# FDMine: a graph mining approach to predict and evaluate food-drug interactions

**Md Mostafizur Rahman**
StFX: Saint Francis Xavier University https://orcid.org/0000-0002-1636-7793
**Srinivas Mukund Vadrev**
StFX: Saint Francis Xavier University https://orcid.org/0000-0001-9403-9528
**Arturo Magana-Mora**
Saudi Aramco: Saudi Arabian Oil Co https://orcid.org/0000-0001-8696-7068
**Jacob Levman**
StFX: Saint Francis Xavier University https://orcid.org/0000-0002-9604-3157
**Othman Soufan** ( osoufan@stfx.ca )
St. Francis Xavier University https://orcid.org/0000-0002-4410-1853

**Research article**

**Keywords:** Food-Drug Interaction, Link Prediction, Graph Mining, Adverse Effect, Structure Similarity Profile

**Posted Date:** January 13th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-613772/v2

# FDMine: a graph mining approach to predict and evaluate food-drug interactions

Md. Mostafizur Rahman[1], Srinivas Mukund Vadrev[1], Arturo Magana-Mora[2], Jacob Levman[1] and Othman Soufan[1]

[1]*Department of Computer Science, St. Francis Xavier University, Nova Scotia, Canada*

[2]*Saudi Aramco, EXPEC Advanced Research Center, Drilling Technology Team, Dhahran, 31311, Saudi Arabia.*

**Corresponding author**

Correspondence to Jacob Levman (jlevman@stfx.ca) or Othman Soufan (osoufan@stfx.ca)

## Abstract

Food-drug interactions (FDIs) arise when nutritional dietary consumption regulates biochemical mechanisms involved in drug metabolism. Towards characterizing

the nature of food's influence on pharmacological treatment, it is essential to detect all possible FDIs. In this study, we propose FDMine, a novel systematic framework that models the FDI problem as a homogenous graph. In this graph, all nodes representing drug, food and food composition are referenced as chemical structures. This homogenous representation enables us to take advantage of reported drug-drug interactions for accuracy evaluation, especially when accessible ground truth for FDIs is lacking. Our dataset consists of 788 unique approved small molecule drugs with metabolism-related drug-drug interactions (DDIs) and 320 unique food items, composed of 563 unique compounds with 179 health effects. The potential number of interactions is 87,192 and 92,143 when two different versions of the graph referred to as disjoint and joint graphs are considered, respectively. We defined several similarity subnetworks comprising food-drug similarity (FDS), drug-drug similarity (DDS), and food-food similarity (FFS) networks, based on similarity profiles. A unique part of the graph is the encoding of the food composition as a set of nodes and calculating a content contribution score to re-weight the similarity links. To predict new FDI links, we applied the path category-based (path length 2 and 3) and neighborhood-based similarity-based link prediction algorithms. We calculated the precision@top (top 1%, 2%, and 5%) of the newly predicted links, the area under the

receiver operating characteristic curve, and precision-recall curve. We have performed three types of evaluations to benchmark results using different types of interactions. The shortest path-based method has achieved a precision 84%, 60% and 40% for the top 1%, 2% and 5% of FDIs identified, respectively. We validated the top FDIs predicted using FDMine to demonstrate its applicability and we relate therapeutic anti-inflammatory effects of food items informed by FDIs. We hypothesize that the proposed framework can be used to gain new insights on FDIs. FDMine is publicly available to support clinicians and researchers.

**Keywords:** Food-Drug Interaction; Link Prediction; Graph Mining; Adverse Effect; Structure Similarity Profile.

# Introduction

Drugs bind to targeted receptors on the surface of the cells or enzymes to regulate the rate of chemical reactions. These chemical reactions may be relied upon to treat different diseases and considerably enhance the patients' prognoses. However, drug overdoses or drug interactions may cause critical adverse health conditions. Although the impact of the drugs depends on the affinity of the drug to bind to a specific cell/enzyme receptor, its effectiveness depends on other factors such as when taken alongside other drugs or food. Ideally, drug effects should be consistent across all patients for whom food ingredients do not affect drug response [1]. However, several studies [2, 3] have demonstrated the impact of certain foods, decreasing or increasing the activity of different drugs (food-drug interactions – FDI).

FDIs often cause changes in drug plasma concentrations, which may significantly increase or decrease the effectiveness of the drug [4]. These alterations can occur in three ways: they can boost the actions of of drugs (i.e., increased metabolism), reduce the actions of drugs (i.e., decreased bioavailability), or create an adverse effect.

FDIs can be classified into two primary mechanisms: pharmacokinetic (PK) interactions, and pharmacodynamic (PD) interactions [5]. PK interactions denote the circumstance when foods alter processes related to absorption, distribution, metabolism, and excretion of medications. For example, for a short time after consumption, grapefruit juice slows the metabolism of cyclosporine (e.g.: cytochrome P450 enzymes) [6, 7]. PD interactions are caused by specific interactions between a drug and a food component that results in a a pharmacological effect [8]. An example of a PD interaction is a diet high in vitamin K that antagonizes the therapeutic effects of warfarin (used for blood clot treatments) [5].

FDIs can affect drug absorption levels and, therefore, are known to impact drug discovery [9]. For example, *Moringa oleifera* leaf extract has been used to inhibit cancer cells and to increase the efficacy of chemotherapy in humans [10, 11, 12]. The roots of Erythroxylum pervillei provide pervilleines A, B, C, and F, effective inhibitors of P-glycoprotein, which is linked to multi-drug resistance and low cancer therapeutic response [13]. These are only a few examples that demonstrate the importance of understanding the interactions of food constituents and dietary supplements (containing different chemicals and phytochemicals) with drugs. Consequently, understanding FDIs has the potential to assist physicians, researchers, and patients in reducing ad-verse drug events (ADEs).

Earlier research has been mostly based on clinical studies or literature reviews that focus on specific drug interactions with a limited set of foods [5, 8, 14, 15]. Based on analyzing PD or PK alterations, these studies examine how food items can affect the efficacy of specific drugs. Some studies have focused on a particular group of patients and examined FDI interactions with the types and number of drugs used (e.g., drugs used for chemotherapy, drugs used as anticoagulants) [16, 17, 18]. Although these studies provided valuable information to physicians about the potentialities of FDIs, the level of novel exploration is limited. Computational approaches can, therefore, potentially be used to predict novel FDIs.

Cheminformatics studies have achieved outstanding results in the fields of drug-drug interactions (DDIs), drug-target interactions (DTIs), and new drug discovery. Multiple computational models have been developed for detecting how a particular drug pair interacts towards new drug discovery. The adoption of different machine learning models is rapidly increasing in drug discovery [19]. These models have been used for finding new DDIs. For example, Lee et al. Proposed a deep learning model to predict the pharmacological effects of DDIs using structural similarity profile (SSP), target gene similarity profiles, and gene ontology (GO) term similarity profiles of known drug pairs [20]. Recently DeepDDI, was developed as a multi-label classification model that calculates structural similarity profiles (SSP) of DDIs and uses principal components analysis to reduce features and feed them into a feed-forward deep neural network (DNN) [21]. Another predictive model [22] was developed to delineate currently unknown biological effects of inactive ingredients and generally recognized as safe compounds present in food. A general-purpose method, named Alternative Drug-Drug Interaction, was developed to predict the DDIs [23]. Three combined methods were used, including deep learning, text mining, and graph clustering. Feng et al. proposed DPDDI to predict DDIs without considering the biological and chemical properties [24]. The authors used graph convolution networks (GCN) and DNN as a

predictor. By identifying the topological association of drugs in the DDI network, GCN explores low-dimensional feature representations of drugs.

Several chemoinformatics studies have successfully demonstrated the application of computational models for predicting DTIs. Yo et al. [25] used a deep learning model to predict DTIs using a network representation. The solution is a linear classification model based on using the least absolute shrinkage and selection operator (LASSO) and LASSO-DNN. LASSO helped in feature extraction to predict DTIs. In one of our previous works, we developed DASPfind [26], a novel computational method to predict the DTIs that uses a simple path (up to 3 lengths) to infer novel drug-protein interactions from a graph structure. The graph was derived from similarities among drug-drug, protein-protein, and known drug-protein interactions. Olayan et al. [27] developed the DDR method for predicting DTIs. The authors constructed a heterogenous graph from the known DTIs and multiple similarities among the drug-drug and target-target interactions, used for feature engineering. The engineered features were later used as inputs for a random forest method to predict the novel DTIs. Different studies have developed link prediction approaches to predict DTIs. Lu et al. [28] used link prediction based on similarity indices to predict DTIs. Fokoue et al. [29] developed the Tiresias framework that uses a large-scale similarity-based link prediction based on different drug data to determine the DDIs. The framework uses a large-scale logistic regression model to predict potential DDIs.

Although the implementation has made significant advances of these chemoinformatics models for DDIs and DTIs, FDIs remain poorly addressed. This is mainly due to the inadequacy of resources regarding FDIs since it is often difficult to extract a sufficient number of curated interactions. In addition, for FDI there is no gold standard dataset yet for evaluation. Recently, FooDB [30, 31] was developed as a well-structured and annotated database listing food items and compound composition. Although there is no gold standard dataset for evaluation as in the field of DTIs, we propose using known DDIs. Given the homogenous nature of our graph representations (i.e., all nodes are chemicals), we can resort to certain subnetworks for evaluation. To the best of our knowledge, this is the first work on developing a homogenous graph mining framework for food-drug interactions.

In this study, we propose FDMine, a framework that analyzes FooDB [30, 31] and DrugBank [32] databases to create a comprehensive dataset of small molecules with known food-food interactions (FFIs), DDIs, and FDIs. FDMine uses the simplified molecular-input-line-entry system (SMILE) description to establish similarity profiles and link prediction algorithms to predict the FDIs. The proposed framework uses two different kinds of representations (disjoint and joint) graphs consisting of three subnetworks connected. These subnetworks are drug-drug similarity, food-drug similarity, and food-food similarity. The

rationale behind this approach is to capitalize on the similarity information of different subnetworks and combine it with building a homogenous graph. We consider a unique representation of food items, their compound composition, and the contribution of each compound. After building the graph network, the framework implements a comprehensive set of different link prediction algorithms to predict potential FDIs. The shortest path-based method has achieved a precision 84%, 60% and 40% for the top 1%, 2% and 5%, respectively. In the joint version of the graph, FDMine recovered 27,448 links on average from 27,612 available (i.e., 99.4% recovery with standard deviation of 5.1e^-4).

# Methods

## Databases and datasets preparation

### *DrugBank*

We used the DrugBank (v 5.1.7) database that contains detailed information for each drug (i.e., chemical, pharmaceutical, and pharmacological data) with extensive drug target information (i.e., sequence, pathway, and structure) [32, 33, 34]. The database contains information for a total of 13,680 different drugs. In Drug-Bank, drugs are grouped into five categories, including approved, experimental, investigational, nutraceu-tical, and withdrawn. Drugs can be differentiated as small molecules or biotechnology-driven. The database provides access to the SMILE strings of the drug molecules and reports drug-drug interactions [33].

In this study, we considered the drugs assigned to the approved drug group and have small molecules. This resulted in 1,683 drugs. We further reduced this set of molecules by considering only those having "metabolism (increase or decrease)" related interactions, resulting in 788 unique approved small molecule drugs. FDI interactions are mainly detected in relation to metabolic mechanisms [14]. The details of the drug extraction procedure from the DrugBank dataset can be found in the Additional file 1: Fig. S1.

### *FooDB*

We used the FooDB Version 1.0 dataset in JSON format [30, 31], containing several datasets related to foods, compounds, nutrients, and health effects. In this study, we considered the FooDB content dataset that directly mapped foods to the chemical compounds' composition. Initially, we created a subset of the content dataset that stored the required attributes (i.e., food id, original food name, source id, source type, among others), yielding a total of 19,867 objects. Then, we filtered the extracted data by removing the list of predicted and unknown data entries by using the conditions "citation type == DATABASE" and "source type == COMPOUND". This provides a more accurate source of information. Finally, we only considered the food items mapped to a compound, resulting in 16,230 objects for further analysis.

After the parsing step, we mapped the resulting dataset with the "Compound" information to collect the required details for each compound, including SMILE description and content contribution. In FoodDB, the content range of each compound within a food item is presented (e.g., Strawberry has a content range of Potassium of 0.000 - 187.000 mg/100 g). Finally, we have the SMILE description of the corresponding compounds and the international chemical key (InChiKey) as a unique identifier.

To relate the food compounds to health effects, we retrieved data from the health effects dataset that enabled us to know which food compound has a health effect on the human body. The resulting dataset contains 8,846 objects including 320 unique foods, and 563 unique food compounds having 179 unique health effects. One extracted example is that benzoic acid from American cranberry has an allergenic health effect.

Since the same compounds can be found in different foods, it is necessary to store these data with a naming convention that allows us to differentiate each food with its composition correctly. In this study, we used the following naming convention: FOODXXXX_FDBXXXXX_CompoundName. For example, the data entries "FOOD00005_ FDB000633_ Kaempferol" and "FOOD00008_ FDB000633_ Kaempferol" refer to the same compound Kaempferol with the compound identifier FDB000633 from two different foods (FOOD00005 and FOOD00008). Each compound can be treated differently based on the reported content range in the food item.

The data-flow diagram of the extraction procedure of the FooDB dataset can be found in the Additional file 1: Fig. S2.

***Food composition and compound contribution***

Each food item is composed of a set of chemical compounds. Clearly, the "amount of the original content" of any compound is not the same for each food. For example, the amount of the phytic acid in carrot is 5270.000 ml/100g and buckwheat is 1800.000 ml/100g. Carrot contains approximately three times more phytic acid than buckwheat by mass. Therefore, the contribution of the phytic acid is different for carrot and buckwheat. Consequently, we used the following equation to calculate the contribution of each compound for each food based on the amount contained in the food:

$$Contributionscore(normalized) = \frac{Compoundoriginalcontent \in afooditem}{Totaloriginalcontentofallcompounds \in afood} \quad (1)$$

The range of the normalized contribution is from 0 to 1. Where 0 and 1 contribution refer to a food compound with no contribution or full contribution, respectively.

In the graph, the food item and its compound composition are represented as separate nodes. The normalized contribution score scales edge weights of links connecting compounds to the food item.

More details and an example on the contribution score of food compounds is given in the Additional file 1: Table S1.

## Homogenous Graph Representation

We consider a set of food compounds, $F = \{f_1, f_2, \dots, f_m\}$ and a set of drugs, $D = \{d_1, d_2, \dots, d_n\}$ where $m$ represents the number of food compounds and n represents the number of drugs. We merged all drugs and food compounds into a single graph. So, in our representation, we have a set of drug and food compounds $FD = \{f_1, f_2, \dots, f_m, d_1, d_2, \dots, d_n\}$. Then, we considered the set of an $m * n$ dimensional structure similarity matrices between drugs, between food compounds, and between food-drug. A score between [0, 1] is the degree of similarity. A similarity score close to 0 means that two items are not identical to each other, where the most similar items are represented by a similarity score close to 1. Using this similarity concept, we derived a homogenous graph. From this homogenous graph, we will apply different path category and neighborhood-based similarity-based algorithms to predict the novel FDIs.

## Structure Similarity Profile

A structural similarity profile (SSP) is a feature vector that contains a unique numerical representation after acquiring structural features of individual food compounds and drugs. The SSP contains pairwise structural similarity scores obtained from the comparison among all the 788 approved small molecule drugs of DrugBank and 8,846 unique food compounds. Structural similarity between a pair of nodes (i.e., drug-drug, food-food, and food-drug) was measured by the Tanimoto coefficient. This coefficient is an efficient way to calculate the structure similarity based on the chemical fingerprint [35, 36]. The Tanimoto coefficient is defined as the number of common chemical fingerprints compared to the number of all chemical fingerprints of the two drugs. Chemical fingerprints of each drug were calculated using Morgan/Circular fingerprints [37] (also known as extended-connectivity fingerprint ECFP4 [38]) that is widely used in different studies. ECFP4 showed the best performing fingerprints in the target prediction benchmarks [39, 40] and in small molecule virtual screening [41]. The calculating procedure of the SSP can be found in the Additional file 1: Fig S3.

## Sparse Matrix Representation

We used the similarity profile to derive the sparse matrix representation, later used for plotting the graphs. In this matrix, we eliminated all the zero entries and applied a threshold since some similarity scores contain trivial values and thus may not indicate significant changes. For determining the threshold, we have considered the distribution of the similarity scores. The majority of similarity values lie between 0.3~0.6, hence selecting a high similarity value may drastically change the data-set size. Also, of note, a high threshold will always lead to potential pairs having increased probability of interaction. Several studies have referred to different values in the range of 0.5-0.85 for applying a similarity threshold for the Tanimoto coefficient [42, 43, 44]. While a higher threshold can lead to more potentially valuable hypotheses, it can limit the number of genuinely novel predictions. Table 1 highlights the number of links of each subnetwork after applying a range of similarity thresholds. Compared to a threshold of 0.6, a value of 0.7 would result in 75% fewer number of possible FDIs. Therefore, we choose 0.6 at this step. It should be noted that this parameter is provided as an input argument for the user of FDMine.

Table 1 Number of links in the graph after applying different Tanimoto similarity thresholds

| Tanimoto Threshold | Total Links | DD Links | FF Links | FD Links |
|---|---|---|---|---|
| >= 0.5 | 5,392,354 | 14,298 | 5,228,607 | 149,449 |
| >= 0.6 | 4,177,383 | 2,926 | 4,167,202 | 7,255 |
| >= 0.7 | 3,834,135 | 920 | 3,831,336 | 1,879 |

## Updating Similarity Scores using Food-Compound Contribution

We obtained a total of 4,177,383 similarities using the SSP. Then, we multiplied the similarity score by the normalized contribution of the food compound (Eq. 1). As illustrated in Table 2, when we have a food-drug pair (see row 1), we multiplied the similarity score by the contribution of the food compound. Similarly, we multiplied the similarity score by the higher contribution of the food compound. For example, the contribution of the FOOD00006_ FDB000474_ L-Lysine is 0.007301117, and the FOOD00006_ FDB000556_ L-Alanine is 0.009780473. So, we have considered the maximum value of 0.009780473 to update the similarity score. For drug pairs, similarity scores were preserved.

$$Score = PriorScore(SSP) * ContributionofFoodCompound \quad (2)$$

Table 2 Calculating New Score based on the Food Compound Contribution

| nodeA | nodeB | Prior Score | Contribution (0~1) | New Score |
|---|---|---|---|---|
| DB00136 | FOOD00165 _ FDB012362 _ beta-Sitosterol | 0.6947674 | 0.3459079 | 0.2403255323224 |
| FOOD00006 _ FDB000474 _ L-Lysine | FOOD00006 _ FDB000556 _ L-Alanine | 0.6 | 0.009780473 | 0.005868284 |

After updating the similarity scores in the graph, we consider another threshold using the contribution score. Here, we consider a more relaxed range (0.3, 0.4, 0.5 and 0.6) as compared to the Tanimoto coefficient threshold. In our literature validation, we prepare and discuss another batch of results using a similarity score of 0.3, though a value of 0.5 has been employed for the generation of our primary findings. For a threshold of 0.5, we ended up with 87,192 interactions and 92,143 for disjoint and joint dataset respectively. Table S2 in Additional file 1 lists the number of interactions for the considered range.

## Link Prediction Algorithms

After applying the similarity thresholds, the generated graph had several disjoint subgraphs. We call this the disjoint version. Some link prediction algorithms cannot handle the disjoint version. Therefore, we considered preparing a joint graph. We chose any node (randomly) from each subgraph and added an edge to link all subgraphs to make the joint graph network. Then, a very small edge weight of 1e-5 was assigned to the newly added links, limiting their effect on generating biased hypotheses. We generated results for both versions. A detailed description is available in the Additional file 1: Fig S4.

### *Path Category-based Algorithm*

Our goal is to predict the novel (unknown) FDIs from the generated homogenous graph. A homogenous graph is one where all nodes are of the same type. Different than DTI heterogenous graphs (e.g., drug-protein), nodes in our graph are chemicals. One class of algorithms is based on running the shortest path to find candidate interactions for the considered food and drug pair. Here, we have used 2-length and 3-length pathways. For example, a 2-length path is "Drug1-Food1-Food2" (see Figure 1) connects the Drug1 node with the Food2 node through the similarity between "Drug1 and Food1" and "Food1 and Food2". This is defined as a D-F-F path. As illustrated in Figure 1, the gold color circle denotes the food node and silver color circle denotes the drug node. There are 8 possible combinations of paths (i.e., Drug-Drug-Drug, Drug-Food-Drug, Food-Food-Food, Food-Drug-Food, Drug-Drug-Food, Drug-Food-Food, Food-Drug-Food, and Food-Food-Food).

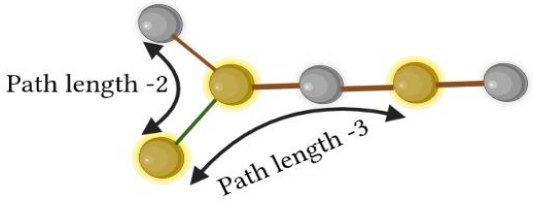


**Figure 1** Example working procedure of the Path Category-based method

For predicting new interactions, any path can be followed. The same applies for 3-length pathway prediction. For example, we can get another new link using 3-path length (Food-Food-Drug-Food). The score for the newly predicted link is calculated according to equation 3, where, $p$ is the path, $n$ is the total number of path and $w$ is the weight of the path:

$$score = Min \sum_{p=1}^{n} P_w \tag{3}$$

Dijkstra's algorithm was used for finding the shortest path where the similarity score is used as the path weight.

***Neighbourhood-based Similarity-based Link Prediction***

In the link prediction, given a graph $G$, the main aim is to predict new edges (drug-food) from the existing graph. Predictions are useful to suggest unknown relations (or interactions) based on edges in the observed graph. In the link prediction, we try to build a similarity measure between pairs of nodes and link the most similar nodes. Link prediction algorithms are very common in many application domains such as, identifying protein-protein interactions [45], drug-drug interactions [29], DTIs [28], social networks [46], reconstructing networks [47], document recommendation, recommendation systems [48], biological networks [49], disease prediction [50], bipartite networks [51], etc.

Here, we applied six different types of link prediction algorithm. They are, Adamic and Adar Coefficient (AA) [50, 52], Common Neighbor (CN) [28, 50, 53], Jaccard Coefficient (JAC) [28, 50, 54], Resource Allocation (RA) [50, 55, 56], Multiple Paths of Length L=3 (L3) [45, 57], and Dice Coefficient (Dice) [58, 59]. All of these algorithms have their scoring function. Each of these algorithms assigns a score for the new predicted links.

***Adamic and Adar Coefficient (AA)***

The Adamic and Adar Coefficient (AA) gives preference to node pairs with more common neighbors but with a lower degree. If there are no common neighbors for a node pair, then the AA score is 0. The AA measure is formulated to connect node pairs that have common neighbors.

$$S_{AA}(a,b) = \sum_{z \in \Gamma(a) \cap \Gamma(b)} \frac{1}{log k_z} \tag{4}$$

Here, $a$ and $b$ are two nodes, and $z$ denotes a common neighbor to both $a$ and $b$. $k$ is the degree of node $z$.

***Common Neighbor (CN)***

In the Common Neighbor (CN) algorithm, the score for link prediction is computed by finding the number ofcommon neighbors between two distinct nodes. Where, $a$ and $b$ are two nodes. $\Gamma(a)$ and $\Gamma(b)$ denote the set of neighbors of nodes $a$ and $b$, respectively.

$$S_{CN}(a,b) = |\Gamma(a) \cap \Gamma(b)| \tag{5}$$

***Jaccard Coefficient (JAC)***

The JAC measure considers only node pairs that have at least one common neighbor. The JAC measure gives equal weight to all common neighbors and does not consider the degree of the common neighbors. The JAC measure gives preferences to node pairs that share a larger fraction of their neighbor. The JAC measure always ranges from 0 to 1 irrespective of the size of the neighborhoods of the vertices. The formula is given below to calculate the JAC. $\Gamma(a)$ and $\Gamma(b)$ denote the set of neighbors of nodes $a$ and $b$, respectively.

$$S_{Jaccard}(a,b) = \frac{|\Gamma(a) \cap \Gamma(b)|}{|\Gamma(a) \cup \Gamma(b)|} \tag{6}$$

***Resource Allocation (RA)***

Resource Allocation (RA) calculates the score based on irregular nodes connecting node $a$ and $b$. The number of resources node $a$ receives from node $b$ through indirect links is called the similarity index. In

the RA each intermediate link contributes a unit of the resource. The RA is also symmetric. $z$ denotes common neighbor of both $a$ and $b$ nodes and k-denotes the degree of node $z$.

$$S_{RAI}(a,b) = \sum_{z \in \Gamma(a) \cap \Gamma(b)} \frac{1}{k_z} \tag{7}$$

***Multiple Paths of Length L=3 (L3)***

Links of high degree nodes prompt multiple and unspecific shortcuts in the network, resulting in biased predictions. This can be avoided by using proper degree of normalization. Such degree of normalization is very important for L3. To eliminate potential degree biases caused by lower degree nodes, we assign a degree normalized L3 score to each node pair $a$ and $b$. Here, u and v are intermediate nodes in the 3-length path.

$$L3_{ab} = \sum_{u,v \in L3} \frac{A_{au} A_{uv} A_{vb}}{\sqrt{k_u k_v}} \tag{8}$$

***Dice Coefficient***

Dice coefficient is similar to the Jaccard Coefficient (JAC). The Dice coefficient is calculated using equation 9, where, $a$ and $b$ are two nodes.

$$S_{Dice}(a,b) = \frac{2 * |a \cap b|}{|a \cup b|} \tag{9}$$

## Performance evaluation

To measure the performance of applied link prediction approaches, we adopted the idea of precision@k [60, 61] or top $k$ predictive rate [53, 62]. This metric is also known as $r$-precision [63, 64, 65, 66]. precision@k is the recommended measure for link prediction algorithms [67]. It refers to the percentage of true positives among only the top $k$ ranked predicted links. Given the ranked output of the graph, we need to evaluate the ranking precision of the methods.

Following [26], we chose the top 1%, 2%, and 5% as the value of $k$. In general, the area under the receiver operating characteristic curve (AUROC) or (AUC) is used to evaluate performance of classification models. Nevertheless, recent studies have shown that AUROC is unsuitable for checking the performance of the link prediction algorithms [56, 68, 69, 70]. Another statistical measure is the area under the precision-recall

curve (PRC), which provides a more accurate assessment especially when dealing with imbalanced datasets [71]. In this study, we used, precision@top, AUC, and PRC as performance metrics.

In order to compute some of the measures, we had to derive true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). To perform this, we ranked the predicted links in descending order based on the rank score given by the link prediction methods. Then, we considered several thresholds as cutoff values. The starting threshold is the minimum score given by the link prediction methods. Then we increase by a step size of 0.1, which was selected to ensure sufficient granularity in computing the area under the curve. We repeated this step until the threshold value is the same as the maximum score given by the link prediction algorithm. For each specific threshold score, if we found the known link in the test dataset matched with the newly predicted link and the score is greater than the threshold, we considered this matching as a true positive (TP) for evaluative purposes. Given an unknown link, which does not match the test dataset, but was predicted by the link prediction algorithm, and the score is greater than the threshold, we consider the case a false positive (FP). Similarly, when we found a known link (same as the test dataset and in the newly predicted links), but the score was below the threshold, we consider this a false negative (FN). Lastly, when we found any unknown link with the score below the threshold, we assign the sample as a true negative (TN). Using the TP, FP, TN, and FN we calculated the "precision@top-1%", "precision@top-2%", "precision@top-5%", AUC, and PRC.

## Data splitting for testing

To evaluating the performance of link prediction algorithms, the test data is generated by excluding a collection of links from the full homogenous networks. Our homogenous network contains drug-drug similarity, food-drug similarity, and food-food similarity. We split 30% of links randomly to make the test data set, while the rest of the 70% of links are used for the training dataset. For stability, we repeat this evaluation ten times and report average performance.

## Ground-truth evaluation using DDS

Contrary to food-protein interactions [26], there is no accessible gold standard for widely confirmed food-drug interactions. Therefore, we resorted to the extracted drug-drug interactions from DrugBank for ground truth evaluation. Since the graph representation in FDMine is homogenous (i.e., all nodes are chemicals), we can consider any part of the graph as a representative set of evaluation. Here, we remove 30% of the drug-drug links in the graph. Then, we execute the framework and report top ranked cases for the precision

368 evaluation. We split 30% DDS links (randomly) for making the test data set, while the rest of the 70% DDS,
369 and all FDS, FFS links are used in the training dataset. Here, we measured the precision in terms of recov-
370 ering the original links in the DDS subgraph. It should be noted that we also performed evaluation using a
371 random subset of any type of links (see Results).
372
373 We have performed three types of evaluations to benchmark the results. In the first evaluation, a drug can
374 have a link with another drug because of certain similarity scores. In the second evaluation, a drug will have
375 a correct link with another drug only if it is reported in the DrugBank database. The difference between the
376 second and third evaluation is that the original links in the second evaluation are assumed based on the
377 established similarity measures. Both evaluations will help us establish a comprehensive overview of link
378 recovery in general and the validity of these recovered links using DrugBank. Although drug-drug interac-
379 tions are examined in these two evaluations, they both provide estimates for the accuracy of food-drug
380 predictions since the graph is homogenous in nature. The following Table 3 lists all the evaluative ap-
381 proaches we have performed in this study.
382
383 Table 3 List of evaluation approaches

| **Title** | **Evaluation** | **Graph** | **Correct predictions** | **Methods** |
|---|---|---|---|---|
| Evaluation 1 | Remove random 30% of links from the DDIs (repeat 10 times) | Comprehensive evaluation for recovery of DDS similarity links | Match predicted links with the actual ones | All methods are applied |
| Evaluation 2 | Remove random 30% of links (repeat 10 times) | Ground Truth using DrugBank | Match predicted links with DrugBank reported interactions | SP_2 (the best from evaluation 1 over disjoint graph) and RA (the best from evaluation 1 over joint graph) |

| | | | | |
|---|---|---|---|---|
| Evaluation 3 | Remove random 30% of links (repeat 10 times) | Whole graph including DDS, FDS, FFS | Match predicted links with the actual ones | SP_2 (the best from evaluation 1 over disjoint graph) and RA (the best from evaluation 1 over joint graph) |

## Implementation

We have deployed the code and run all experiments on a server with RAM 64 GB, and Intel(R) Core(TM) i9-7980XE CPU @ 2.60GHz (18 Cores, 36 Threads). For DrugBank data preprocessing, we used Compute Canada cluster and to calculate SSP we used Google Colaboratory (a product from Google Research).

## Our Proposed FDMine Framework

The FDMine framework (see Figure 2) is composed of several phases. In **Phase 1**, raw data is parsed from DrugBank and FooDB databases. In **Phase 2**, we execute two steps including a) building a homogenous network based on the structure similarity profile and b) updating the weights of the homogenous network using food compound contributions. Next, the graph is prepared with nodes representing drugs, food and food compounds' composition. In the graph, links are weighted by similarity and contribution scores (see **Phase 3** in Figure 2). When applying the similarity thresholds, the homogenous network produces multiple subgraphs (disjoint graph). We build another version called the joint homogenous graph network and consider executing several link prediction algorithms including applied path category-based and neighborhood-based similarity-based approaches. In the final **Phase 4**, we rank the newly predicted link (based on the score given by our methods), test the performance of the applied methods with the test dataset and finally, consult the literature to validate the top FDIs found using the different methods. For testing, we perform comparison using ground-truth and report literature validation for our leading findings (see Results and Discussion section).

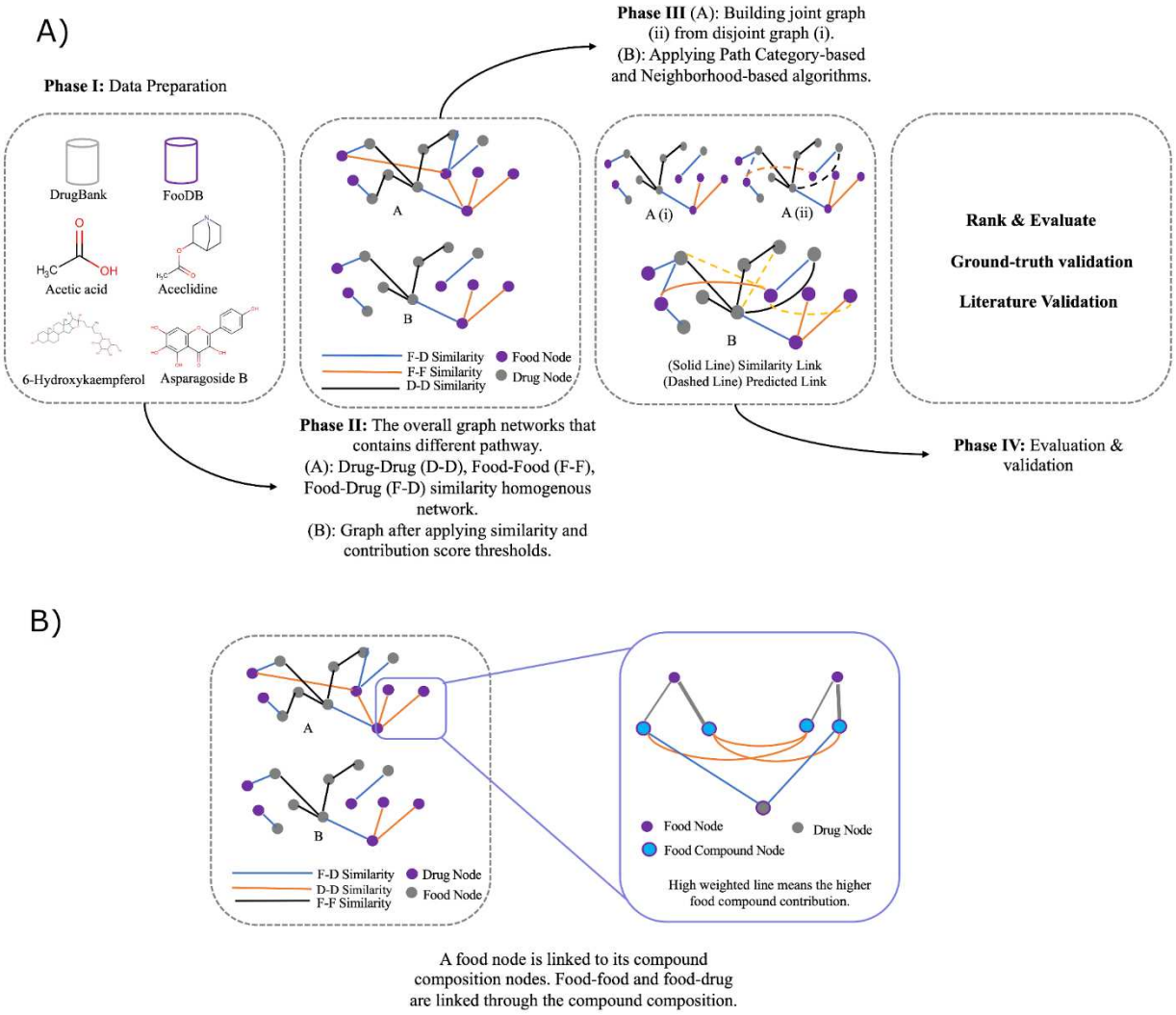


**Figure 2** The framework of FDMine. The main steps are 1) preparing a comprehensive dataset describing FDIs by analyzing the whole DrugBank and FooDB databases with a unique representation of food composition 2) defining a scoring function for computing chemical compound contribution in food items, 3) implementing a set of path category-based (path length 2 and 3) and different neighborhood-based similarity-based algorithms to discover new FDIs from two different homogenous (disjoint and joint) graph networks, and 4) used the precision@k metric and calculated the precision@top (top 1%, 2%, and top 5%) for drug-drug links to verify the accuracy of the algorithms with the test dataset.

## Results and Discussion

The next subsections describe in detail the FDMine performance evaluation and the analysis of the novel FDI predictions.

## Prediction Results of FDMine

***Evaluation 1: Comprehensive evaluation for the recovery of DDS similarity links***

As explained earlier, DDS similarity links are a priority in our evaluation setup as it establishes a ground truth evaluation (see Evaluation 2 results). Here, drug-drug links are based on the similarity scorings we computed. We have applied two different link prediction approaches over two different types of homogenous graph networks. One is the disjoint graph network, and the other is the joint graph network. The applied methods are the path category-based and neighborhood-based similarity-based link prediction algorithms. We used path lengths 2 and 3 for the path category-based algorithm. SP_2 and SP_3 are used to describe (Path length 2), and (Path length 3), respectively. From neighborhood-based similarity-based link prediction, we applied Academic Adar (AA), Common Neighbor (CN), Jaccard Index (JAC), Dice Coefficient (Dice), Resource Allocation (RA), and Multiple paths of length l=3 (L3).

Table 4 provides a summary of different models over the disjoint graph network. For the disjoint graph, the SP_2 outperformed other methods. The precision rate for the top 1% (i.e., precision@top-1) is 84% for SP_2 while RA, the second best has achieved 64%. For precision@top-2, SP_2 achieved the best results with 60% and L3, the second best 42%. The highest value for the precision@top-5 was achieved by the SP_2 (40%). In the disjoint version of the graph, neighborhood-based similarity-based methods achieved, on average 17% with variant standard deviation each. However, SP_3 always showed a low performance (05%, 03%, 02% for precision@top-1, precision@top-2, and precision@top-5 respectively) compared to all other methods. SP_2 achieved 52% and 26% AUC and PRC, respectively. All neighborhood-based similarity-based methods achieved more than 80% (AUC) except L3 which had a reported precision of 60%. The PRC scores of the RA, AA, and CN were 70%, 65%, and 60% respectively.

When considering the joint version of the graph, different results were attained. The neighborhood-based similarity-based methods showed best results for the top precision@top-1, precision@top-2, and precision@top-5. For the precision@top-1, the RA achieved the best result (71%), followed by AA (67%). For the precision@top-2, L3 and RA both yielded similar performance (39%). Additionally, all neighborhood-based similarity-based methods produced the same result (16%) for precision@top-5. Contrary to the case of the disjoint version of the graph, the performance of SP_2 was weak. The SP_2 achieved, 23%, 15%, and 9% for the precision@top-1, precision@top-2, and precision@top-5 respectively. For the joint graph, the neighborhood-based similarity-based algorithms achieved AUC of more than 90% except L3 (65%). The value of the PRC is also high for the neighborhood-based similarity-based methods. The PRC scores for the RA, AA, CN were 87%, 86%, and 84% respectively. However, SP_3 always (disjoint and joint

graphs) showed the weakest results in terms of all performance metrics (precision@top, AUC, and PRC). Table 5 summarizes the different models over the joint graph network. The comparison graph for the precision@top-1%, precision@top-2%, and precision@top-5% are provided in Figure 3. For more details, see the Additional File 1 Figures S6 and S7.

Table 4 Comparison of the precision@top (average), AUC, PRC over eight different methods on the disjoint graph network

| **Method** | **'Precision@Top-1 (%)** | **Precision@Top-2 (%)** | **Precision@Top-5 (%)** | **AUC (%)** | **PRC (%)** |
|---|---|---|---|---|---|
| SP_2 | **84 (±6.3)** | **60 (±5.3)** | **40 (±2.5)** | 52 (±1.0) | 26 (±1.0) |
| SP_3 | 05 (±5.6) | 03 (±3.1) | 02(±1.4) | 59 (±23.0) | 03 (±3.0) |
| AA | 56 (±1.6) | 36 (±1.0) | 17 (±0.6) | 88 (±.01) | 65 (±1.7) |
| CN | 53 (±1.5) | 33 (±1.1) | 17 (±0.4) | 88 (±1.0) | 60 (±1.6) |
| RA | 64 (±1.7) | 40 (±1.4) | 17 (±0.6) | 80 (±3.5) | **70 (±1.7)** |
| L3 | 58 (±1.9) | 42 (±1.2) | 17 (±0.6) | 60 (±4) | 30 (±3.1) |
| JAC | 40 (±1.6) | 31 (±0.5) | 17 (±0.5) | 94 (±0.4) | 34 (±1.8) |
| Dice | 40 (±1.6) | 31 (±0.5) | 17 (±0.5) | **97 (±0.7)** | 35 (±2.0) |

Table 5 Comparison of the precision@top (average), AUC, PRC over eight different methods on the joint graph network

| **Method** | **Precision@Top-1 (%)** | **Precision@Top-2 (%)** | **Precision@Top-5 (%)** | **AUC (%)** | **PRC (%)** |
|---|---|---|---|---|---|
| SP 2 | 23 (±1.8) | 15 (±1.5) | 09 (±0.9) | 38 (±1) | 08 (±0.07) |
| SP 3 | 0.1 (±0.2) | 0.1 (±0.1) | 0.1 (±0.0) | 88 (±31) | 00 (±0) |

| | | | | | |
|---|---|---|---|---|---|
| AA | 67 (±0.9) | 37 (±0.5) | 16 (±0.2) | 95 (±0.6) | 86 (±0.6) |
| CN | 65 (±0.9 | 36 (±0.5) | 16 (±0.2) | 94 (±0.4) | 84 (±0.7) |
| RA | **71 (±0.9)** | **39 (±0.5)** | **16 (±0.2)** | 92 (±2.4) | **87 (±1.9)** |
| L3 | 31 (±1.2) | 39 (±0.5) | 16 (±0.2) | 65 (±3.9) | 23 (±2.0) |
| JAC | 59 (±0.7) | 35 (±0.4) | 16 (±0.2) | 97 (±0.3) | 66 (±1.4) |
| Dice | 59 (±0.7) | 35 (±0.4) | 16 (±0.2) | **98 (±0.2)** | 65 (±1.4) |

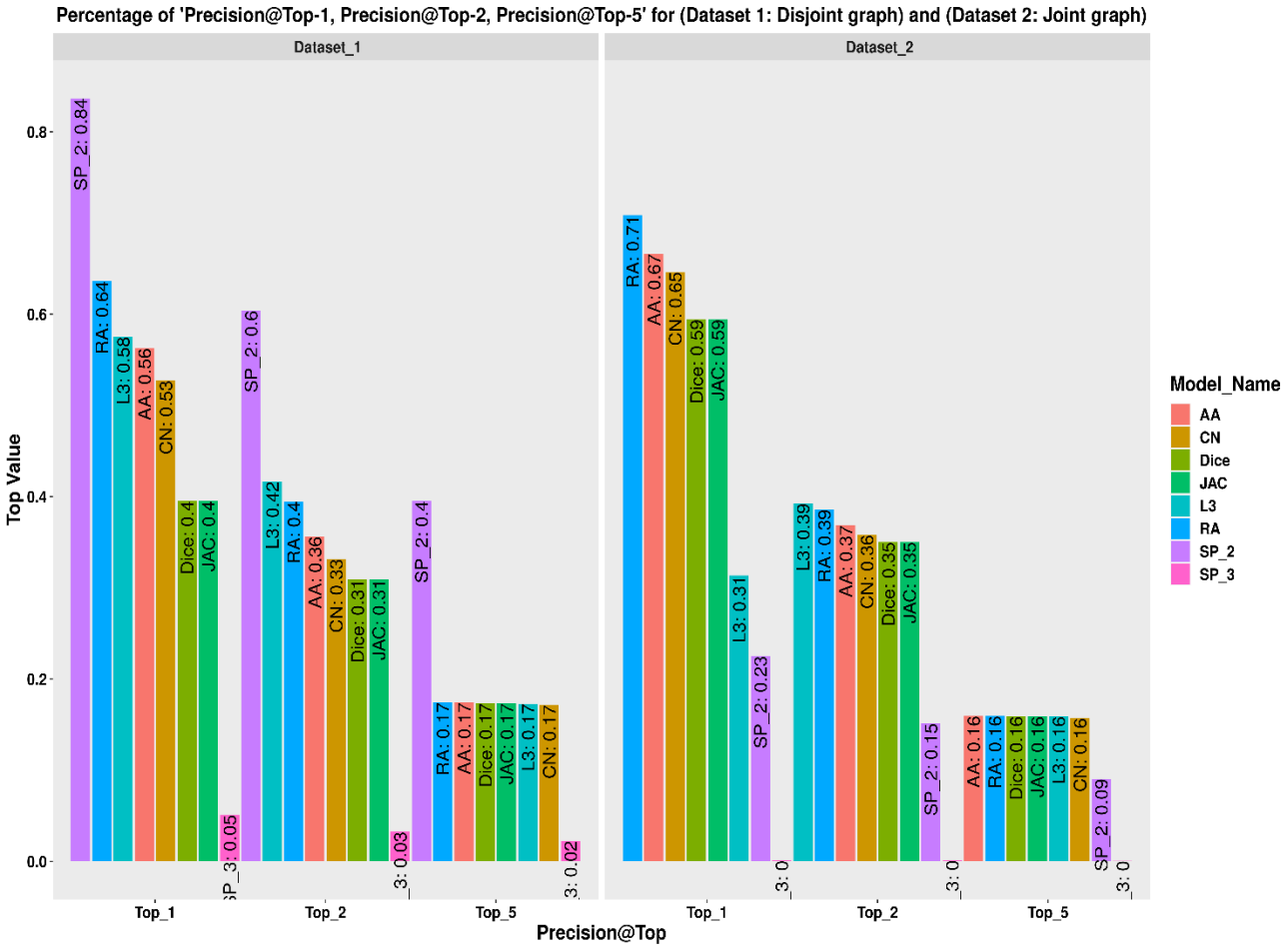


**Figure 3** Comparison of the precision@top over eight methods and two different graph networks

***Evaluation 2: Ground truth evaluation using DrugBank***

The dataset we constructed using DrugBank and FooDB contains drug-drug links. The disjoint and joint dataset contains 2,926, and 6,581 drug-drug links respectively. From evaluation 1, out of 2,926, and 6,581, our method managed to discover 1,706, and 4,178 of those links respectively, reported as DDIs in the DrugBank. We have considered these 1,706 and 4,178 as known DDIs and as ground truth. To cross-validate the performance of FDMine we excluded a portion of known DDIs (or ground truth) as a test dataset from the main dataset and the rest of the dataset was used to train the models. Then, we calculate the precision@top-1%, precision@top-2%, and precision@top-5% and found approximately the same performance of FDMine with the disjoint dataset and slightly better results for the joint dataset. Here, we have chosen only the best models, SP_2 for the disjoint dataset and RA for the joint dataset. Table 6 and Table 7 provides the performance of FDMine with the ground truth test dataset.

Table 6 Performance evaluation of ground truth using disjoint dataset and path category-based (path length-2) method

| Method | Proportion | #Test DDI | #Matched DDI | Precision@ Top-1 (%) | Precision@ Top-2 (%) | Precision@ Top-5 (%) |
|---|---|---|---|---|---|---|
| SP_2 | 0.6 | 1023 | 864.8 (±13.85) | 84.49 (±5.09) | 72.29 (±6.59) | 47.11 (±4.00) |
| | 0.5 | 853 | 750.7 (±9.91) | 78.21 (±7.50) | 64.73 (±4.86) | 42.20 (±2.79) |
| | 0.4 | 682 | 613.5 (±6.06) | 76.31 (±5.77) | 57.51 (±5.53) | 36.81 (±3.88) |
| | 0.3 | 511 | 469.1 (±4.93) | 60.60 (±9.06) | 43.69 (±5.44) | 28.09 (±2.57) |

Table 7 Performance evaluation of ground truth using joint dataset and Neighborhood-based Similarity-based (RA) Method

| Method | Proportion | #Test DDI | #Matched DDI | Precision@ Top-1 (%) | Precision@ Top-2 (%) | Precision@ Top-5 (%) |
|---|---|---|---|---|---|---|
| RA | 0.6 | 2506 | 2413.0 (±9.12) | 94.93 (±0.30) | 93.16 (±0.71) | 51.55 (±0.71) |
| | 0.5 | 2089 | 2027.4 (±12.01) | 95.99 (±0.35) | 86.64 (±1.29) | 40.63 (±1.01) |
| | 0.4 | 1671 | 1628.4 (±6.97) | 96.75 (±0.49) | 72.15 (±1.07) | 31.64 (±0.54) |
| | 0.3 | 1253 | 1223.3 (±4.18) | 90.96 (±1.05) | 54.59 (±0.86) | 22.97 (±0.43) |

***Evaluation 3: prediction results for whole graph (DDS, FFS, FDS)***

Here we randomly assigned 30% of all (DD, FF, FD) links from the whole dataset to make the test dataset, and the rest of the 70% was used to train the model. We applied 'shortest path length 2' over the disjoint and 'RA' over joint graph. The 30% test dataset from the disjoint and joint dataset contains 26,157 and 27,612 links respectively. The FDMine was able to recover an average of 9612.6 (±5723.06) and 27448.4

(±14.20) links respectively from the disjoint and joint dataset using 'shortest path length 2' and 'RA' methods respectively.

## New Food Drug Interaction Prediction

After comparing the different approaches for link prediction, we executed the FDMine framework to find top candidates for FDIs. In the framework, we consider taking the top results from the joint and disjoint versions. At the final stage of FDMine, we surveyed the literature to find supporting evidence to the generated predictions. We have performed two batches using different contribution scores (i.e., 0.5 and 0.3, respectively). The default value in the FDMine framework is a 0.5 contribution score. The results as listed in Additional file 2, have shown some repeated drugs in top findings due to a higher threshold value. A high threshold value will lead to removing more connections in the graph. This will lead to more disjoint subgraphs and nodes with higher connections within the subgraphs gain higher rank scores. Therefore, we consider a more relaxed threshold and generate Batch-2 results (i.e., contribution score of 0.3). In this batch, we see more diversity in results. Additional file 2 lists all Batch-1 results, and Additional file 3 lists all Batch-2 results with a description of the experiments used in each. We analyzed all results of both batches and discussed here the insights driven from two types of evidence including: 1) linking food to anti-inflammatory effects based on known biological pathways and 2) linking food to pharmacological effects based on matching functions of a drug and a chemical substructure found in food.

### Food compound compositions with Anti-inflammatory effects (biological pathway driven evidence)

The results in this section are part of Batch-1 results (see Additional file 2). Our findings using a literature review indicate possible pairing of drug and nutraceutical food components. As per the summary in Table 8, the interactions we obtained appear to affect key biological pathways including - Prostaglandin biosynthesis for inflammatory response [72], beta-adrenergic signaling for cardiac output modulation [73] and GABA pathway [74] - a GABA based inhibitory neurotransmitter that down-regulates CNS stimulation [75]. After examining the results in Table 8, we have found that dietary fatty acids like Oleic acid (FDB012858), Erucic acid (FDB004287), (Z,Z)-9,12-Octadecadienoic (FDB012760) and Elaidic acid (FDB002951) available in foods like Onions - FOOD00006, Garden Cress - FOOD00099, Pomegranate-FOOD00151, etc. can affect prostaglandin biosynthesis via PPAR mediated mechanism and Gabaergic pathway. Figures 4 and 5 highlight the list of these compounds and their interaction with Peroxisome pro-

liferator-activated receptor (PPAR) and GABA-mediated effects, respectively. Similarly, we found evidence of food components like Eugenol (FDB012171), Carvacrol (FDB014512), which can potentially substantiate hypotensive effects when taken with beta adrenergic drugs. For example, Eugenol has been known to cause vasodilation via vanilloid TRPV4 receptors found on endothelial muscles in arteries. Beta-adrenergic drugs are prescribed to patients suffering from hypertension to decrease blood pressure (BP). So, when combined, this can cause an elevated drop in BP.

Prostaglandins are compounds that play a role in the anti-inflammatory pathway during injury [76]. An essential molecular building block in humans is arachidonic acid. It interacts with the Peroxisome proliferator-activated receptor (PPAR) to form various prostaglandins [76] or anti-inflammatory compounds. Various dietary fatty acids (see Table 8; Oleic acid, Linoleic acid, Erucic Acid, Eldaic acid) are also absorbed via the exogenous chylomicron pathway and hydrolysed for various tissues to absorb them for further processing [77]. Some of our predicted compound items include Oleic acid - FDB012858, and Erucic acid - FDB004287, that are similar to Arachidonic acid and are analogous [78] structures, belonging to the fatty acid group and are found in many dietary sources including Celery - FOOD00015, Peanuts (FOOD00016) and Burdock - FOOD00017 (See Table 8). Our literature review has highlighted reported evidence on the influence of these dietary fatty acids on the Arachidonic acid cycle. Arachidonic acid is a precursor for the synthesis of various other biomolecules, associated with anti-inflammatory pathways [79]. During injury, inflammation occurs and causes arachidonic acid to bind with PPAR-gamma receptors as shown in Figure 4 to form prostaglandins or protective anti-inflammatory agents to curb the injury [80]. Fatty acids (see Table 8) also compete with arachidonic acid during injury or inflammation to produce various substituted prostaglandins belonging to a family of derivative compounds known as eicosanoids [81], via PPAR [82]. Since the substituted prostaglandins are not exactly derived from arachidonic acid, they show slightly fewer anti-inflammatory profiles as compared to other eicosanoids produced directly from arachidonic acid [83]. It is worth noting that arachidonic acid belongs to the list of essential fatty acids including alpha-linoleic acid and docosahexaenoic acid [83]. There has been evidence to show that dietary sources such Linoleic acid, Erucic acid and Elaidic acid (see Table 8) did increase PPAR gene expression in healthy subjects [84]. In 2012 Hung-Tsung Wu et al. also showed the interaction of oleic acid with PPAR-g receptors [85]. These results may suggest taking drugs like Doconexent - DB03756 with foods such as FOOD00099 - Garden Cress, FOOD00151 - Pomegranate, FOOD00009 - Chives, FOOD00062 - Hazelnut, FOOD00525 - Macadamia can alter the normal dynamics of anti-inflammatory responses. Arachidonic acid is biosynthesized from dietary linoleic acid and released by phospholipases during inflammation. This pathway is also known as the COX or Cyclooxygenase pathway [86].

Table 8 Depicts some of our top correlations of food substances that can potentially be involved in food drug interactions when combined with a drug with a similar activity. Each food component can link to any of the drugs as long as they are in the same batch.

| Food component | Food source ID | Food name | Pharmacological actions | Drug | References | Batch |
|---|---|---|---|---|---|---|
| **Oleic acid** | FOOD00006 | Garden Onion | Dietary fatty acids like Oleic acid can compete with arachidonic acid by interacting with PPAR-g receptor to form prostaglandins<br><br>They can also cross the blood brain barrier and interact with GABA receptors to induce anxiolytic & possible anti-epileptic effects | Vigabatrin, Pregabalin, Gabapentin Doconexent | [82, 85, 89, 90, 91, 92] | Top 10 in joint and disjoint - batch 1 (supplementary file 2) |
| | FOOD00009 | Chives | | | | |
| | FOOD00011 | Cashew Nus | | | | |
| | FOOD00012 | Pineapple | | | | |
| | FOOD00015 | Wild celery | | | | |
| | FOOD00016 | Peanuts | | | | |
| | FOOD00017 | Burdock | | | | |
| | FOOD00021 | Asparagus | | | | |
| | FOOD00024 | Brazil Nut | | | | |
| | FOOD00026 | Borage | | | | |
| **Erucic acid** | FOOD00099 | Garden Cress | | | | |
| **Elaidic acid** | FOOD00151 | Pomegranate | | | | |
| **(Z,Z)-9,12-Octadecadienoic acid** | FOOD00009 | | | | | |
| **Eugenol** | FOOD00179 | Cloves | Euginol causes vasodilation via vanilloid | Betaxolol, Atenelol, | [86, 87] | Top 20 in joint and |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | TRPV4 receptors found on endothelial muscles found on arteries. Euginol & Capsacin have a vanilloid ring. TRPV4 is involved in BP regulation via various mechanisms | Esmolol, Bisprolol, Metoprolol | | disjoint - batch 2 (supplementary file 3) |
| **Isopropyl-2-methylphenol** | FOOD00089 | Hyssop | p-Cymene has been reported to cause smooth muscle vasodilation and has antihypertensive effects | | | |
| **1-Isopropyl-4-methylbenzene** | FOOD00013 | Dill | Also known as p-cymene. It has been shown to cause sedative effects via GABA adrenergic receptors and also causes vasodilation of smooth arterial muscles | | | |
| **1-Methoxy-4-(2-propenyl)benzene** | FOOD00137<br>FOOD00019 | Anise<br>Tarragon | Methyl Chavicol has been reported as an adjunct therapy for treatment of hypertension, found in anise. | | | |

554

555

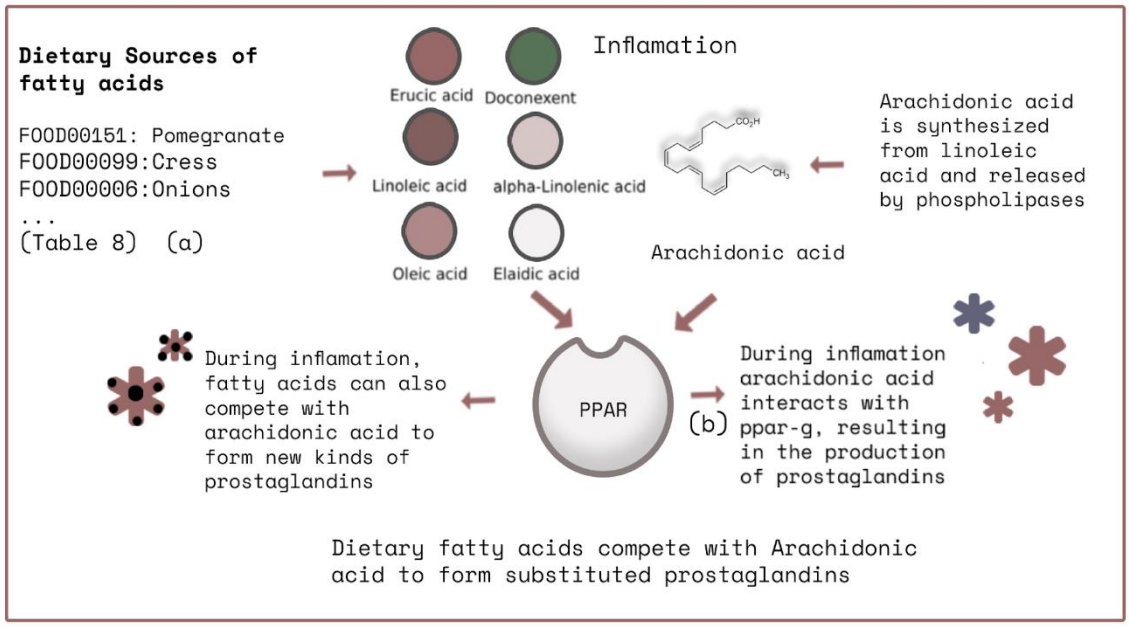


**Figure 4** An illustration depicting the effect of dietary fatty acids on COX pathway a) Various foods are rich sources of dietary fatty acids b) During inflammation, Arachidonic acid interacts with PPAR to produce prostaglandins c) Dietary Fatty acids can compete with Arachidonic acid during inflammation at PPAR to form substituted prostaglandin variants.

**Food compound composition with pharmacological effects (similar function-driven evidence)**

Here, we relaxed the contribution score to 0.3 (i.e., Batch-2) to obtain a diverse set of results (see Additional file 3). In this part of our literature validation, we analyze the potential of similar functions of drugs and food compounds on specific diseases. The results in Table 8 highlight some correlations with a group of drugs called beta-adrenergic drugs and essential oils. Our top correlated pairs of food and drug observed that both of them caused reduced blood pressure. Beta-blockers are used to treat hypertension in patients. Beta-blockers consist of b1, b2, and b3 subtype receptors. Beta-blockers can fall into various categories based on the extent of selectivity of binding across these subtypes. For example, Atenolol (DB00335), Bisoprolol (DB00612), Metoprolol (DB00264) and Esmolol (DB00187) are b1 selective blockers [88]. The effects of b1 blockade include a decrease in cardiac output by inhibiting the SA and AV nodes, thereby decreasing stroke volume [89]. Propranolol (DB00571) and Penbutolol (DB01359), on the other hand, are non-selective beta-adrenergic blockers. Studies have also observed that beta-blockers may also contribute to GABA turnover in the cerebrum [90].

The results suggest that beta-blocker drugs like Atenolol, Betaxolol, Esmolol, Oxprenolol, Penbutalol, and Propranolol can interact in the form of synergism when combined with a specific compound composition including p-Cymene - FDB014512, Eugenol - FDB012171, and Carvacrol (terpenoid substances). For example, Marcio et al. 2011 reported that monoterpenoids like p-Cymene - FDB014512 and Carvacrol have vasorelaxant effects [86].

We were able to confirm that fatty acids (Oleic acid (FDB012858), Erucic acid (FDB004287), (Z,Z)-9,12-Octadecadienoic (FDB012760) and Elaidic acid (FDB002951) ) can cross the blood-brain barrier and be beneficial to relieve anxiety [91]. They are also believed to act via stimulation of GABA-A based receptors. Benzodiazepines, barbiturates [92] and some anticonvulsants act by modulating the GABA receptors [93]. The inhibitory effects of GABA help relieve seizures. However, drugs like Pregabalin and Gabapentin instead act by blocking calcium or sodium channels to help stabilize seizures. Although this is not directly interacting with GABA receptors, it helps reduce excitatory neurotransmitters. Thus, they may help substantiate antiepileptic activity by increasing amounts of GABA.

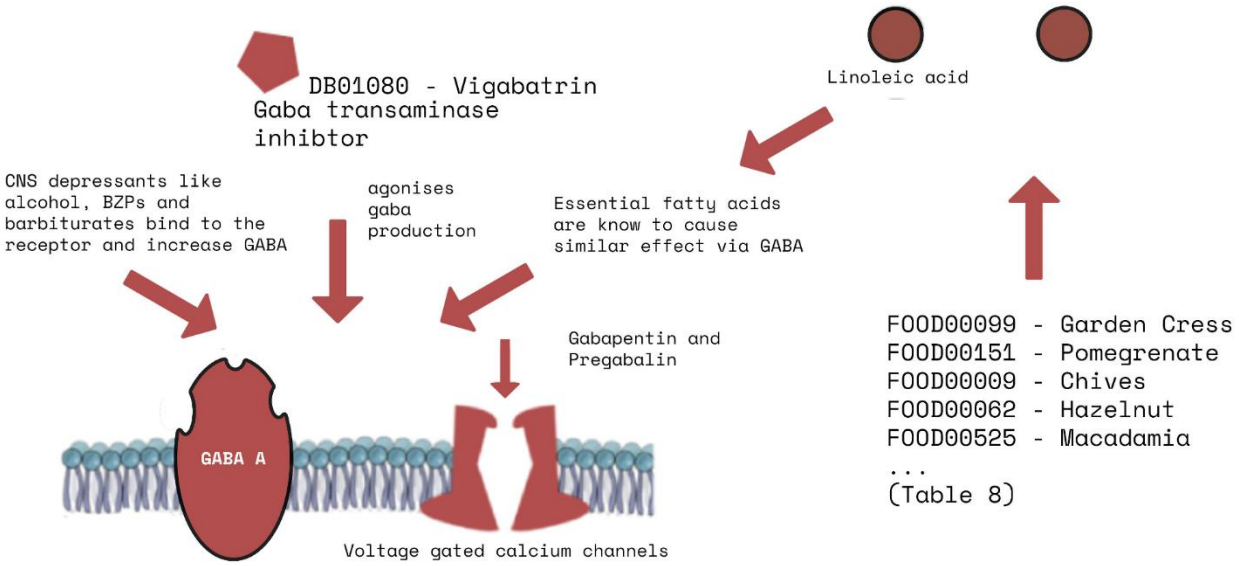


**Figure 5** An illustration depicting Gabaergic drug mechanisms. Dietary sources containing fatty acids increase the production of GABA. Taking drugs like Vigabatrin, pregabalin & Gabapentin with such a diet can increase Gabaergic effects.

In summary, the discussed pairs of food ingredients and drugs can influence their own pharmacokinetics. For example, taking beta-adrenergic drugs with food containing terpenes like Eugenol and Methyl chavicol can potentially cause more pronounced antihypertensive effects. Taking antiepileptic medications along with foods containing fatty acids can potentially elevate overall GABA levels significantly than when they are taken individually. Moreover, dietary fatty acids can also interact with the PPAR receptor during inflammation to produce variations of prostaglandins. This demonstrates the feasibility of using our FDMine framework to identify potential food and drug interactions.

## Conclusion

In this study, we introduced FDMine as a framework to infer the interaction between food compounds and drugs using a homogenous graph representation. We considered several resources to construct food-drug, drug-drug, and food-food similarity profiles. FDMine uses established path category-based and neighborhood-based similarity methods to predict FDIs efficiently. A subset of Drug-drug interactions was used as ground-truth evaluations. This proposed methodology is based on encoding all entities including drug and food into a homogenous graph of chemical nodes. Therefore, any part of this graph can then be used as a representative evaluation, potentially informative to clinicians and researchers. We have performed additionally two types of evaluations to benchmark results using different parts of the graph. The shortest path-based method has achieved a precision 84%, 60% and 40% for the top 1%, 2% and 5%, respectively. FDMine was able to achieve an average 99.4% recovery rate from 27,612 available links in the joint version of the graph. We validated the top FDIs predicted using FDMine to demonstrate the applicability of the model. In the literature validation, we discussed the therapeutic effects of a group of food items. We observed that a set of FDIs may reduce blood pressure, have anti-inflammatory effects or reduce seizure. The benchmark results and literature review suggest that FDMine can help to identify FDIs precisely and may represent an advanced strategy in drug discovery.

## Availability of data and materials

The code and datasets supporting the conclusions of this article are included within the article (and its additional files) or is made available at https://github.com/mostafiz67/FDMine_Framework

## Competing interests

The authors declare that they have no competing interests.

## Author contributions

MR and OS conceptualized the problem. MR was responsible for solution development and implementation. SV and AM were responsible for validating the new predictions. AM, JL and OS reviewed the text and the evaluation of the work. JL and OS supervised the study.

## Funding information

This work was supported by a Natural Science and Engineering Research Council of Canada, Canada Research Chair grant (grant number 231266) to JL, a Canada Foundation for Innovation and Nova Scotia Research and Innovation Trust infrastructure grant to JL, and a Natural Science and Engineering Research Council of Canada Discovery Grant to JL. This research was supported in part by the Heaps Chair Endowment Fund at St. Francis Xavier University through the Dr. H. Stanley & Doreen Alley Heaps Chairship. This research was enabled in part by support provided by Compute Canada (www.computecanada.ca) and by Google Cloud under the GCP research credits program.

## Acknowledgements

Authors would like to acknowledge Sumaiya Amin for some preliminary data preparation in this work.

## References

1. Bushra, R., Aslam, N., Khan, A.Y.: Food-drug interactions. Oman medical journal 26(2), 77 (2011)
2. Schmidt, L.E., Dalhoff, K.: Food-drug interactions. Drugs 62(10), 1481–1502 (2002)
3. Won, C.S., Oberlies, N.H., Paine, M.F.: Mechanisms underlying food–drug interactions: inhibition of intestinal metabolism and transport. Pharmacology & therapeutics 136(2), 186–201 (2012)
4. Mouly, S., Morgand, M., Lopes, A., Lloret-Linares, C., Bergmann, J.: Drug-food interactions in internal medicine: What physicians should know? La Revue de medecine interne 36(8), 530–539 (2015)
5. Ased, S., Wells, J., Morrow, L.E., Malesker, M.A.: Clinically significant food-drug interactions. The Consultant Pharmacist® 33(11), 649–657 (2018)
6. Hollander, A.A., van Rooij, J., Lentjes, E.G., Arbouw, F., van Bree, J.B., Schoemaker, R.C., van Es, L.A., van der Woude, F.J., Cohen, A.F.: The effect of grapefruit juice on cyclosporine and prednisone metabolism in transplant patients. Clinical Pharmacology & Therapeutics 57(3), 318–324 (1995)

7. Dahan, A., Altman, H.: Food–drug interaction: grapefruit juice augments drug bioavailability—mechanism, extent and relevance. European journal of clinical nutrition 58(1), 1–9 (2004)
8. Koziolek, M., Alcaro, S., Augustijns, P., Basit, A.W., Grimm, M., Hens, B., Hoad, C.L., Jedamzik, P., Madla, C.M., Maliepaard, M., et al.: The mechanisms of pharmacokinetic food-drug interactions–a perspective from the ungap group. European Journal of Pharmaceutical Sciences 134, 31–59 (2019)
9. Goldstein, L.H., Elias, M., Ron-Avraham, G., Biniaurishvili, B.Z., Madjar, M., Kamargash, I., Braunstein, R., Berkovitch, M., Golik, A.: Consumption of herbal remedies and dietary supplements amongst patients hospitalized in medical wards. British journal of clinical pharmacology 64(3), 373–380 (2007)
10. Berkovich, L., Earon, G., Ron, I., Rimmon, A., Vexler, A., Lev-Ari, S.: Moringa oleifera aqueous leaf extract down-regulates nuclear factor-kappab and increases cytotoxic effect of chemotherapy in pancreatic cancer cells. BMC complementary and alternative medicine 13(1), 1–7 (2013)
11. Hermawan, A., Nur, K.A., Dewi, D., Putri, P., Meiyanto, E., et al.: Ethanolic extract of moringa oleifera increased cytotoxic effect of doxorubicin on hela cancer cells. Journal of Natural remedies 12(2), 108–114 (2012)
12. Al-Asmari, A.K., Albalawi, S.M., Athar, M.T., Khan, A.Q., Al-Shahrani, H., Islam, M.: Moringa oleifera as an anti-cancer agent against breast and colorectal cancer cell lines. PloS one 10(8), 0135814 (2015)
13. Nirmala, M.J., Samundeeswari, A., Sankar, P.D., et al.: Natural plant resources in anti-cancer therapy-a review. Res Plant Biol 1(3), 01–14 (2011)
14. Mouly, S., Lloret-Linares, C., Sellier, P.-O., Sene, D., Bergmann, J.-F.: Is the clinical relevance of drug-food and drug-herb interactions limited to grapefruit juice and saint-john's wort? Pharmacological research 118, 82–92 (2017)
15. de Boer, A., Van Hunsel, F., Bast, A.: Adverse food–drug interactions. Regulatory Toxicology and Pharmacology 73(3), 859–865 (2015)
16. Segal, E.M., Flood, M.R., Mancini, R.S., Whiteman, R.T., Friedt, G.A., Kramer, A.R., Hofstetter, M.A.: Oral chemotherapy food and drug interactions: a comprehensive review of the literature. Journal of oncology practice 10(4), 255–268 (2014)
17. Di Minno, A., Frigerio, B., Spadarella, G., Ravani, A., Sansaro, D., Amato, M., Kitzmiller, J.P., Pepi, M., Tremoli, E., Baldassarre, D.: Old and new oral anticoagulants: food, herbal medicines and drug interactions. Blood reviews 31(4), 193–203 (2017)
18. Gupta, R.C., Chang, D., Nammi, S., Bensoussan, A., Bilinski, K., Roufogalis, B.D.: Interactions between antidiabetic drugs and herbs: an overview of mechanisms of action and clinical implications. Diabetology & metabolic syndrome 9(1), 1–12 (2017)
19. Stephenson, N., Shane, E., Chase, J., Rowland, J., Ries, D., Justice, N., Zhang, J., Chan, L., Cao, R.: Survey of machine learning techniques in drug discovery. Current drug metabolism 20(3), 185–193 (2019)
20. Lee, G., Park, C., Ahn, J.: Novel deep learning model for more accurate prediction of drug-drug interaction effects. BMC bioinformatics 20(1), 1–8 (2019)

21. Ryu, J.Y., Kim, H.U., Lee, S.Y.: Deep learning improves prediction of drug–drug and drug–food interactions. Proceedings of the National Academy of Sciences 115(18), 4304–4311 (2018)
22. Reker, D., Shi, Y., Kirtane, A.R., Hess, K., Zhong, G.J., Crane, E., Lin, C.-H., Langer, R., Traverso, G.: Machine learning uncovers food-and excipient-drug interactions. Cell reports 30(11), 3710–3716 (2020)
23. Allahgholi, M., Rahmani, H., Javdani, D., Weiss, G., M´odos, D.: Addi: Recommending alternatives for drug–drug interactions with negative health effects. Computers in Biology and Medicine 125, 103969 (2020)
24. Feng, Y.-H., Zhang, S.-W., Shi, J.-Y.: Dpddi: a deep predictor for drug-drug interactions. BMC bioinformatics 21(1), 1–15 (2020)
25. You, J., McLeod, R.D., Hu, P.: Predicting drug-target interaction network using deep learning model. Computational biology and chemistry 80, 90–101 (2019)
26. Ba-Alawi, W., Soufan, O., Essack, M., Kalnis, P., Bajic, V.B.: Daspfind: new efficient method to predict drug–target interactions. Journal of cheminformatics 8(1), 1–9 (2016)
27. Olayan, R.S., Ashoor, H., Bajic, V.B.: Ddr: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. Bioinformatics 34(7), 1164–1173 (2018)
28. Lu, Y., Guo, Y., Korhonen, A.: Link prediction in drug-target interactions network using similarity indices. BMC bioinformatics 18(1), 1–9 (2017)
29. Fokoue, A., Sadoghi, M., Hassanzadeh, O., Zhang, P.: Predicting drug-drug interactions through large-scale similarity-based link prediction. In: European Semantic Web Conference, pp. 774–789 (2016). Springer
30. Naveja, J.J., Rico-Hidalgo, M.P., Medina-Franco, J.L.: Analysis of a large food chemical database: Chemical space, diversity, and complexity. F1000Research 7 (2018)
31. FooDB: Foodb version 1.0 (2017)
32. Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., Woolsey, J.: Drugbank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research 34(suppl 1), 668–672 (2006)
33. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al.: Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic acids research 46(D1), 1074–1082 (2018)
34. Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., Hassanali, M.: Drugbank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research 36(suppl 1), 901–906 (2008)
35. Bajusz, D., R´acz, A., H´eberger, K.: Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? Journal of cheminformatics 7(1), 1–13 (2015)
36. Alazmi, M., Kuwahara, H., Soufan, O., Ding, L., Gao, X.: Systematic selection of chemical fingerprint features improves the gibbs energy prediction of biochemical reactions. Bioinformatics 35(15), 2634–2643 (2019)

37. Morgan, H.L.: The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. Journal of Chemical Documentation 5(2), 107–113 (1965)
38. Rogers, D., Hahn, M.: Extended-connectivity fingerprints. Journal of chemical information and modeling 50(5), 742–754 (2010)
39. Awale, M., Reymond, J.-L.: Web-based tools for polypharmacology prediction. In: Systems Chemical Biology, pp. 255–272. Springer, (2019)
40. Awale, M., Reymond, J.-L.: Polypharmacology browser ppb2: target prediction combining nearest neighbors with machine learning. Journal of chemical information and modeling 59(1), 10–17 (2018)
41. Riniker, S., Landrum, G.A.: Open-source platform to benchmark fingerprints for ligand-based virtual screening. Journal of cheminformatics 5(1), 1–17 (2013)
42. Rai, A., Kumar, V., Jerath, G., Kartha, C., Ramakrishnan, V.: Mapping drug-target interactions and synergy in multi-molecular therapeutics for pressure-overload cardiac hypertrophy. NPJ systems biology and applications 7(1), 1–11 (2021)
43. Gottlieb, A., Stein, G.Y., Oron, Y., Ruppin, E., Sharan, R.: Indi: a computational framework for inferring drug interactions and their associated recommendations. Molecular systems biology 8(1), 592 (2012)
44. Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., Friedman, C.: Drug—drug interaction through molecular structure similarity analysis. Journal of the American Medical Informatics Association 19(6), 1066–1074 (2012)
45. Kov´acs, I.A., Luck, K., Spirohn, K., Wang, Y., Pollis, C., Schlabach, S., Bian, W., Kim, D.-K., Kishore, N., Hao, T., et al.: Network-based prediction of protein interactions. Nature communications 10(1), 1–8 (2019)
46. Al Hasan, M., Chaoji, V., Salem, S., Zaki, M.: Link prediction using supervised learning. In: SDM06: Workshop on Link Analysis, Counter-terrorism and Security, vol. 30, pp. 798–805 (2006)
47. Guimer`a, R., Sales-Pardo, M.: Missing and spurious interactions and the reconstruction of complex network. Proceedings of the National Academy of Sciences 106(52), 22073–22078 (2009)
48. Chen, H., Li, X., Huang, Z.: Link prediction approach to collaborative filtering. In: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05), pp. 141–142 (2005). IEEE
49. Clauset, A., Moore, C., Newman, M.E.: Hierarchical structure and the prediction of missing links in networks. Nature 453(7191), 98–101 (2008)
50. Folino, F., Pizzuti, C.: Link prediction approaches for disease networks. In: International Conference on Information Technology in Bio-and Medical Informatics, pp. 99–108 (2012). Springer
51. Daminelli, S., Thomas, J.M., Dur´an, C., Cannistraci, C.V.: Common neighbours and the local-community-paradigm for topological link prediction in bipartite networks. New Journal of Physics 17(11), 113037 (2015)
52. Adamic, L.A., Adar, E.: Friends and neighbors on the web. Social networks 25(3), 211–230 (2003)
53. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. Journal of the American society for information science and technology 58(7), 1019–1031 (2007)

762 54. Jaccard, P.: ’E comparative study of floral distribution in a portion of the alps and jura. Bull Soc Vaudoise
763 Sci Nat 37, 547–579 (1901)
764 55. Zhou, T., Lu¨, L., Zhang, Y.-C.: Predicting missing links via local information. The European Physical
765 Journal B71(4), 623–630 (2009)
766 56. Yang, Y., Lichtenwalter, R.N., Chawla, N.V.: Evaluating link prediction methods. Knowledge and
767 Information Systems 45(3), 751–782 (2015)
768 57. Chen, Y., Wang, W., Liu, J., Feng, J., Gong, X.: Protein interface complementarity and gene duplication
769 improve link prediction of protein-protein interaction network. Frontiers in genetics 11 (2020)
770 58. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26(3), 297–302 (1945)
771 59. Sorensen, T.A.: A method of establishing groups of equal amplitude in plant sociology based on similarity
772 of species content and its application to analyses of the vegetation on danish commons. Biol. Skar. 5, 1–34
773 (1948)
774 60. Crichton, G., Guo, Y., Pyysalo, S., Korhonen, A.: Neural networks for link prediction in realistic biomedical
775 graphs: a multi-dimensional evaluation of graph embedding-based approaches. BMC bioinformatics 19(1),
776 1–11 (2018)
777 61. Lu¨, L., Zhou, T.: Link prediction in weighted networks: The role of weak ties. EPL (Europhysics Letters)
778 89(1), 18001 (2010)
779 62. Chen, H., Zhang, Z., Zhang, J.: In silico drug repositioning based on the integration of chemical, genomic
780 and pharmacological spaces. BMC bioinformatics 22(1), 1–12 (2021)
781 63. Wang, C., Satuluri, V., Parthasarathy, S.: Local probabilistic models for link prediction. In: Seventh IEEE
782 International Conference on Data Mining (ICDM 2007), pp. 322–331 (2007). IEEE
783 64. O’Madadhain, J., Hutchins, J., Smyth, P.: Prediction and ranking algorithms for event-based network data.
784 ACM SIGKDD explorations newsletter 7(2), 23–30 (2005)
785 65. Backstrom, L., Leskovec, J.: Supervised random walks: predicting and recommending links in social
786 networks. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining,
787 pp. 635–644 (2011)
788 66. Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N.V., Rao, J., Cao, H.: Link prediction and recommendation
789 across heterogeneous social networks. In: 2012 IEEE 12th International Conference on Data Mining, pp.
790 181–190 (2012). IEEE
791 67. Kerrache, S., Alharbi, R., Benhidour, H.: A scalable similarity-popularity link prediction method. Scientific
792 reports 10(1), 1–14 (2020)
793 68. Muscoloni, A., Michieli, U., Cannistraci, C.V.: Local-ring network automata and the impact of hyperbolic
794 geometry in complex network link-prediction. arXiv preprint arXiv:1707.09496 (2017)
795 69. Garcia-Gasulla, D., Ayguad´e, E., Labarta, J., Cort´es, U.: Limitations and alternatives for the evaluation of
796 large-scale link prediction. arXiv preprint arXiv:1611.00547 (2016)

70. Wang, W., Cai, F., Jiao, P., Pan, L.: A perturbation-based framework for link prediction via non-negative matrix factorization. Scientific reports 6(1), 1–11 (2016)
71. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 233–240 (2006)
72. Kumar, N.G., Contaifer, D., Madurantakam, P., Carbone, S., Price, E.T., Van Tassell, B., Brophy, D.F., Wijesinghe, D.S.: Dietary bioactive fatty acids as modulators of immune function: implications on human health. Nutrients 11(12), 2974 (2019)
73. do Vale, G.T., Ceron, C.S., Gonzaga, N.A., Simplicio, J.A., Padovan, J.C.: Three generations of β-blockers: history, class differences and clinical applicability. Current hypertension reviews 15(1), 22–31 (2019)
74. Tritsch, N.X., Granger, A.J., Sabatini, B.L.: Mechanisms and functions of gaba co-release. Nature Reviews Neuroscience 17(3), 139–145 (2016)
75. Jorgensen, E.M.: Gaba. WormBook: The Online Review of C. elegans Biology [Internet] (2005)
76. Ricciotti, E., FitzGerald, G.A.: Prostaglandins and inflammation. Arteriosclerosis, thrombosis, and vascular biology 31(5), 986–1000 (2011)
77. Engelking, L.R.: Textbook of Veterinary Physiological Chemistry, Updated 2/e. Academic Press, (2010)
78. Di Pasquale, E., Chahinian, H., Sanchez, P., Fantini, J.: The insertion and transport of anandamide in synthetic lipid membranes are both cholesterol-dependent. PLoS One 4(3), 4989 (2009)
79. Higgins, A., Lees, P.: The acute inflammatory process, arachidonic acid metabolism and the mode of action of anti-inflammatory drugs. Equine Veterinary Journal 16(3), 163–175 (1984)
80. Delves, P.J., Roitt, I.M.: Encyclopedia of Immunology, pp. 2024–2027. Academic Press, (1998)
81. Baker, R.R.: The eicosanoids: a historical overview. Clinical biochemistry 23(5), 455–458 (1990)
82. Varga, T., Czimmerer, Z., Nagy, L.: Ppars are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1812(8), 1007–1022 (2011)
83. Kaur, N., Chugh, V., Gupta, A.K.: Essential fatty acids as functional components of foods-a review. Journal of food science and technology 51(10), 2289–2303 (2014)
84. Ortun˜o Sahagu´n, D., M´arquez-Aguirre, A., Quintero-Fabi´an, S., L´opez-Roa, R., Rojas-Mayorqu´ın, A.: Modulation of ppar-γ by nutraceutics as complementary treatment for obesity-related disorders and inflammatory diseases. PPAR research 2012 (2012)
85. Wu, H.-T., Chen, W., Cheng, K.-C., Ku, P.-M., Yeh, C.-H., Cheng, J.-T.: Oleic acid activates peroxisome proliferator-activated receptor δ to compensate insulin resistance in steatotic cells. The Journal of nutritional biochemistry 23(10), 1264–1270 (2012)
86. Santos, M.R., Moreira, F.V., Fraga, B.P., Souza, D.P.d., Bonjardim, L.R., Quintans-Junior, L.J.: Cardiovascular effects of monoterpenes: a review. Revista Brasileira de Farmacognosia 21(4), 764–771 (2011)

87. Peixoto-Neves, D., Wang, Q., Leal-Cardoso, J.H., Rossoni, L.V., Jaggar, J.H.: Eugenol dilates mesenteric arteries and reduces systemic bp by activating endothelial cell trpv 4 channels. British journal of pharmacology 172(14), 3484–3494 (2015)
88. Farzam, K., jan, A.: Beta blockers. StatPearls (2021)
89. Alhayek, S., Preuss, C.V.: Beta 1 receptors. StatPearls (2021)
90. Remiszewska, M., Jastrzebski, Z., Czyzewska-Szafran, H., Wutkiewicz, M.: Antihypertensive treatment with beta blockers and gabaergic transmission in rat brain. Acta Poloniae Pharmaceutica 52(2), 185–186 (1994)
91. Bernal-Morales, B., Cueto-Escobedo, J., Guill´en-Ruiz, G., Rodr´ıguez-Landa, J.F., Contreras, C.M.: A fatty acids mixture reduces anxiety-like behaviors in infant rats mediated by gabaa receptors. BioMed research international 2017 (2017)
92. Campo-Soria, C., Chang, Y., Weiss, D.S.: Mechanism of action of benzodiazepines on gabaa receptors. British journal of pharmacology 148(7), 984–990 (2006)
93. Czapinski, P., Blaszczyk, B., Czuczwar, S.J.: Mechanisms of action of antiepileptic drugs. Current topics in medicinal chemistry 5(1), 3–14 (2005)

# Supplementary Information

**Additional file 1: Figure S1**. The DrugBank Dataset extraction procedure. **Figure S2.** The FooDB dataset extraction procedure. **Figure S3**. Calculating Structure Similarity Profile Using Tanimoto Coefficient. **Table S1.** Calculating the contribution of a food compound in a food. **Figure S4**. Disjoint and Joint Graph. **Figure S5**. Precision@Top. **Figure S6.** Precision@top comparison of eight different methods over the disjoint graph network. **Figure S7.** Precision@top comparison of eight different methods over the joint graph network. **Figure S8.** Area Under the Curve (AUC) for path category-based (dataset 1: disjoint graph). **Figure S9.** Area Under the Curve (AUC) for path category-based (dataset 2: joint graph). **Figure S10.** Area Under the Curve (AUC) for neighborhood-based similarity-based (dataset 1: disjoint graph). **Figure S11.** Area Under the Curve (AUC) for neighborhood-based similarity-based (dataset 2: joint graph). **Figure S12.** Precision-Recall Curve (PRC) for path category-based (dataset 1: disjoint graph). **Figure S13.** Precision-Recall Curve (PRC) for path category-based (dataset 2: joint graph). **Figure S14.** Precision-Recall Curve (PRC) for path neighborhood-based similarity-based (dataset 1: disjoint graph). **Figure S15.** Precision-Recall Curve (PRC) for neighborhood-based similarity-based (dataset 2: joint graph). **Table S2**. Number of links in the graph after applying different food compound contribution score.

**Additional file 2: Table S1.** Top 10 FDIs found from path category-based (path length-2) method over disjoint graph. Records might appear repeated, but food item IDs are different in this table. **Table S2.** Top 20 FDIs from path category-based (path length-2) method over disjoint and joint graph. **Table S3.** Top 10 FDIs from Neighborhood-based Similarity-based method over joint graph.

**Additional file 3: Table S1.** Top 25 FDIs found from path category-based (path length-2) method over disjoint graph. **Table S2.** Top 25 FDIs found from path category-based (path length-2) method over joint graph. **Table S3.** Top 20

866 common FDIs found from path category-based (path length-2) method over disjoint and joint graph. **Table S4.** Top
867 25 FDIs from Neighborhood-based similarity-based methods method over the joint graph.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.docx
- AdditionalFile2.docx
- AdditionalFile3.docx

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.docx
- AdditionalFile2.docx
- AdditionalFile3.docx