRB3	adrenoceptor beta 3	–
20	hsa:151	ADRA2B	adrenoceptor alpha 2B	

Table 8
Statistics of Test2 for 30 pseudo-new drugs achieved by Heter-LP. # K is the number of known targets of given drug, # P is the number of correct predicted targets by Heter-LP.

No.	Drug ID	# K	# P	No.	Drug ID	# K	# P	No.	Drug ID	# K	# P
1	D00066	2	2	11	D00300	3	1	21	D00674	1	0
2	D00110	3	1	12	D00326	4	2	22	D00760	1	1
3	D00113	5	3	13	D00336	3	1	23	D00779	2	2
4	D00182	1	1	14	D00426	4	4	24	D00930	1	1
5	D00195	9	1	15	D00437	9	9	25	D01712	5	0
6	D00228	2	1	16	D00494	11	7	26	D01973	1	1
7	D00274	3	3	17	D00537	3	2	27	D02756	6	2
8	D00279	2	0	18	D00546	4	1	28	D03365	8	3
9	D00283	5	3	19	D00577	5	1	29	D03621	4	4
10	D00293	4	2	20	D00585	4	3	30	D05353	4	3
Sum of # P/Sum of # K						65/119					

As you can see in Table 7, DT-Hybrid could not predict any target for this pseudo-new drug (all of the entries of D00232 in its output were zero). We repeated this experiment several times for D00232 as well as other drugs and targets. The results were similar (all zero) in all cases. In [62], the authors of DT-Hybrid also declared that it could not predict the targets of new drugs and they suggest possible workarounds for this problem.

Similar examinations were conducted for some other drugs (like D05353, D00227) and for a variety of targets, creating pseudo-new targets by removing all known interactions with drugs. Results were largely consistent with those of Test2, but results are excluded due to space limitation.

We selected 30 different drugs randomly from our dataset and performed Test2 for each drug separately. Table 8 provides the results of these analyses. The results demonstrate the strength of Heter-LP in predicting the interactions for new drugs.

3.4. Experimental analysis

Statistical analysis confirms the ability of the proposed method in predicting potential interactions, now is the time to investigate its practical effectiveness. In this regard, we used all the data as a training set and examined new predicted interactions. Novel interactions were ranked by score and the top 20 predictions were extracted. These novel interactions were checked manually using the online version of DrugBank,[13] Supertarget,[14] KEGG Drug[15] and Therapeutic Target Database (TTD).[16].

This test is performed two times, one based on the gold standard dataset and the other one using the introduced independent datasets (Section 2.2.2). Because of space limitations, the full predicted lists are placed in supplementary materials.

We categorized the novel predictions in two groups, trivial and non-trivial ones. Trivial predictions could be predicted by straightforward and primary investigations of the input data. Non-trivial could not be quickly discovered from input data. It seems that an effective method should be able to identify both types sufficiently. Some examples are represented in the following sections.

3.4.1. Experimental analysis based on gold standard dataset

We sorted the predicted list of unknown drug-target interactions of each group (E, GPCR, IC, NR) and extracted the top 20 ones of each group separately. Because of space limitations, only the results of GPCR are represented here in Table 9; others are available in supplementary materials. A similar investigation was done by some of the other methods like BLM [28], KBMF2 k [21], LapRLS [31], LPMIHN [30], NetCBP [26], NRWRH [45], RLS-Kron[17] [28], WNN [29]. Table 7 of supplementary materials is a brief comparison of the results of experimental analysis of different methods.

Verified predictions in Table 9 are denoted by the name of the related source. As you can see, seven of 20 GPCR interactions are verified. Furthermore, a number of the non-validated predictions have additional supporting biological evidence that are beyond the scope of this paper. One non-trivial example prediction is discussed in the following case study.

Although we also had predicted drug-disease and disease-target interactions, we do not discuss them here since none of the compared methods are capable of making such predictions. Instead, these interactions will be discussed below in the experimental analysis based on independent datasets.

On the other hand, we claim that our proposed method could predict trivial and non-trivial interactions. We will explain two case studies to demonstrate this claim.

A trivial case study: (D02358 & hsa:154)

D02358 is the KEGG id of drug Metoprolol (USAN/INN) and hsa:154 is the KEGG id of protein target Beta-2 adrenergic receptor which is also known by its UniProt name ADRB2_HUMAN. No interaction is defined between D02358 and hsa:154 when using our gold standard dataset as input (the corresponding entry in $G_{1,3}$ is zero).

We searched the SuperTarget website in August 2016 and found hsa:154 to be a target of D02358. We therefore consider this to be a trivial prediction since it could have been predicted via simple research. According to the input similarity matrix of protein targets, hsa:154 is the most similar target to hsa:153. However, according to our input interaction matrix, hsa:153 is the sole target of D02358. It is reasonable to introduce the pair (D02358 & hsa:154) as one the most probable candidates (row 2 of Table 9) and new experimental research (listed in SuperTarget) indeed verifies this interaction.

A non-trivial case study: (D00673 & hsa:3269)

D00673 is the KEGG id of drug Ranitidine hydrochloride and hsa:3269 is the KEGG id of protein target Histamine H1 receptor (UniProt name HRH1_HUMAN). No interaction is defined between D00673 and hsa:3269 in our gold standard input dataset.

We examined the input matrices in two ways to establish that this predicted interaction is non-trivial. First, we found the interacted targets of D00673 from the drug-target interaction input matrix then investigated their similar targets using the target similarity input matrix. Second, we found the interacted drugs of hsa:3269 from the drug-target interaction input matrix and identified similar drugs using the drug-drug similarity input matrix.

The only target of D00673 is hsa:3274. The most similar target to hsa:3274 is hsa:3360 and between 95 distinct targets, hsa:3269 ranks 37th in similarity to hsa:3274. Clearly, the predicted interaction (D00673 & hsa:3269) could not be predicted solely through target similarity.

We then found the drugs predicted to interact with hsa:3269 using the drug-target interaction input matrix. Its related drugs are:

[13] http://www.drugbank.ca/.
[14] http://insilico.charite.de/supertarget/.
[15] http://www.genome.jp/kegg/drug/.
[16] http://bidd.nus.edu.sg/group/cjttd/.
[17] Regularized Least Squares (RLS) with Kronecker sum kernel.

Table 9
Top 20 new predicted interactions in GPCR dataset.

No.	Pair		Annotation			Verification source
	Drug	Target	Drug	Target	UniprotName of target	
1	D00542	hsa:338442	Halothane (JP17/USP/INN)	G-protein coupled receptor 109A	NIAR1_HUMAN	
2	D02358	hsa:154	Metoprolol (USAN/INN)	Beta-2 adrenergic receptor	ADRB2_HUMAN	SuperTarget
3	D04625	hsa:154	Isoetharine (USP) Isoetarine (INN)	Beta-2 adrenergic receptor	ADRB2_HUMAN	KEGG
4	D02614	hsa:154	Denopamine (JAN/INN)	Beta-2 adrenergic receptor	ADRB2_HUMAN	
5	D02147	hsa:153	Albuterol (USP) Salbutamol	Beta-1 adrenergic receptor	ADRB1_HUMAN	SuperTarget
6	D02359	hsa:153	Ritodrine (USAN/INN)	Beta-1 adrenergic receptor	ADRB1_HUMAN	
7	D00683	hsa:153	Albuterol sulfate (USP) Salbutamol sulfate (JP17)	Beta-1 adrenergic receptor	ADRB1_HUMAN	SuperTarget
8	D05792	hsa:153	Salmeterol (USAN/INN)	Beta-1 adrenergic receptor	ADRB1_HUMAN	
9	D00688	hsa:153	Terbutaline sulfate (JP17/USP)	Beta-1 adrenergic receptor	ADRB1_HUMAN	SuperTarget
10	D00684	hsa:153	Bitolterol mesylate (USAN) Bitolterol mesilate (JAN)	Beta-1 adrenergic receptor	ADRB1_HUMAN	
11	D01386	hsa:153	Ephedrine hydrochloride (JP17/USP)	Beta-1 adrenergic receptor	ADRB1_HUMAN	KEGG
12	D00687	hsa:153	Salmeterol xinafoate (JAN/USAN)	Beta-1 adrenergic receptor	ADRB1_HUMAN	SuperTarget
13	D00673	hsa:3269	Ranitidine hydrochloride (JP17/USP)	Histamine H1 receptor	HRH1_HUMAN	
14	D03503	hsa:3269	Cimetidine hydrochloride (USP)	Histamine H1 receptor	HRH1_HUMAN	
15	D00422	hsa:3269	Ranitidine (USAN/INN)	Histamine H1 receptor	HRH1_HUMAN	
16	D00440	hsa:3269	Nizatidine (JP17/USP/INN)	Histamine H1 receptor	HRH1_HUMAN	
17	D00295	hsa:3269	Cimetidine (JP17/USP/INN)	Histamine H1 receptor	HRH1_HUMAN	
18	D00765	hsa:1128	Rocuronium bromide (JAN/USAN/INN)	Muscarinic acetylcholine receptor M1	ACM1_HUMAN	
19	D01346	hsa:3269	Bentiromide (JAN/USAN/INN)	Histamine H1 receptor	HRH1_HUMAN	
20	D00760	hsa:1128	Doxacurium chloride (USAN/INN)	Rocuronium bromide (JAN/USAN/INN)	Muscarinic acetylcholine receptor M1	

D00234, D00283, D00300, D00364, D00454, D00480, D00493, D00494, D00520, D00521, D00665, D00666, D01242, D01295, D01324, D01332, D01713, D01717, D01782, D02327, D02354, D02361, D02566, D03621, D04979, D05129.

The most similar drug to D00673 is D00422, and from above mentioned drugs, the most similar one is D00480 which ranks 30th in terms of similarity with D00673. Clearly, the D00673 & hsa:3269 interaction could not have bene predicted based on drug similarity alone.

Now we will show that this prediction is plausible and should be considered as a good candidate for experimental validation. As verified by SuperTarget, D00673 now has 14 known targets, two of which are HRH2_HUMAN and HRH4_HUMAN. Our predicted target, HRH1_HUMAN, is highly similar to HRH4_HUMAN (SuperTarget data and DrugBank documentation clarify their similar aspects).

3.4.2. Experimental analysis based on independent datasets

The gold standard data had value in that it enabled us to compare our proposed method with a wide variety of methods evaluated using the same data. However, the gold standard data is somewhat obsolete (2008) and incomplete (e.g. lacking drug-disease interaction data). We, therefore, created an updated and complete dataset to fully evaluate the capabilities of our proposed method, Heter-LP. These data were used as input of the proposed method and their results were analyzed. We discuss one of its interesting predictions as a case study.

Case study: Osteoarthritis with mild chondrodysplasia

Osteoarthritis with mild chondrodysplasia is a type of skeletal disease due to the mutation of type II procollagen (COL2A1). It causes a progressive degeneration of the articular cartilage of joints with mild spinal chondrodysplasia.[18,19].

Supplementary materials presents its associated drugs and their corresponding targets (which we used as input data). The only known target of this disease is "hsa:1280" for which there is no known drug. The most similar diseases (similarity higher than 0.3) to "Osteoarthritis with mild chondrodysplasia" and their associated drugs and targets are listed in supplementary materials.

We have predicted two new drugs for the treatment of this disease: Alendronate sodium and Alendronic acid. Alendronate sodium is a salted form of Alendronic acid and, as expressed in DrugBank, is "for the treatment and prevention of osteoporosis in women and Paget's disease of bone in both men and women". Its treatment effects have two sides, one by its affinity for hydroxyapatite and the other, its inhibiting effect on FPP[20] synthase. Hydroxyapatite is part of the mineral matrix of bone and inhibition of FPP will inhibit osteoclast activity and reduce bone resorption.[21]

In our input datasets both of Alendronate sodium and Alendronic acid are assigned for the treatment of Osteoporosis-pseudoglioma syndrome (OPPG), which is a skeletal disease "characterized by severe congenital osteoporosis with blindness".[22]

We assert that this prediction is plausible and merits further experimental validation because "OPPG" and "Osteoarthritis with mild chondrodysplasia" share the same category and the mechanism of action of Alendronic acid and Alendronate sodium. It is necessary to mention that this prediction is not a trivial one. As you can see in supplementary materials "OPPG" and "Osteoarthritis with mild chondrodysplasia" are not similar diseases in input dataset. Also, their input drugs and their input targets are not the same. The only target for "OPPG" in input datasets is hsa:4041 and, as mentioned before, the only known target of "Osteoarthritis with mild chondrodysplasia" is hsa:1280. To establish that this prediction could not have been made trivially on the basis of drug similarity, we provide a list of most similar drugs to "Alendronate" from input dataset in supplementary materials. No target similarity comparison is represented here because of the non-existence of hsa:4041 in input target similarity matrix.

[18] "Chondrodysplasia is a heterogeneous group of bone dysplasias, the common characteristic of which is stippling of the epiphyses in infancy." http://medical-dictionary.thefreedictionary.com/chondrodysplasia.
[19] http://www.kegg.jp/kegg/disease/.
[20] Farnesyl pyrophosphate.
[21] http://www.drugbank.ca/drugs/DB00630.
[22] http://www.kegg.jp/kegg/disease/.

4. Conclusion

Label propagation is an efficient technique to utilize both local and global features in a network for semi-supervised learning [57]. In spite of the growing interest in the use of heterogeneous networks in various scientific disciplines, there is insufficient attention being paid to label propagation on these networks. In this paper, we introduce a new label propagation algorithm on heterogeneous networks named Heter-LP. Its convergence to an optimal solution is discussed and proved. We have shown that there are fewer iteration loops in Heter-LP in comparison to other heterogeneous label propagation algorithms and the time complexity is acceptable.

The Heter-LP algorithm was applied to the problem of drug repositioning to demonstrate its applicability. It was shown that Heter-LP can infer new interactions for disease-drug, drug-target, and disease-target relationships successfully through integrating heterogeneous information obtained from various types of resources at different levels of biological detail. In fact, we used both local and global features together by using label propagation. Furthermore, an advantage of the proposed model is that it does not require negative interactions for training, as experimental analysis rarely reports negative samples.

We provide a comprehensive statistical analysis of performance by using of 10-fold cross validation testing. The achieved AUC and AUPR outperform most existing state-of-the-art methods for drug repositioning. Although these parameters are weaker than some methods in some sense, our experimental analysis has demonstrated some attractive abilities. It is shown that Heter-LP can predict interactions of new drugs, targets and diseases correctly. Moreover, in spite of some methods that can predict only trivial interactions, Heter-LP can predict both trivial and non-trivial ones.

In total, the analysis demonstrates that label propagation is an effective algorithm to predict the new drug-target-disease interactions. The possible combination of this approach with network attributes, such as topological ones (e.g. different centrality measures) or the addition of more types of data, such as anatomical therapeutic chemical (ATC) codes of drugs, and continuous sequence similarities of proteins. Lastly, we are confident that performance will continue to increase as more accurate and complete input data become available.

Conflicts of interest

The authors have no conflicts of interest to declare.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2017.03.006.

References

[1] P. Zhang, P. Agarwal, Z. Obradovic, Computational drug repositioning by ranking and integrating multiple data sources, in: H. Blockeel, K. Kersting, S. Nijssen, F. Železný (Eds.), Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23–27, 2013, Proceedings, Part III, Springer, Berlin, Heidelberg, 2013, pp. 579–594.
[2] F. Emmert-Streib, S. Tripathi, R.d.M. Simoes, A.F. Hawwa, M. Dehmer, The human disease network, Syst. Biomed. 1 (1) (2013) 20–28.
[3] T.C. Silva, L. Zhao, Machine Learning in Complex Networks, Springer, 2016.
[4] Y.-F. Dai, X.-M. Zhao, A survey on the computational approaches to identify drug targets in the postgenomic era, Biomed. Res. Int. 2015 (2015) 9.
[5] D. Emig, A. Ivliev, O. Pustovalova, L. Lancashire, S. Bureeva, Y. Nikolsky, M. Bessarabova, Drug target prediction and repositioning using an integrated network-based approach, PLoS ONE 8 (4) (2013) e60618.
[6] S.-H. Yeh, H.-Y. Yeh, V.-W. Soo, A network flow approach to predict drug targets from microarray data, disease genes and interactome network – case study on prostate cancer, J. Clin. Bioinform. 2 (2012) 1.
[7] J. Setoain, M. Franch, M. Martinez, D. Tabas-Madrid, C.O. Sorzano, A. Bakker, E. Gonzalez-Couto, J. Elvira, A. Pascual-Montano, NFFinder: an online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning, Nucleic Acids Res. 43 (W1) (2015) W193–W199.
[8] R. Chang, R. Shoemaker, W. Wang, A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention, IEEE/ACM Trans. Comput. Biol. Bioinformat. 8 (5) (2011) 1170–1182 (IEEE, ACM).
[9] S. Imoto, Y. Tamada, C.J. Savoie, S. Miyano, Analysis of gene networks for drug target discovery and validation, Methods Mol. Biol. (Clifton, NJ) 360 (2007) 33–56.
[10] A.S. Brown, S.W. Kong, I.S. Kohane, C.J. Patel, ksRepo: a generalized platform for computational drug repositioning, BMC Bioinformatics 17 (1) (2016) 78.
[11] H.-R. Chen, D.H. Sherr, Z. Hu, C. DeLisi, A network based approach to drug repositioning identifies plausible candidates for breast cancer and prostate cancer, BMC Med. Genomics 9 (1) (2016) 1–11.
[12] J. Zhang, J. Huan, Analysis of network topological features for identifying potential drug targets, in: Proc 9th Intl Workshop Data Mining Bioinformatics (BIOKDD 2010), 2010.
[13] M. Koyuturk, Using protein interaction networks to understand complex diseases, Computer 45 (3) (2012) 31–38.
[14] Z. Wu, Y. Wang, L. Chen, Network-based drug repositioning, Mol. BioSyst. 9 (6) (2013) 1268–1281.
[15] Z. Li, R.S. Wang, X.S. Zhang, Two-stage flux balance analysis of metabolic networks for drug target identification, BMC Syst. Biol. 5 (Suppl 1) (2011) S11.
[16] O. Folger, L. Jerby, C. Frezza, E. Gottlieb, E. Ruppin, T. Shlomi, Predicting selective drug targets in cancer through metabolic networks, Mol. Syst. Biol. 7 (2011) 501.
[17] A.K. Chavali, K.M. D'Auria, E.L. Hewlett, R.D. Pearson, J.A. Papin, A metabolic network approach for the identification and prioritization of antimicrobial drug targets, Trends Microbiol. 20 (3) (2012) 113–123.
[18] S. Fakhraei, B. Huang, L. Raschid, L. Getoor, Network-based drug-target interaction prediction with probabilistic soft logic, IEEE/ACM Trans. Comput. Biol. Bioinformat. 11 (5) (2014) 775–787.
[19] F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W. Zhou, J. Huang, Y. Tang, Prediction of drug-target interactions and drug repositioning via network-based inference, PLoS Comput. Biol. 8 (5) (2012) e1002503.
[20] S. Alaimo, A. Pulvirenti, R. Giugno, A. Ferro, Drug–target interaction prediction through domain-tuned network-based inference, Bioinformatics (Oxford, England) 29 (16) (2013) 2004–2008.
[21] M. Gonen, Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization, Bioinformatics (Oxford, England) 28 (18) (2012) 2304–2310.
[22] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug-target interaction networks from the integration of chemical and genomic spaces, Bioinformatics (Oxford, England) 24 (2008).
[23] K. Bleakley, Y. Yamanishi, Supervised prediction of drug-target interactions using bipartite local models, Bioinformatics (Oxford, England) 25 (2009).
[24] J.-P. Mei, C.-K. Kwoh, P. Yang, X.-L. Li, J. Zheng, Drug-target interaction prediction by learning from local information and neighbors, Bioinformatics (Oxford, England) (2012).
[25] K. McGarry, U. Daniel, Data mining open source databases for drug repositioning using graph based techniques, Drug Discov. World 16 (2015) 64–71.
[26] H. Chen, Z. Zhang, A semi-supervised method for drug-target interaction prediction with consistency in networks, PLoS ONE 8 (5) (2013) e62975.
[27] M. Re, G. Valentini, Network-based drug ranking and repositioning with respect to DrugBank therapeutic categories, IEEE/ACM Trans. Comput. Biol. Bioinformat. 10 (6) (2013) 1359–1371.
[28] T. van Laarhoven, S.B. Nabuurs, E. Marchiori, Gaussian interaction profile kernels for predicting drug-target interaction, Bioinformatics (Oxford, England) 27 (21) (2011) 3036–3043.
[29] T. van Laarhoven, E. Marchiori, Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile, PLoS ONE 8 (6) (2013) e66952.
[30] X.-Y. Yan, S.-W. Zhang, S.-Y. Zhang, Prediction of drug-target interaction by label propagation with mutual interaction information derived from heterogeneous network, Mol. BioSyst. 12 (2) (2016) 520–531.
[31] Z. Xia, L.-Y. Wu, X. Zhou, S.T. Wong, Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces, BMC Syst. Biol. 4 (2) (2010) 1–16.
[32] X. Zheng, H. Ding, H. Mamitsuka, S. Zhu, Collaborative matrix factorization with multiple similarities for predicting drug-target interactions, in: Book Collaborative Matrix Factorization with Multiple Similarities for Predicting Drug-Target Interactions, ACM, 2013, pp. 1025–1033.
[33] L. Perlman, A. Gottlieb, N. Atias, E. Ruppin, R. Sharan, Combining drug and gene similarity measures for drug-target elucidation, J. Comput. Biol. 18 (2) (2011) 133–145.
[34] H. Luo, P. Zhang, H. Huang, J. Huang, E. Kao, L. Shi, L. He, L. Yang, DDI-CPI, a server that predicts drug-drug interactions through implementing the chemical-protein interactome, Nucleic Acids Res. 42 (2014) W46–W52.
[35] J.T. Dudley, T. Deshpande, A.J. Butte, Exploiting drug–disease relationships for computational drug repositioning, Brief. Bioinform. 12 (4) (2011) 303–311.

[36] J. Li, S. Zheng, B. Chen, A.J. Butte, S.J. Swamidass, Z. Lu, A survey of current trends in computational drug repositioning, Brief. Bioinform. 17 (1) (2016) 2–12.
[37] H. Chen, H. Zhang, Z. Zhang, Y. Cao, W. Tang, Network-based inference methods for drug repositioning, Comput. Math. Methods Med. 2015 (2015) 7.
[38] A. Gottlieb, G.Y. Stein, E. Ruppin, R. Sharan, PREDICT: a method for inferring novel drug indications with application to personalized medicine, Mol. Syst. Biol. 7 (2011) 496.
[39] C. Wu, R.C. Gudivada, B.J. Aronow, A.G. Jegga, Computational drug repositioning through heterogeneous network clustering, BMC Syst. Biol. 7 (5) (2013) 1–9.
[40] K. Yang, H. Bai, Q. Ouyang, L. Lai, C. Tang, Finding multiple target optimal intervention in disease-related molecular network, Mol. Syst. Biol. 4 (2008) 228.
[41] H. Ye, Q. Liu, J. Wei, Construction of drug network based on side effects and its application for drug repositioning, PLoS ONE 9 (2) (2014) e87864.
[42] Y. Wang, S. Chen, N. Deng, Y. Wang, Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data, PLoS ONE 8 (11) (2013) e78518.
[43] Y. Yamanishi, M. Kotera, M. Kanehisa, S. Goto, Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework, Bioinformatics (Oxford, England) 26 (12) (2010) i246–254.
[44] W. Wang, S. Yang, X. Zhang, J. Li, Drug repositioning by integrating target information through a heterogeneous network model, Bioinformatics (Oxford, England) 30 (20) (2014) 2923–2930.
[45] X. Chen, M.-X. Liu, G.-Y. Yan, Drug-target interaction prediction by random walk on the heterogeneous network, Mol. BioSyst. 8 (7) (2012) 1970–1978.
[46] P. Zhang, F. Wang, J. Hu, R. Sorrentino, Label propagation prediction of drug-drug interactions based on clinical side effects, Sci. Rep. 5 (2015) 12339.
[47] H. Ding, I. Takigawa, H. Mamitsuka, S. Zhu, Similarity-based machine learning methods for predicting drug-target interactions: a brief review, Brief. Bioinform. 15 (5) (2014) 734–747.
[49] J. Li, Z. Lu, A new method for computational drug repositioning using drug pairwise similarity, in: Proceedings. IEEE International Conference on Bioinformatics and Biomedicine, 2012, pp. 1–4.
[50] J. Piñero, N. Queralt-Rosinach, À. Bravo, J. Deu-Pons, A. Bauer-Mehren, M. Baron, F. Sanz, L.I. Furlong, DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes, Database: J. Biol. Databases Curation 2015 (2015) bav028.
[51] M.A. van Driel, J. Bruggeman, G. Vriend, H.G. Brunner, J.A.M. Leunissen, A text-mining analysis of the human phenome, Eur. J. Hum. Genet. 14 (5) (2006) 535–542.
[52] E. Pauwels, V. Stoven, Y. Yamanishi, Predicting drug side-effect profiles: a chemical fragment-based approach, BMC Bioinformatics 12 (2011) 169.
[53] M. Kuhn, I. Letunic, L.J. Jensen, Peer Bork; The SIDER database of drugs and side effects. Nucleic Acids Res (2016) 44 (D1): D1075-D1079. http://dx.doi.org/10.1093/nar/gkv1075.
[54] G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, S. Wang, GOSemSim: an R package for measuring semantic similarity among GO terms and gene products, Bioinformatics (Oxford, England) 26 (7) (2010) 976–978.
[55] H. Wickham, Reshaping Data with the reshape Package, J. Stat. Softw 21 (12) (2007) 1–20.
[56] T. Zhou, J. Ren, M. Medo, Y.C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E: Stat., Nonlin, Soft Matter Phys. 76 (4 Pt 2) (2007) 046115.
[57] T. Hwang, R. Kuang, A heterogeneous label propagation algorithm for disease gene discovery, in: Proceedings of the 2010 SIAM International Conference on Data Mining, pp. 583–594.
[58] T. Zhou, J. Ren, M. Medo, Y.-C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E 76 (4) (2007) 046115.
[59] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J.R. Wakeling, Y.-C. Zhang, Solving the apparent diversity-accuracy dilemma of recommender systems, Proc. Natl. Acad. Sci. 107 (10) (2010) 4511–4515.
[60] M. Newman, Networks: An Introduction, Oxford University Press, Inc., 2010.
[61] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[62] Z. Xu, A. Almudevar, D.H. Mathews, Statistical evaluation of improvement in RNA secondary structure prediction, Nucleic Acids Res. 40 (4) (2012) e26.