Drug-Drug Interaction Discovery: Kernel Learning from Heterogeneous Similarities

Devendra Singh Dhami[a,*], **Gautam Kunapuli**[a], **Mayukh Das**[a], **David Page**[b], and **Sriraam Natarajan**[a,c]

[a]Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas, United States

[b]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, United States

[c]School of Informatics, Computing & Engineering, Indiana University Bloomington, United States (On Leave)

Abstract

We develop a pipeline to mine complex drug interactions by combining different similarities and interaction types (molecular, structural, phenotypic, genomic etc). Our goal is to learn an optimal kernel from these heterogeneous similarities in a supervised manner. We formulate an extensible framework that can easily integrate new interaction types into a rich model. The core of our pipeline features a novel kernel-learning approach that tunes the weights of the heterogeneous similarities, and fuses them into a Similarity-based Kernel for Identifying Drug-Drug interactions and Discovery, or SKID^3. Experimental evaluation on the DrugBank database shows that SKID^3 effectively combines similarities generated from chemical reaction pathways (which generally improve precision) and molecular and structural fingerprints (which generally improve recall) into a single kernel that gets the best of both worlds, and consequently demonstrates the best performance.

Keywords

1. Introduction

Drug-drug interactions (DDIs) occur when multiple medications are co-administered and can potentially cause adverse effects on the patients. DDIs have emerged, around the world, as a major cause of hospital admissions, rehospitalizations, emergency room visits, and even death [1]. These numbers are even more stark among older adults, who are more likely to be prescribed multiple medications; the study by Becker et al. [1] identifies that the elderly have an increased risk factor of as much as 8.5 times over the general population.

---

*Corresponding author: devendra.dhami@utdallas.edu (Devendra Singh Dhami).

Consequently, DDIs contribute to increased hospital stays and increasing costs of health care, even though up to 50% of these adverse drug effects (ADEs) are preventable [2]. While regulatory agencies such as the U. S. Food and Drug Administration have rigorous approval processes for new drugs, controlled clinical trials do not always uncover all possible drug interactions. For example, the last stage of the FDA approval process involves a Phase III clinical trial, which typically enrolls 1000–5000 individuals, while the drug may be prescribed to millions of patients after approval. In addition to clinical trials, *in vitro* and *in vivo* experiments are also used to identify DDIs. However, these approaches are highly labor-intensive, costly and time-consuming. Another factor is that many known DDIs involve medications such as anti-inflammatories or anticoagulants, which are prescribed for common and chronic conditions. Other confounding factors that make studying DDIs a difficult challenge include dosage variations and demographic variability.

All of these challenges have led to a shift in research towards *in silico* approaches that leverage advances in AI and machine learning [3] for DDI discovery. These approaches for DDI can be viewed in one of two ways.

- The feature-based view, which roughly categorizes the approaches based on the type of DDI features used. These are either text-based (which involves the analysis of abstracts or EHRs) or structure-based (which involve the study of chemical, molecular and pharmacological properties). Our approach is structure-based.

- The algorithm-based view distinguishes between approaches as classification (which treat DDI discovery as a binary classification problem) and clustering (which assume that similar drugs may interact with a same drug). Our approach is a hybrid of both these paradigms.

In our prior work, we have performed ADE discovery and subgroup discovery from electronic health records (EHR) [4] and text-mining of medical journal abstracts [5]. These approaches address the problem of *post-marketing surveillance*, that is, they seek to exploit the new information available after a drug has been approved and has been prescribed to larger, more diverse populations. In this work, we address *pre-trial discovery*, that is, we reframe the problem as one of studying drug-drug interactions, rather than taking a single drug and finding adverse events associated with it. The primary motivation is to preemptively identify potential DDI and ADE risks during drug design. As we show in this work, our novel formulation incorporates elements of *both classification and similarity based algorithms*, which improves discovery as well as explainability. The result is a kernel that we call SKID[3] (Similarity-based Kernel for Identifying Drug-Drug interactions and Discovery).

Our problem setting differs from current approaches in three significant ways, that motivated us to develop SKID[3]:

- A majority of current work focuses on drug-interaction discovery through *information extraction*, specifically through text mining. These approaches attempt to identify drug interactions from various unstructured text-based sources

such as biomedical journals and semi-structured sources such as electronic health records.

We approach the problem by looking at structured sources of information for insights into drug interactions. That is, we characterize drug similarities using different properties of drugs such as molecular structure and pharmacological interaction pathways. This allows us to pose the DDI discovery problem as a structure prediction [6] *task.*

- The approaches that do use structural information generally aim to extract explicit vector representations of properties such as 3-d structure, which allows the application of off-the-shelf machine-learning techniques such as support vector machines and kernel learning [7].

  We, instead, analyze structural similarities between drugs in ways that are natural and intuitive to their representation (such as random walks on chemical interaction pathways), rather than forcing an artificial and uninterpretable embedding in a vector space.

- Finally, many current approaches focus on a single type of interaction or similarity, whether it is discovered from text sources or from structural analysis. This is a significant drawback, as this analysis approach ignores the diverse pharmacological facets to drug-drug interaction to look at one (or a few) interaction types in isolation.

  We develop a general and extensible framework that admits heterogeneous characterizations arising from any source including text-based, molecular structure, pharmacological, phenotypic, genomic, therapeutic similarities. This allows us to exploit diverse characterizations of drug similarities from various perspectives, fusing them into one coherent, interpretable model.

We make the following contributions with our proposed solution to address the above limitations:

- We characterize molecular similarity between two drugs using a novel approach: *knowledge-refined random walks* to measure the *reachability* of one drug from another; reachability informs the intuition that drugs that are more reachable are more interactive. As far as we are aware, this is the first work on exploiting bias knowledge to characterize drug similarities for DDI discovery.

- We develop a novel framework that combines multiple similarity measures into unified kernel that exploits and fuses their potential. In addition to our novel reachability measure (described above), we also use four other measures that capture molecular and chemical similarities through SMILES strings and MACCS fingerprints.

- We formulate DDI as a *kernel-learning problem* that fuses heterogeneous similarity measures. Our formulation enables us to treat each similarity as a different *view of drug interactions*. By fusing similarities from different sources, our formulation aims to reconcile various (molecular, pharmacological etc.,)

views into a single model. Further, our formulation incorporates terms to capture both individual as well as neighborhood interactions, leading to greater robustness.

- From a machine-learning standpoint, our formulation is general in that it admits a *variety of regularization and loss functions*. In this work, we show our approach for a specific formulation that attempts to simultaneously align the optimal kernel with the heterogeneous similarity measures as well as predict the drug-drug interactions.

- From an optimization standpoint, our formulation is a *bilinear program*, which is a non-convex optimization problem. We illustrate an alternating minimization approach for solving this problem; this approach identifies robust and relevant local solutions for DDI discovery and scales well with the underlying drug database size.

- Our empirical evaluation on a data set constructed from DrugBank uncovers previously known drug-drug interactions with high accuracy. Furthermore, a closer inspection of "false positives" and "false negatives" identified by SKID[3] reveals that it has identified drug-drug interactions, missing from Drug-Bank, but existing in other independent sources. This clearly demonstrates the potential of our approach to perform DDI discovery. More specifically, it also offers us a path forward: DDI discovery via active learning with semi-supervised data, which is the real-world problem setting.

The rest of the paper is organized as follows: after reviewing related work in the next section, we define the problem of DDI prediction/discovery. We then present similarity measures and formulate kernel learning for DDI discovery. Next, we present our comprehensive experimental evaluation before concluding the paper by motivating interesting research directions.

2. Related Work

The interactions of a drug can be specified in two ways: (1) the drug has an adverse effect on the human body, called adverse drug events (ADEs), and (2) the drug interacts with another drug called drug-drug interactions (DDIs). Most recent research has focused on finding ADEs from text. Different approaches have been taken in order to identify and discover ADEs in the machine learning community, especially from the natural language processing (NLP) perspective. Chee et al. [8] make use of ensemble classifiers to extract ADEs, while Liu et al. [9] used transductive SVMs to extract ADEs from online health forums. Gurulingappa et al. [10] use NLP with support vector machines (SVMs) to extract ADEs from MEDLINE casereports. Karlsson et al. [11] and Page et al. [4] perform ADE information extraction from EHR data. More recently, Kang et al. [12] took a knowledge-based approach for extracting ADEs from bio-medical text, while Natarajan et al. [13] use Markov logic networks for the same problem.

The problem of DDI prediction and discovery has received far less attention, although similarity-based methods have proven to be very popular. The problem of DDI discovery is a

pairwise classification task, which lends itself very well to kernel-based methods [7]. Kernels are naturally suited to representing pairwise similarities, and are constructed directly from the data vectors during pre-processing. Most similarity-based methods for DDI discovery/prediction also use text sources such as biomedical research literature as the underlying data source, and construct NLP-based kernels from these medical documents [14, 15]. Our work differs considerably from such approaches as we do not restrict ourselves to corpus-based NLP kernels for similarity, but rather *focus on molecular and structural similarities*. It should be noted, however, that our framework can easily support such NLP-based similarities as it is designed to work with heterogeneous similarity measures; this will be an interesting next step.

Fusing multiple kernels has also been studied as a viable approach for DDI discovery. Chowdhury and Lavelli [16] combined linguistic and NLP kernels for the DDIExtraction2011 challenge. While this work used multiple similarities (kernels), they were all constructed from the same data, making their approach homogeneous. The work of Cheng et al. [17] is closest to our heterogeneous approach; they consider four types of drug-drug similarities (phenotypic, therapeutic, structural and genomic). However, a significant difference from our approach is that they treat pairwise similarities directly as features for use with standard machine-learning models such as SVMs and *k*-nearest neighbor classifiers. This approach destroys the structural and neighborhood information inherent in drug-drug similarity matrices; this means that their model does not capture the true complexity of the DDI manifold. Our method differs from Cheng et al's as we combine heterogeneous similarities *jointly* and (locally) optimally, rather than combining kernels into a single feature set.

It should also be noted that other multiple kernel approaches *do not learn relative weights of similarities*, that is, kernel combination is not a part of the learning process and is performed *a priori* using fixed weights. This is a significant difference, as our approach learns a kernel as well as relative weights between similarities to show which ones have the most influence on the final kernel. Molecular structure similarity analysis has been studied in the context of DDIs before, where Vilar et al [18] used SMILES code and MACCS fingerprints, with a matrix multiplication method thresholded by a Tanimoto coefficient cutoff to predict new DDIs. Similarity-based kernels were also used in the different task of drug-target interactions prediction [19]. The work of Tatonetti et al. [20], Thomas et al. [21] and Percha et al. [22] are also relevant, though they were applied to drug-target interaction prediction.

From a machine-learning standpoint, our work is closely related to multiple kernel learning, which combines the power of multiple kernels together to learn a linear or non-linear kernel combination. The work of Lanckriet et al. [23] optimizes over a linear combination of multiple kernels through semidefinite programming. In this seminal work, Lanckriet et al., test their method on two data sets and demonstrate that learning a combination of kernels is indeed better than learning single kernels for classification. Bach et al. [24] built upon this work and proposed more efficient algorithms for multiple kernel learning. Sonnenburg et al. [25] further generalized the formulation by posing the multiple kernel learning problem as an semi-infinite linear program that is easier to solve. In recent years, the multiple kernel learning has also been extended to multi-class problems [26] and localized kernels [27],

where kernels are learned more precisely using the local information available. These and other methods are discussed by Gönen and Alpaydın [28]. These approaches all rely on the fact that multiple kernel learning can be equivalently cast in terms of the SVM dual; thus these approaches are used for *individual classification* of training examples. Our framework is considerably different, however, as we are interested in *pairwise classification* of training examples to identify interactions.

Our framework, instead, relies on *kernel alignment*, which serves to regularize a kernel learning problem. Specifically, we seek to learn a *single kernel* from multiple similarities by aligning the kernel with the labels [29] as well as local neighborhood [30]. At a high level, our approach seeks to perform manifold regularization [31] and alignment, to fuse information from various similarity measures into one kernel.

3. Drug-Drug Interactions

Before we define the DDI task, we give terse definitions of various entities involved when two or more drugs interact. The *target* or drug target is the protein modified by the drug in order to achieve the desired effect once the drug is administered to the body. *Enzymes* are catalysts that accelerate biological reactions, while *transporters* are proteins that help drugs reach the intended target [32], and also help in determining whether the drug will be absorbed, distributed or eliminated.

DDIs can be either *synergistic* (positive, and help increase the effect of the drugs) or *antagonistic* (negative, cause serious side effects). In this work, we do not differentiate between these two types of interactions. DDIs themselves can be inherently classified into two categories [33]:

- **Pharmacokinetic**: is the effect that a drug goes through when administered, for example, it is absorbed or metabolized. In case of DDIs, pharmacokinetic refers to the (synergystic or antagonistic) effect of one drug on the other drug's absorption, distribution, metabolism and excretion when administered simultaneously or within a short time span of one another.

- **Pharmacodynamic**: is the effect that body goes through when a drug is administered. In case of DDIs, pharmacodynamic refers to the effect of one drug on another drug when they are operating on the same target or even different targets, but with similar behaviour towards the different targets i.e do they inhibit the tendency of the the target to act which can cause an unwanted interaction.

The pharmacokinetic category consists of metabolism interactions like enzyme inhibitors and substrates. Target, enzyme and transporter inhibitors are chemical molecules that bind to the target (or enzyme, or transporter resp.), and inhibit its activity. Enzyme/transporter substrates are molecules which react with the enzyme/transporter, and are converted into different molecules called products. The pharmacodynamic category, on the other hand, occurs due to the agonists and antagonists. An agonist binds to a target, and evokes a response, while an antagonist binds to the target and inhibits a response.

We build our approach based on these two categories with the motivation that if two drugs interact, then there should exist a "path of relationships" describing the molecular and structural properties of the drugs, especially when there is an interaction. Thus, we extract relations as shown in Table 1 from the DrugBank database, whose general schema is shown in figure 1. These relations ensure that we are in the domain of pharmacokinetic and pharmacodynamic categories of the DDIs. Another motivation for using these relations is that the effect of enzymes on DDIs, especially the cytochrome P450, have been well studied extensively in medical literature [34, 35]. Thus, the use of such relations becomes natural in DDI prediction, and can be considered a form of domain expertise.

4. Kernel Learning for Drug Drug Interactions

In classical multiple kernel learning [24, 23], kernels are typically constructed in two different ways. First, multiple kernels can be constructed from the same data source (homogeneous), or from different data sources (heterogeneous). These multiple kernels are then combined in a linear or non-linear fashion. A point to keep in mind in such a multiple-kernel learning setting is the assumption that we have the complete data vectors x_i from which we can construct multiple kernels. Our method diverges considerably from this representation since we never have an explicit representation or embedding of the drugs. Instead, we have several different similarity measures, from which we construct a single kernel for our prediction/discovery task.

4.0.1. Reachability

A key component describing drug-drug interactions is the charaterization of how two drugs react with each other. This is captured using a *directed graph of known chemical reactions* between drugs and enzymes, transporters etc. using ADMET (absorption, distribution, metabolism, excretion and toxicity) features. The idea of *reachability* follows from the intuition that two drugs are likely to interact with one another if one is reachable from the other via one or more paths in an ADMET knowledge graph.

While there exist numerous approaches in graph theory for reachability analysis on graphs [36, 37], our choice is guided by the fact that we operate on multi-relational, directed, relatively sparse graphs involving several thousands of entities/nodes representing drugs, enzymes, targets etc. An iterative search within such a large graph may be intractable. We are inspired by the success of randomized approaches in computational statistics and the seminal work on the path ranking algorithm (PRA, [38]). These approaches show that random walks on a knowledge graphs can be used to generate robust predictive models for relation extraction and reachability analysis. We adapt a similar approach to construct our reachability measure. The estimation of reachability between 2 drugs in a given drug pair proceeds as follows (Figure 2):

(a) Preprocessing—A knowledge graph is constructed for known chemical reactions using ADMET features.

(b) Guided (Parameterized) Random Walk Generation—Parameterized random walks are sequences of relations with shared arguments, where the arguments are entity

classes (not entity instances) starting and ending in the drug entity. Essentially, parameterized random walks are paths in the relational schema of chemical reactions (Figure 1). Similar to PRA [38], our random-walk generation allows for walking against the implicit direction of the relation. Thus, the relations prefixed with represent the inverse of a given relation. An example of a random walk through an ADMET graph looks like:

```
TargetInhibitor(d0, t0)) ∧ _TargetInhibitor(t0, d1)) ∧
TransporterSubstrate(d1, t2)) ∧ _Transporter Inhibitor(t2, d3).
```
We impose certain restrictions on the walks, including disallowing same relation types from following each other (a relation and its inverse are considered different types). We generate several random walks of varying length. Guidance is induced via refining the parameterized walks using domain knowledge [39] (Table 2) that indicate certain types of chemical reactions (or a series), which when present in the walks, increases the likelihood of an interaction between the two drugs at the start and end of the path.

(c) Instantiation—Instantiation of a parameterized walk, $\mathbb{W}$, is the process of finding all possible paths, satisfying $\mathbb{W}$, that exist in the network of chemical reactions $\mathcal{G}$ between two drugs of a given pair$\langle \mathbf{d}_1, \mathbf{d}_2 \rangle$ (Figure 3). If we consider paths as subgraphs, and $\mathbb{W}$ a *motif*, then set of instances $\mathcal{I}_{\langle \mathbf{d}_1, \mathbf{d}_2 \rangle} = \left\{ \forall_g \middle| g \subseteq \mathcal{G}, g \middle| = \mathbb{W}, \mathbf{d}_1 \wedge \mathbf{d}_2 \in g \right\}$. Searching for the set of instances is a combinatorially hard problem (*#P*-complete). We exploit the power of graph databases to compute this. The network of reactions is represented as an RDF[1] graph and the parameterized walks are posed as SPARQL queries [40]. Some example groundings for a couple of given random walks are shown in table 3 and table 4.

(d) Measure/Score generation—The reachability measure is generated for every drug pair $\langle \mathbf{d}_1, \mathbf{d}_2 \rangle$ by obtaining the cardinality (count) of the instance set $\mathcal{I}_{\langle \mathbf{d}_1, \mathbf{d}_2 \rangle}$.

4.0.2. Similarities based on SMILES and SMARTS strings

The simplified molecular-input line-entry system (SMILES) is a commonly-used specification for describing chemical and molecular structure using ASCII strings. The SMILES arbitrary target specification (SMARTS) is an extension of SMILES that is also commonly used for specifying molecular sub-structures precisely. We extract four similarity measures based on molecular and chemical properties of the drug (specified by SMILES and SMARTS strings) using the package rdkit[2]. We compute four similarity measures from SMILES strings [41], which have been previously proven useful in various bio-computing tasks [42, 43, 44]:

(S1) Molecular **Feature Similarity** (FS) compares the chemical properties of two drugs using 19 features extracted from their SMILES strings. These features include the number of valence electrons, number of aromatic rings and number of hydrogen donors and receptors, which are important for determining the reactiveness of a

[1]The Resource Description Framework (RDF) was developed by the WWW Consortium (W3C) for knowledge representation and management on the web.
[2]http://www.rdkit.org/

molecule. We use the Jaccard distance between all features as the similarity between two drugs.

(S2) SMILES **String Similarity** (SS) is the similarity between the SMILES strings themselves, which is calculated using edit distance between the strings.

(S3) **Molecular Fingerprint** (FP) similarity is computed between the fingerprints, which are bit-string representations of the molecular structure.

(S4) **Molecular ACCess System** (MACCS) keys are a particular type of fingerprint generated from SMARTS strings. Similarities on MACCS are commonly used in the drug discovery domain, though they have been proven to be useful on the DDI domain as well [45].

4.1. Notation and Problem Description

Before describing our approach in detail, we formalize our notation. Given a drug database with N drugs, we are interested in discovering whether a pair of drugs $\mathbf{d}_i$ and $\mathbf{d}_j$ **interact** with each other. Recall that we do not distinguish between synergistic and antagonistic interactions. Let all possible drug pairs in the database be $\mathcal{P} = \left\{ (\mathbf{d}_i, \mathbf{d}_j) \mid 1 \le i, j \le N \right\}$, and we use the short-hand notation ij to denote the drug-drug pair $(\mathbf{d}_i, \mathbf{d}_j)$. As mentioned previously, our problem setting is considerably different from the classical multiple kernel learning framework. We do not attempt to construct an explicit vector representation or embedding of a drug $\mathbf{d}_i$. Instead, given N drugs, we construct M pairwise similarity matrices S_m, for $m = 1,$ …, M. As described above, these similarities can be constructed using various drug properties that represent the potential for interactions such as molecular structure, pharmacological attributes etc. Since these "similarities" represent potential for interactions, they can also be constructed from natural language text extracted from such diverse sources as electronic health records [4] and journal abstracts [5].

Our approach seeks to combine different interaction measures and similarities, S_m, from various sources into one coherent kernel. Note that the only requirement on the similarity matrices is that $S_m \in \mathbb{S}^N$, the space of all $N \times N$ *symmetric* matrices. We *do not assume positive semi-definiteness* (psd[3]) of similarity measures; as we show below, it is possible to align a psd kernel with non-psd similarity matrices. Thus, any symmetric scoring function $\sigma_m(\mathbf{d}_i, \mathbf{d}_j)$ can be used to generate a similarity matrix S_m. This allows our approach to be *agnostic to multiple representations* of a drug. For example, σ_1 can be string alignment similarity of the genomic strings of two drugs, while σ_2 can be the bag-of-words co-occurence count of the two drugs in a biomedical corpus. Broadly, any scoring function that measures similarity of a potential for interaction can be considered a candidate similarity measure.

---

[3]A symmetric matrix is positive semi-definite if its eigenvalues are all non-negative (0), and positive definite if its eigenvalues are strictly positive (> 0). Positive semi-definiteness allows us to manipulate kernels instead of explicitly transforming the data into a higher dimensional space.

The (i, j)-th element of S_m is denoted s_m^{ij}, and describes the interaction between $\mathbf{d}_i$ and $\mathbf{d}_j$ according to interaction measure S_m. The interaction label $y_{ij} = +1$ if the drugs $\mathbf{d}_i$ and $\mathbf{d}_j$ interact adversely with each other and $y_{ij} = -1$ otherwise. We denote the matrix of all drug-drug interactions as $Y \in \mathbb{S}^N$, the symmetric matrix whose (i, j)-th entry is the interaction label y_{ij}. Generally, we only know the true labels for a small subset of drug pairs, $\mathcal{L} \subset \mathcal{P}$, and our goal is to learn a model on $\mathcal{L}$ in order to discover drug-drug interactions in the remaining pairs $\mathcal{U} = \mathcal{P}\backslash\mathcal{L}$. Our problem can be formulated as follows:

Given: For N drugs, M interaction similarities S_m, a small subset of *known* interactions y_{ij} for pairs $ij \in \mathcal{L} \subset \mathcal{P}$,

Learn: A kernel $Z \succeq 0$, and interaction similarity combination weights $\alpha_m \geq 0$, $\sum_{m=1}^{M} \alpha_m = 1$,

Predict/Discover: Previously unknown pairwise drug-drug interactions $\hat{y}_{ij} = \text{sign}(z_{ij})$, for pairs $ij \in \mathcal{U} = \mathcal{P}\backslash\mathcal{L}$.

Our novel formulation addresses kernel learning at an element-wise, local and global level, enabling us to learn robust models for discovery of new drug-drug interactions.

4.2. Incorporating Neighborhood Information

We view each interaction/similarity measure as a *graph* that provides a *different view of the neighborhood* of a drug. That is, each similarity matrix S_m represents a fully-connected graph with s_m^{ij} representing the edge weight between drugs $\mathbf{d}_i$ and $\mathbf{d}_j$. Since each S_m measures similarities differently, the neighborhood of a drug $\mathcal{N}_m(\mathbf{d}_i)$ with respect to different S_m will be different. In order to effectively incorporate this multi-view neighborhood information, we construct graph Laplacians L_m, $m = 1, \ldots, M$, for each similarity. Laplacians are naturally locality-preserving [46, 47], that is, they preserve the neighborhood structure in the data. This allows us to learn a kernel that fuses neighborhood information $\mathcal{N}_m$ from multiple interaction types. Without loss of generality, we set the diagonal of S_m to zero: $\texttt{diag}(S_m) = \mathbf{0}$, reflecting that drugs do not interact with themselves. The Laplacian can be constructed as

$$L_m = (1 + \delta)I_N - D^{-\frac{1}{2}}S_m D^{-\frac{1}{2}}, \quad (1)$$

where I_N is an $N \times N$ identity matrix and D is a diagonal matrix with entries $d_{ii} = \sum_{j=1}^{N} s_{ij}^m$ (the row sum of the similarity matrix S_m).

We formulate the following kernel learning problem:

$$\underset{L, Z, \alpha}{\text{minimize}} \quad \overbrace{\langle L, Z \rangle + \langle L, Y \rangle}^{\text{alignment}} + \overbrace{\lambda_1 r(\alpha)}^{\text{regularization}}$$

$$\underbrace{\lambda_2 (\ell_1(Z, Y) + \ell_2(L, Y))}_{\text{loss functions}} \qquad (2)$$

$$\text{subject to } L = \sum_{m=1}^{M} \alpha_m L_m, \alpha \geq 0, \mathbf{e}'\alpha = 1, Z \succeq 0.$$

We highlight the various components of the formulation (2):

The variable $L = \sum_{m=1}^{M} \alpha_m L_m$ is a *convex combination* of the Laplacians L_m arising from the various interaction similarities. The matrix variable L is introduced purely for convenience of notation and can easily be eliminated from the objective function of (2). We select a convex rather than a linear or conic combination in order to *improve interpretability* [28]. That is, positive a_m enable us to intuitively interpret the importance of one similarity relative to the others. The formulation attempts to identify a combination weights $\boldsymbol{a}$ as well as a kernel $Z \succeq 0$, which ensures that Z is psd.

The *alignment* terms are inspired by the success of alignment-based regularization for kernel learning [29]. Hoi et al [30] observe that these alignment terms essentially perform manifold regularization [31], which has the effect of incorporating local neighborhood information encoded in the different Laplacians as well as the labels into learning $\boldsymbol{a}$ and Z. Specifically, $\langle L, Y \rangle$ encourages the weights on the Laplacians a to be consistent with the labels Y. The impact of labels is also propagated into Z by the $\langle L, Z \rangle$ term.

The entries of the learned kernel z_{ij} directly provide a unified interaction score and we predict drug interactions as

$$\hat{y}_{ij} = \text{sign}(z_{ij}). \qquad (3)$$

While the unified kernel Z is positive semi-definite, it's entries can still be negative, which is a fact that we exploit here. Enforcing positive semi-definiteness also naturally imposes symmetry on the learned kernel.

In order that the elements of Z capture interactions effectively into a score, we require a loss function that ensures that the *interaction margin* is maximized. We use the *hinge loss* to ensure that $y_{ij} z_{ij} \geq 1$ holds. Intuitively, these constraints ensure that $z_{ij} \geq 1$ when $y_{ij} = +1$ and $z_{ij} \leq -1$ when $y_{ij} = -1$. The interaction margin behaves very similarly to the margin in SVMs [7]. Thus, we select ℓ_1 to be the hinge loss in (2), which is applied to the drug pairs with known labels (indexed by) $ij \in \mathcal{L}$:

$$\ell_1(Z, Y) = \sum_{ij \in \mathscr{L}} \max(1 - y_{ij}z_{ij}, 0). \quad (4)$$

The loss function ℓ_1 ensures element-wise consistency between the learned kernel Z and the labels Y. In a similar vein, the loss function ℓ_2 aims to propagate this consistency into the combination weights $\boldsymbol{a}$. To this end, we measure the element-wise deviation of the weighted Laplacian with the labels as well, through the Frobenius norm:

$$\ell_2(L, Y) = \frac{1}{2}\|L - Y\|_F^2. \quad (5)$$

Finally, we also add a regularization term over $\boldsymbol{a}$, typically to ensure robustness in weight learning. In this work, we chose the classical L_2 regularizer, $r(\boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\alpha}\|_2^2$. Other norms can also be used, depending on what properties of $\boldsymbol{a}$ are desired. For instance, the L_1 regularizer, $r(\boldsymbol{a}) = \|a\|_1$ encourages sparsity, while the L_∞ regularizer, $r(\boldsymbol{a}) = \|a\|_\infty$ encourages the model to select the single best kernel. We use L_2 regularization here, and defer the exploration of the properties of the other regularizers to future work.

We formulate the following kernel learning problem:

$$\begin{aligned}
\underset{L, Z, \boldsymbol{\alpha}}{\text{minimize}} \quad & \langle L, Z \rangle + \langle L, Y \rangle + \frac{\lambda_1}{2}\|\boldsymbol{\alpha}\|_2^2 \\
& \lambda_2 \sum_{ij \in \mathscr{L}} \xi_{ij} + \frac{\lambda_2}{2}\|L - Y\|_F^2 \\
\text{subject to} \quad & y_{ij}z_{ij} - 1 + \xi_{ij} \geq 0, \xi_{ij} \geq 0, \forall ij \in \mathscr{L}, \\
& L = \sum_{m=1}^{M} \alpha_m L_m, \boldsymbol{\alpha} \geq 0, \mathbf{e}'\boldsymbol{\alpha} = 1, Z \geq 0
\end{aligned} \quad (6)$$

The *slack variables* $\xi_{ij} \geq 0$ measure the hinge loss of the pairwise interaction fit between the labels and the entries of Z as shown in equation (4). These slack variables function in a manner very similar to the slack variables in SVMs: if the prediction z_{ij} and the label y_{ij} have the same sign, then the model correctly identifies the interaction for drugs $\mathbf{d}_i$ and $\mathbf{d}_j$. In this case, we will have, $y_{ij}z_{ij} > 0$ and consequently, $\xi_{ij} = 0$. However, for misidentified interactions, $\xi_{ij} = 1 - y_{ij}z_{ij} > 0$. Thus, by minimizing ξ_{ij}, we are able to minimize the misclassification of drug-drug interactions. The formulation (6) is an instance of a bilinear program, owing to the terms $\langle L, Z \rangle = \sum_{m=1}^{M} \alpha_m \langle L_m, Z_m \rangle$

We solve (6) using *alternating minimization* [48]. At the t-th iteration, we fix the current estimate of the similarity weights $\hat{\alpha}^t$ (note that when $\boldsymbol{a}$ are fixed, this also fixes L, owing to

the equality constraint in eq. 6). This allows us to infer the new interactions scores $\hat{Z}_{i,j}^{t+1}$ by solving the following sub-problem, which we denote SubProbE $(Z|\hat{\alpha}^t)$. This can be interpreted as the expectation step of an EM procedure, where we identify the hidden variables, in this case, the drug-drug interactions Z. We can now fix $Z = \hat{Z}^{t+1}$ in (6), which gives us a sub-problem we denote SubProbM $(\alpha|\hat{Z}^{t+1})$. Again, this step can be considered equivalent to the maximization step of an EM procedure, where we estimate the parameters (here, α, which parameterize the influence of the various similarities on the final kernel). This procedure is summarized in Algorithm 1. Both sub-problems were solved using SDPT3 [49].

Algorithm 1

Alternating Minimization for Learning SKID[3]

1:   $\hat{\alpha}^0 = 1/m \triangleright$ Initialize weights uniformly

2:   $\hat{Z}^0 = I_N \triangleright$ Initialize kernel to identity matrix

3:   **for** t $t_{\max}$ **do**

4:    $\hat{Z}^{t+1} \leftarrow$ SubProbE(Z|$\hat{\alpha}^t$) $\triangleright$ Update Z

5:    $\hat{\alpha}^{t+1} \leftarrow$ SubProbE$\left(\alpha|\hat{Z}^{t+1}\right) \triangleright$ Update α

6:    **If** $\dfrac{1}{N^2}\left\|\hat{Z}^{t+1} - \hat{Z}^t\right\|_F^2 + \dfrac{1}{m}\left\|\hat{\alpha}^{t+1} - \alpha^t\right\|_2^2 \leq \tau_{\text{tol}}$ **then**

7:     **break** $\triangleright$ Converged to tolerance

8:    **end if**

9:    $t \leftarrow t + 1$

10:   **end for**

---

5. Experiments

In this section, we aim to answer the following questions, which address the effectiveness of our proposed approach:

(**Q1**) How effective are the similarity measures on their own for the task of identifying drug-drug interactions?

(**Q2**) Is kernel learning effective for the DDI task?

(**Q3**) Is combining multiple similarity measures more advantageous than using a single similarity measure? How do the learned weights change with increasing database size?

(**Q4**) Does our work motivate further clinical investigations?

(**Q5**) How scalable is our method?

Our data set consists of 78 drugs obtained from DrugBank[4]. This gives rise to 3003 possible interactions[5]. All our reported results were obtained across five runs with a held-out test set of 603 drug pairs. Different methods were trained with increasing number of drug pairs ranging from 400 to 2400, chosen randomly for each run.

The results of our experiments are shown in Figure 5. Note that all the metrices shown in the results are averaged over 5 runs on a hold-out set. Figures 5(a)–5(d) show that a kernels learned from each individual similarity measure (described in Sec. 4.0.1 and 4.0.2) are able to perform reasonably well on the **DDI** prediction task, thereby answering **Q1** affirmatively.

We also learn a single kernel ($Z \equiv \texttt{SKID}^3$) as well as the weights for the five similarity measures (a_m). It is evident that learning from multiple similarity measures provide a more stable learning curve that performs well. Our initial hypothesis was that the similarities generated from molecular structures (SS, FS, FP and MACCS) and chemical reaction pathways (**RW**) fused into a single kernel could combine the advantages of both. That is, our hypothesis was that similarity fusion should achieve the high precision of the molecular structures similarity as well as the high recall of the chemical reaction pathways similarity. The results clearly confirm this, thereby answering **Q2** and **Q3** affirmatively. Figure 5(e) shows the change of the learned weights as the number of training drug pairs increases. A key observation from Figure 5(e) is that the influence of the random walk (RW) similarity decreases, while the weight of the molecular structure similarities increases. This suggests that RW similarities are particularly effective in smaller databases, for targeted identification of interactions.

Q1–**Q3** evaluate the performance of our approaches and *confirm existing interactions* as provided by DrugBank. Our goal with **Q4** was to see if $\texttt{SKID}^3$ is able to *discover new interactions*. In order to answer **Q4**, it is necessary that our analysis goes beyond ground truth that we are considering in constructing the model. Thus, we look closely at the false positives and false negatives, under the intuition that DrugBank (or any other database) is never fully complete or accurate.

In Table 5 we present *a few drug pairs* that are supposedly "incorrectly classified" by our method using the DrugBank ground truth[6]. Table 5 shows that the interactions discovered by our approach can be supported by independent sources or research. Specifically, according to the ground truth, 6 interactions were flagged as false positives. On the contrary, according to literature, these are likely true interactions. Thus, we answer **Q4** affirmatively. This is a crucial observation in the task of drug surveillance: many sources of DDIs need to be carefully and continuously curated for updating this ground truth. This result highlights the fact that $\texttt{SKID}^3$ can indeed not only classify DDIs, but can help in knowledge refinement as well as knowledge discovery. Validating this hypothesis more fully requires large-scale evaluation, which is an interesting direction for future research.

---

[4]https://www.drugbank.ca/

[5]Given *n* drugs, since each drug can interact with every other drug except itself, there will be a total of $\binom{n}{2} = \frac{n(n-1)}{2}$ interactions.

[6]In the previous instance of the Drugbank database download in April 2017, this instance was not present whereas in February 2018, when checked again, this interaction was added. We use the previous instance as ground truth.

Finally, Figure 5(f) shows the time taken by our method. The training time increases linearly with the number of drug pairs, showing the scalability of our method and answering **Q5** affirmatively. This result has practical implications for scalable DDI discovery with full drug databases.

6. Conclusion and Future work

We consider the problem of drug-drug interaction discovery, and develop a framework to exploit deeper structures and drug features using kernel learning. Our extensible framework can fuse information from multiple views including chemical reaction pathways and molecular structure, which we have demonstrated here. Furthermore, our formulation can easily admit other types of interactions as similarities including phenotypic, pharmacological, genomic and text, to name a few.

Our evaluations on the DrugBank database established the superiority of our proposed approach, which is distinct from many current approaches that generally ignore drug properties and instead seek interactions through text mining of existing literature. Extending this work to include more features including other semantic similarity metrics is an interesting direction. Combining the results of learning from DrugBank with other NLP based extraction techniques is another direction. Finally, using other labeling techniques such as weak supervision or distant supervision can potentially lead to larger training sets and can make the discovery process more effective.

Acknowledgments

References

1. Becker ML, et al. Hospitalisations and emergency department visits due to drugdrug interactions: a literature review. Pharmacoepidemiology and Drug Safety. 16

2. Gurwitz JH, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. JAMA.

3. Percha B, Altman RB. Informatics confronts drug-drug interactions. Trends Pharmacol Sci. 2013; 34:178–84. [PubMed: 23414686]

4. Page D, Natarajan S, Costa V SantosPeissig P, Barnard A, Caldwell M. Identifying adverse drug events from multi-relational healthcare data. AAAI; 2012.

5. Odom P, Bangera V, Khot T, Page D, Natarajan S. Extracting adverse drug events from text using human advice. Holmes JH, Bellazzi R, Sacchi L, Peek N, editorsAIME; 2015.

6. Bakir GH, Hofmann T, Schölkopf B, Smola AJ, Taskar B, Vishwanathan SVN. Predicting Structured Data. The MIT Press; 2007.

7. Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. Cambridge Univ. Press; 2004.

8. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. AMIA Annual Symposium Proceedings. 2011

9. Liu X, Chen H. Azdrugminer: an information extraction system for mining patient-reported adverse drug events in online patient forums. ICSH; 2013.

10. Gurulingappa H, Mateen-Rajpu A, Toldo L. Extraction of potential adverse drug events from medical case reports. Journal of biomedical semantics. 2012; 3(1):15. [PubMed: 23256479]

11. Karlsson I, Zhao J, Asker L, Boström H. Predicting adverse drug events by analyzing electronic patient records. AIME; 2013.

12. Kang N, Singh B, Bui C, Afzal Z, van Mulligen EM, Kors JA. Knowledge-based extraction of adverse drug events from biomedical text. BMC bioinformatics.

13. Natarajan S, Bangera V, Khot T, Picado J, Wazalwar A, Costa VS, Page D, Caldwell M. Markov logic networks for adverse drug event extraction from text. KIS.

14. Segura-Bedmar I, Martinez P, de Pablo-Sánchez C. Using a shallow linguistic kernel for drug–drug interaction extraction. Journal of biomedical informatics.

15. Chowdhury MFM, Lavelli A. Fbk-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information. SEM. 2013

16. Chowdhury FM, Lavelli A. Drug-drug interaction extraction using composite kernels. 2011:27–33.

17. Cheng F, Zhao Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. JAMIA.

18. Vilar S, Harpaz R, Uriarte E, Santana L, Rabadan R, Friedman C. Drugdrug interaction through molecular structure similarity analysis. JAMIA.

19. Ding H, Takigawa I, Mamitsuka H, Zhu S. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. Briefings in Bioinformatics.

20. Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. JAMIA.

21. Thomas P, Neves M, Solt I, Tikk D, Leser U. Relation extraction for drug-drug interactions using ensemble learning. Training.

22. Percha B, Garten Y, Altman RB. Discovery and explanation of drug-drug interactions via text mining. PSB. 2012

23. Lanckriet GR, Cristianini N, Bartlett P, Ghaoui LE, Jordan MI. Learning the kernel matrix with semidefinite programming. JMLR.

24. Bach FR, Lanckriet GR, Jordan MI. Multiple kernel learning, conic duality, and the smo algorithm. ICML. 2004

25. Sonnenburg S, Rätsch G, Schäfer C, Schölkopf B. Large scale multiple kernel learning. JMLR.

26. Zien A, Ong CS. Multiclass multiple kernel learning. ICML. 2007

27. Gönen M, Alpaydin E. Localized multiple kernel learning. ICML. 2008

28. Gönen M, Alpaydın E. Multiple kernel learning algorithms. JMLR. 2011; 12:2211–2268.

29. Cortes C, Mohri M, Rostamizadeh A. Algorithms for learning kernels based on centered alignment. JMLR.

30. Hoi SCH, Jin R, Lyu MR. Learning nonparametric kernel matrices from pairwise constraints. ICML. 2007

31. Belkin M, Matveeva I, Niyogi P. Regularization and semi-supervised learning on large graphs. Learning Theory. 2004

32. Nigam SK. What do drug transporters really do? Nature reviews Drug discovery. 2015; 14(1):29. [PubMed: 25475361]

33. August JT, Murad F, Anders M, Coyle JT, Li AP. Drug-Drug interactions: scientific and regulatory perspectives. Vol. 43. Academic Press; 1997.

34. Guengerich FP. Role of cytochrome p450 enzymes in drug-drug interactions. Advances in pharmacology. 1997

35. Ogu CC, Maxa JL. Drug interactions due to cytochrome p450. Baylor University Medical Center Proceedings. 2000

36. Lü L, Zhou T. Link prediction in complex networks: A survey, Physica A: statistical mechanics and its applications.

37. Taskar B, Wong MF, Abbeel P, Koller D. Link prediction in relational data. NIPS. 2004

38. Lao N, Cohen WW. Relational retrieval using a combination of path-constrained random walks. Machine learning. 2010; 81(1):53–67.

39. US Food and Drug Administration. Center for Drug Evaluation and Research. Bethesda, MD: Drug interaction studies-study design, data analysis, implications for dosing, and labeling recommendations.

40. Grobe M. Rdf, jena, sparql and the 'semantic web'. SIGUCCS. 2009

41. Anderson E, Veith GD, Weininger D. SMILES, a line notation and computerized interpreter for chemical structures, US Environmental Protection Agency. Environmental Research Laboratory. 1987

42. Helma C, Cramer T, Kramer S, De Raedt L. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. J Chem Inform Comput Sci.

43. Arimoto R, Prasad MA, Gifford EM. Development of cyp3a4 inhibition models: comparisons of machine-learning techniques and molecular descriptors. Journal of biomolecular screening.

44. Cao DS, Zhao JC, et al. In silico toxicity prediction by support vector machine and smiles representation-based string kernel. SAR and QSAR in Environmental Research. :23.

45. Vilar S, Uriarte E, Santana L, Lorberbaum T, Hripcsak G, Friedman C, Tatonetti NP. Similarity-based modeling in large-scale prediction of drug-drug interactions. Nature protocols.

46. He X, Niyogi P. Locality preserving projections. NIPS. 2003

47. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation.

48. Csiszár I, Tusnády G. Information Geometry and Alternating minimization procedures. Statistics and Decisions.

49. Toh KC, Todd MJ, Tütüncü RH. SDPT3 – A Matlab software package for semidefinite programming, v. 1.3. OMS.

50. Multiple, Drugs.com. https://www.drugs.com/drug-interactions/flomax-with-limbitrol-2146-1397-169-8640.html

51. Nies AT, Hofmann U, Resch C, Schaeffeler E, Rius M, Schwab M. Proton pump inhibitors inhibit metformin uptake by organic cation transporters (octs). PLoS One.

52. Multiple, rxlist.com. https://www.rxlist.com/drug-interactions/glyburide-metformin-oral-and-omeprazole-oral-interaction.htm

53. Thoolen M, Wilfert B, Jonge A, Timmermans P, Zwieten P, et al. Effect of salbutamol and the pde-inhibitor ra 642 on the clonidine withdrawal syndrome in rats. Auton Autacoid Pharmacol.

54. Ali SM, Hadad NS, Jawad AM. Effect of amoxicillin and cefalexin on the pharmacokinetics of diclofenac sodium in healthy volunteers. The Medical J Basrah University.

55. Pavici M, Van Winkelhoff A, De Graaff J. Synergistic effects between amoxicillin, metronidazole, and the hydrox-ymetabolite of metronidazole against actinobacillus actinomycetemcomitans. Antimicrobial agents and chemotherapy.
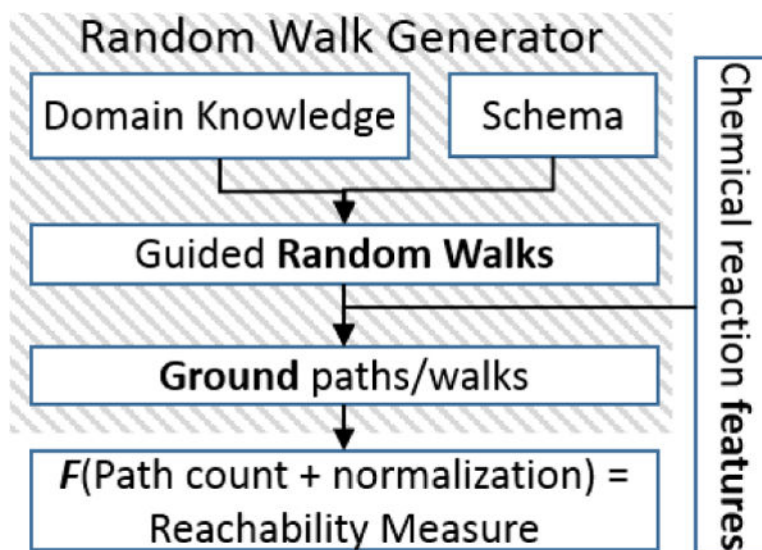
56. Multiple, drugs.com. https://www.drugbank.ca/drugs/DB00333
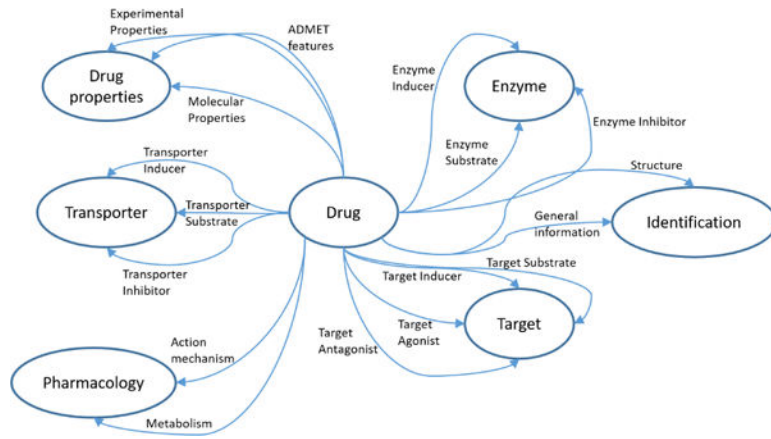
Figure 1.
A general schema representation of the DrugBank database
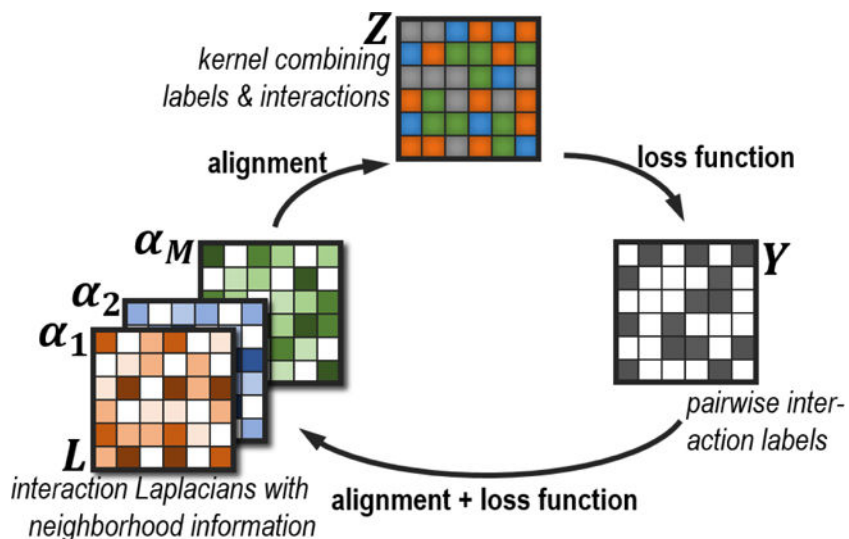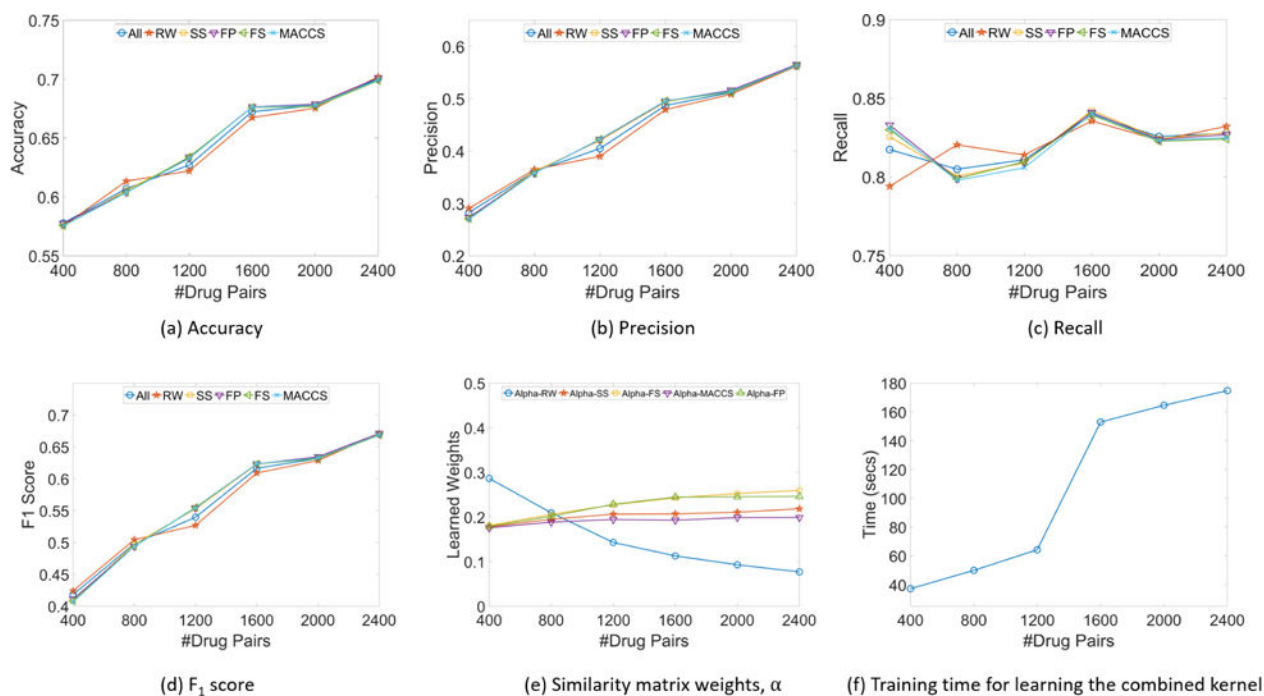
Figure 2.
Reachability measure generation
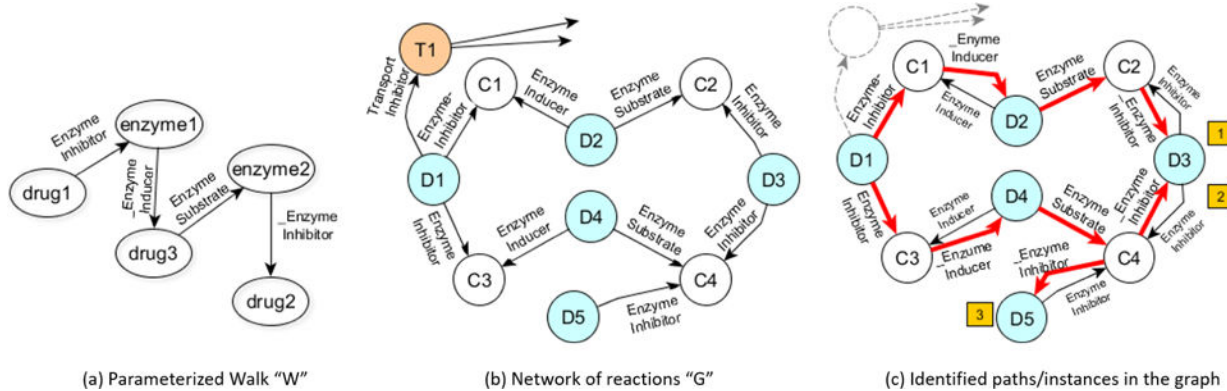
Figure 3.
Instantiation process of a parameterized random walk $\mathbb{W}$ (left) is equivalent to sub-graph matching for a given motif. The graph $\mathcal{G}$ (middle) shows a part of the chemical reaction network (D_X, C_X & T_X indicate drugs, enzymes and transporters resp.). The rightmost figure shows how 3 different instances/paths (marked in red) have been identified that satisfy $\mathbb{W}$.

Figure 4.
Our formulation aims to learn a positive semi-definite kernel Z and combination weights a_m for various similarity measures. These similarities represent interaction scores, which help determine how likely two drugs are to interact. The similarities can be constructed from diverse sources (such as molecular, structural, genomic, text). The similarity measures are expressed through Laplacians, which view the interactions as a *neighborhood graph*. In this manner, we can incorporate local information into the kernel. The loss functions ensure that the learned Z is element-wise consistent with the labels.

(a) Accuracy

(b) Precision

(c) Recall

(d) F_1 score

(e) Similarity matrix weights, α

(f) Training time for learning the combined kernel

Figure 5.
Experimental results, with kernels learned using **All** similarity measures (SKID^3), random-walk reachability (**RW**), SMILES string similarity (**SS**), molecular feature similarity (**FS**), molecular fingerprint similarity (**FP**) and MACCS fingerprint similarity (**MACCS**). (a)–(d) Classification performance of each kernel on a hold-out test set as the size of the training set increases; (e) Weights of each individual kernel in **All**, as training set size increases; (f) Training time for learning the combined kernel (SKID^3) that fuses all the similarity measures, as training set size increases.

Table 1

Initial relations

Initial Relations
Enzyme(enzyme,drug)
Target(target,drug)
Transporter(transporter,drug)
EnzymeSubstrate(drug, enzyme)
EnzymeInhibitor(drug, enzyme)
EnzymeInducer(drug, enzyme)
TargetSubstrate(drug, target)
TargetAntagonist(drug, target)
TargetInducer(drug, target)
TargetInhibitor(drug, target)
TargetAgonist(drug, target)
TransporterSubstrate(drug, transporter)
TransporterInhibitor(drug, transporter)
TransporterInducer(drug, transporter)

Table 2

Domain knowledge

```
        Enzyme Inhibitor(drug, enzyme) ∧ _EnzymeInducer(enzyme, drug)
  Enzyme Inhibitor(drug, enzyme) ∧ _TransporterInhibitor(transporter, drug)
      Enzyme Inhibitor(drug, enzyme) ∧ _EnzymeInhibitor(enzyme, drug)
     Enzyme Inhibitor(drug, enzyme) ∧ _EnzymeInhibitor(enzyme, drug) ∧
                      EnzymeInhibitor(drug, enzyme)
```

Table 3

Few of the groundings generated for the random walk `TargetAgonist (d0, e1) ∧ _TargetAgonist (e1, d1) ∧ EnzymeSubstrated1, e2)(∧ _EnzymeInhibitor (e2, d2)`

```
   Tramadol,Mu-type opioid receptor,Morphine,Cytochrome P450 2C8,Pravastatin
    Morphine,Mu-type opioid receptor,Tramadol,Cytochrome P450 3A4,Tadalafil
 Hydromorphone,Mu-type opioid receptor,Morphine,Cytochrome P450 2D6,Amlodipine
  Methadone,Mu-type opioid receptor,Oxycodone,Cytochrome P450 3A4,Risperidone
Oxycodone,Mu-type opioid receptor,Hydromorphone,Cytochrome P450 2D6,Risperidone
```

Table 4

Few of the groundings generated for the random walk `TransporterSubstrate(d0, t1)`^

`Transporter(t1, d1)`^ `TargetInhibitor(d1, e1)`^ `Enzyme(e1, d2)`

```
Pravastatin,Multidrug resistance protein 1,Acetaminophen,H synthase
               1,Hydromorphone

Metoprolol,Multidrug resistance protein 1,Diclofenac,H synthase 2,Ibuprofen

Venlafaxine,Multidrug resistance protein 1,Acetylsalicylic acid,H synthase
               1,Diphenhydramine

Cephalexin,Solute carrier family 22 member 6,Naproxen,H synthase 1,Zolpidem

Levothyroxine,Solute carrier organic anion transporter family member
           1C1,Diclofenac,H synthase 1,Hydromorphone
```

Table 5

Table depicting example drug pairs where the prediction does not match the ground truth. However, we additionally cite sources (last column) that support our prediction.

Drug 1	Drug 2	DrugBank Ground Truth	Predicted Class	Independent Source
Amitriptyline	Tamsulosin	Not interacting	Interacting	Drugs.com [50]
Omeprazole	Metformin	Not interacting	Interacting	Nies et al. [51], rxlist.com [52]
Salbutamol	Clonidine	Not interacting	Interacting	Thoolen et al. [53]
Cephalexin	Diclofenac	Not interacting	Interacting	Ali et al. [54]
Amoxicillin	Metronidazole	Not interacting	Interacting	Pavicic et al. [55]
Amphetamine	Salbutamol	Not interacting	Interacting	DrugBank[6]
Cephalexin	Methadone	Interacting	Not interacting	Drugs.com [56]