Differential Compound Prioritization via Bi-Directional Selectivity Push with Power

Junfeng Liu† and Xia Ning*,†,‡

†*Indiana University-Purdue University, Indianapolis*

723 West Michigan St, SL 280, Indianapolis, IN 46202, USA

‡*Center of Computational Biology and Bioinformatics, Indiana University School of Medicine*

410 West 10th St, HITS 5000, Indianapolis, IN 46202, USA

E-mail: xning@iupui.edu

Abstract

Effective *in silico* compound prioritization is a critical step to identify promising drug candidates in the early stages of drug discovery. Current computational methods for compound prioritization usually focus on ranking the compounds based on one property, typically activity, with respect to a single target. However, compound selectivity is also a key property which should be deliberated simultaneously so as to minimize the likelihood of undesired side effects of future drugs. In this paper, we present a novel machine-learning based differential compound prioritization method dCPPP. This dCPPP method learns compound prioritization models that rank active compounds well, and meanwhile, preferably rank selective compounds higher via a bi-directional selectivity push strategy. The bi-directional push is enhanced by push powers that are determined by ranking difference of selective compounds over multiple bioassays. Our experiments demonstrate that the new method dCPPP achieves significant improvement on prioritizing selective compounds over baseline models.

This is the author's manuscript of the article published in final edited form as:
Liu, J., & Ning, X. (2017). Differential Compound Prioritization via Bi-Directional Selectivity Push with Power. Journal of Chemical Information and Modeling. https://doi.org/10.1021/acs.jcim.7b00552

1 Introduction

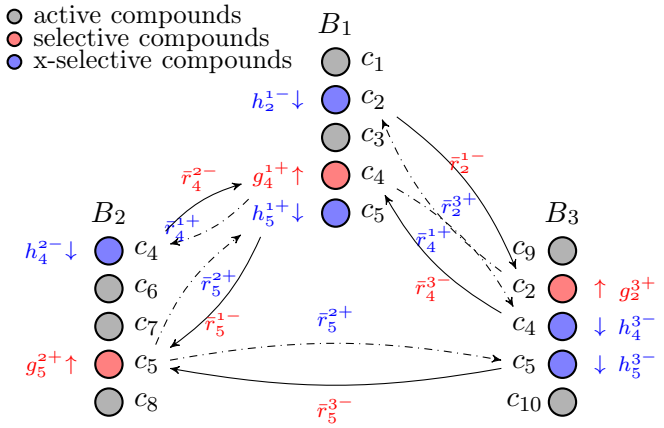
Drug discovery is time-consuming and costly: it approximately takes at least 10 to 15 years and $500 million to $2 billion to fully develop a new drug.[1] To accelerate this process, *in silico* methods[2] have been extensively developed as alternatives, particularly for identifying promising drug candidates in the early stages of drug discovery. *In silico* compound prioritization, which learns computational models to rank compounds in terms of their drug-like/disease-specific properties (e.g., efficacy, specificity), has been recently attracting increasing attention, due to the emerging precision medicine.[3] In many applications of precision medicine (e.g., cancer drug selection[4]), before precise measurements of disease-specific compound properties need to be considered, a set of promising compounds (typically drugs) should be first selected for future investigation. In this paper, we tackle the problem of differential compound prioritization for better ranking selective compounds for drug candidate selection.

Current compound prioritization typically focuses on one single compound property,[5] for example, biological activity. Biological activity of a compound can be initially tested in a target-specific bioassay* by measuring whether the compound binds with high affinity to the protein target that it is aimed to affect. Activity is a critical property that a compound needs to exhibit in order to act efficaciously as a successful drug. Compound prioritization in terms of activity needs to rank most active compounds on top of less active compounds.

Compound selectivity is another key property that successful drugs need to exhibit.[6] Selectivity measures how a compound can differentially bind to only the target of interest with high affinity (i.e., high activity) while bind to other proteins with low affinities. Therefore, the compound selectivity prioritization needs to consider the prioritization difference of a compound in the activity prioritization structures of multiple targets. Specifically, the compound selectivity prioritization needs to follow a combinatorial ranking criterion that 1). it ranks all the compounds well based on their activities; and meanwhile, 2). it ranks

*https://en.wikipedia.org/wiki/Bioassay

strongly selective compounds preferably higher, probably even higher than more active compounds that are not selective. These criterion correspond to that in real applications, active and highly selective compounds are preferred over highly active but also highly promiscuous compounds[7] to minimize the likelihood of undesirable side effects.



B/c represents a bioassay/a compound.

g_i^{k+} right next to a compound c_i represents the push-up (indicated by $\uparrow$) power on c_i as a selective compound in B_k. The solid arrowed lines represent that the push-up power on a selective compound (pointed by the arrows, e.g., c_5 in B_2) is determined by the ranking position of the compound in a different bioassay (pointed by the line ends, e.g., c_5 in B_1). The lines are annotated with such ranking positions (e.g., $\bar{r}_5^{1-}$ on the solid arrowed line from c_5 in B_1 to c_5 in B_2).

h_i^{k-} right next to a compound c_i represents the push-down (indicated by $\downarrow$) power on c_i as an x-selective compound in B_k. The dashed arrowed lines represent that the push-down power on an x-selective compound (pointed by the arrows, e.g., c_5 in B_1) is determined by the ranking position of the compound in a different bioassay (pointed by the line ends, e.g., c_5 in B_2). The lines are annotated with such ranking positions (e.g., $\bar{r}_5^{2+}$ on the dashed arrowed line from c_5 in B_2 to c_5 in B_1).

Figure 1: Overall scheme of dCPPP

In this paper, we present an innovative machine learning method to conduct *in silico* compound prioritization that is able to achieve both the above goals, with a particular focus on better prioritizing selective compounds. This method consists of three components:

1. A compound scoring function, which produces a score for each compound in a bioassay that will be used to rank the compound in the bioassay. The scoring function uses bioassay-specific compound features to calculate the scores.

2. An activity ranking model, which learns the compound scoring function and approxi-

mates the ranking structure among all compounds in a bioassay. The learning is via minimizing the pairwise ordering errors introduced by the scoring function.

3. A bi-directional selectivity push strategy, which preferably pushes up selective compounds in the activity ranking model of a bioassay, and pushes down the compounds in the model that are selective in a different bioassay. The bi-directional push strategy leverages the ranking difference of selective compounds across multiple bioassays and alters the activity ranking by pushing selectivity-related compounds in two directions with specific powers.

These three components will be learned simultaneously within one optimization formulation. This *d*ifferential *C*ompound *P*rioritization via bi-directional selectivity *P*ush with *P*ower method is denoted as dCPPP. Figure 1 presents the overall scheme of dCPPP. To the best of our knowledge, this is the first work in which the activity and selectivity are both tackled within one differential prioritization model that integrates multiple bioassays simultaneously.

The rest of the paper is organized as follows. Section 2 presents the related work to the new method. Section 3 presents the definitions and notations used in the paper. Section 4 presents the new method of activity-selectivity differential ranking with bi-directional powered push. Section 5 presents the materials used for experimental evaluation. Section 7 presents the experimental results. Section 8 and 6 present the discussions and conclusions, respectively.

2 Related Work

2.1 *In Silico* Methods for Drug Discovery

A first step in drug discovery is to conduct bioassays that screen a large set of promising compounds. The outcomes from these bioassays inform the following drug discovery steps.[1] Significant amount of research efforts in knowledge discovery from bioassay data is on estab-

lishing the relationship between the structures of chemical compounds and their bio-chemical properties, for example, Structure-Activity Relationship (SAR)[2] and Structure-Selectivity Relationship (SSR),[8] expressed in the bioassays.

Classification and regression dominate the *in silico* machine learning methods in bioassay analysis, particularly in finding SAR and SSR. In these methods, compounds are typically represented by certain chemical fingerprints, for example, Extended Connectivity Fingerprints (ECFP)[†] and Maccs keys [‡]. Compound activity and selectivity are used as a label/numerical target of the compounds. Popular classification and regression methods include Support Vector Machines (SVM),[9] Partial Least-Squares,[10] random forests,[11] Bayesian matrix factorization,[12] and Naïve Bayesian classifiers,[13] etc. Ranking methods, compared to classification and regression, are less developed for bioassay analysis.

2.2 Structure-Activity-Relationship Modeling

A recent trend in SAR modeling is through leveraging the information from multiple bioassays. A class of methods along this line identifies multiple bioassays and leverages information therefrom to improve SAR qualities. In Ning *et al.*,[14] the SAR classification methods first identify a set of targets related to the target of interest, and then employ various machine learning techniques (e.g., semi-supervised learning,[15] multi-task learning,[16] and classifier ensemble[17]) to utilize activity information from these targets for a better SAR model. In Liu and Ning,[18] compound activity ranking models are developed by leveraging multiple bioassays. In these methods, assistance bioassays and assistance compounds are identified and incorporated to build models that can accurately prioritize active compounds in a bioassay.

A different class of methods is via the multi-assay based "affinity fingerprints". In the Target-Related Affinity Profiling (TRAP) method,[19] the affinity profiles of compounds against a set of diverse bioassays are used as the fingerprints of the compounds. In Bender *et al.*,[20] Bayes scores produced from empirical Bayesian SAR models over a set of targets

[†]Scitegic Inc, http://www.scitegic.com.
[‡]Accelrys, http://accelrys.com

are used as the affinity fingerprints for compounds. Similarly, Lessel *et al.*[21] use the docking scores of compounds against a set of reference binding sides as compound fingerprints. All these existing methods that utilize multiple bioassays in SAR use them homogeneously and cannot utilize the differential signals therein effectively.

2.3 Structure-Selectivity-Relationship Modeling

Existing SSR methods include multi-step classification based approaches,[22] in which compounds that are classified as active are further classified by a selectivity classifier; multi-class classification based approaches,[23] in which compound activity and selectivity are considered as two classes in a common multi-class classifier; compound similarity based approaches,[24] in which compounds that are similar to known selective compounds are considered as selective; etc. A unique thread of research on SSR is using multi-task learning to learn compound activity and selectivity simultaneously.[25] The multi-task method incorporates both activity and selectivity models into one multi-task model to better differentiate compound activity and selectivity. Unfortunately, these existing methods cannot produce activity prioritization and selectivity prioritization simultaneously, or cannot leverage the prioritization structures among multiple bioassays to improve SSR modeling.

2.4 Learning to Rank

Learning to rank (LETOR)[26] focuses on developing ranking models via learning. It has achieved tremendous success in Information Retrieval (IR). Existing LETOR methods fall into three categories: 1). pointwise methods,[27] which learn individual scores that are used later for sorting; 2). pairwise methods,[28] which model pairwise ranking relations; and 3). listwise methods,[29] which model the full combinatorial structures of ranking lists. A recent focus on LETOR is to improve the ranking performance on top of the ranking lists.[30,31]

The idea of using LETOR approaches to prioritize compounds has also drawn some attention recently. For example, Agarwal *et al.*[32] developed bipartite ranking to rank chemical

structures such that active compounds and inactive compounds are well separated in the ranking lists. Jorissen *et al.*[33] used pointwise methods within SVMs to rank compounds in a bioassay to detect active compounds and perform similarity search, respectively. Liu and Ning[18] used SVMRank[34] to build compound activity prioritization models. However, LETOR for compound selectivity prioritization is less developed compared to its use for compound activity prioritization.

3 Definitions and Notations

Table 1: Notations

notations	meanings
$c/B/t$	compound/bioassay/target
c_i^{k+}/c_i^{k-}	selective/non-selective compound c_i in B_k
C_k	the set of compounds in B_k
S_k	the set of selective commpounds in B_k ($S_k = \{c_i^{k+}\}$)
A_k	the set of non-selective commpounds in B_k ($A_k = C_k \setminus S_k$)
S_k^x	the set of x-selective commpounds in B_k ($S_k^x = \{c_i^{k-} \vert \exists B_l, c_i^{k-} \in S_l\}$)
$s_i^k/s_i^{k+}/s_i^{k-}$	score of $c_i/c_i^{k+}/c_i^{k-}$ in B_k
$r_i^k/r_i^{k+}/r_i^{k-}$	percentile ranking of $c_i/c_i^{k+}/c_i^{k-}$ in B_k
p_i^k	ranking position of c_i in B_k
R_i^{k+}/H_j^{k-}	reverse height of c_i^{k+}/ height of c_j^{k-}
g_i^{k+}/h_j^{k-}	push-up power for $c_i^{k+} \in S_k$/push-down power for $c_j^{k-} \in S_k^x$

A compound c is active in a bioassay B with protein target t if the IC_{50} value (i.e., the concentration of the compound that is required for 50% inhibition of the target under consideration; lower IC_{50} value indicates higher activity[§]) of c for t is less than 1 μM. A compound c is selective in a bioassay B with protein target t if the following two conditions are satisfied:[25]

1. c is active for t (i.e., $\mathrm{IC}_{50}(c,t) < 1\mu M$); and
2. $\min\limits_{\forall t_k \neq t} \dfrac{\mathrm{IC}_{50}(c,t_k)}{\mathrm{IC}_{50}(c,t)} \geq 50$,

[§]http://www.ncgc.nih.gov/guidance/section3.html

that is, c needs to be active for t, and its activity on t is at least 50-fold higher than its activity on any other targets.

In this paper, each of the bioassays that are used for model training has only one single protein target. Thus, activity/selectivity with respect to bioassays and with respect to targets will be used interchangeably. When a compound is indicated as selective, by default it is with respect to one certain bioassay/protein target, and the bioassay/protein target is neglected when no ambiguity is raised. A compound can be selective in at most one bioassay. A compound c_i that is selective in a bioassay B_k is denoted as c_i^{k+}. A compound c_i that is not selective in a bioassay B_k (either active and not selective, or inactive in B_k) is referred to as non-selective in B_k and denoted as c_i^{k-}. A compound that is non-selective in a bioassay B_k but selective in another bioassay B_l is referred to as x-selective in B_k. The set of compounds in B_k is denoted as C_k. The set of selective compounds in B_k is denoted as S_k. The set of non-selective compounds in B_k is denoted as A_k. The set of x-selective compounds in B_k is denoted as S_k^x. Table 1 lists the notations that are used in this paper.

4 Methods

4.1 Compound Scoring

In dCPPP, the compound prioritization among a bioassay uses a linear scoring function as in Equation 1,

$$\tilde{s}_i^k = \boldsymbol{w}_k^\mathsf{T} \boldsymbol{x}_i, \tag{1}$$

where $\boldsymbol{w}_k$ is a weighting vector for bioassay B_k, $\boldsymbol{x}_i$ is the feature vector of the compound c_i, and $\tilde{s}_i^k$ is the score of compound c_i in B_k. Each compound in a bioassay is first scored using their features, and the compounds which have larger scores will be ranked higher in the bioassay. The weighting vector $\boldsymbol{w}_k$ will be learned for each bioassay B_k.

4.2 Activity Prioritization

The dCPPP method will produce a ranking of compounds in a bioassay that ranks compounds well based on their activities. That is, in general, compounds that are more active will be ranked higher than those that are less active. To quantitatively measure the activity ranking quality, we use a metric non-Concordance Index (denoted as `nCI`) as follows,

$$\texttt{nCI}(\{\tilde{s}_i^k\}, C_k) = \frac{1}{|\mathcal{P}_k|} \sum_{(c_i \succ c_j) \in \mathcal{P}_k} \mathbb{I}(\tilde{s}_i^k \leq \tilde{s}_j^k), \tag{2}$$

where $\mathcal{P}_k = \{c_i \succ c_j | c_i, c_j \in B_k\}$ is the set of all possible ordered compound pairs in B_k, $\mathbb{I}(\cdot)$ is the indicator function:

$$\mathbb{I}(x) = \begin{cases} 1, & \text{if } x \text{ is true,} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

In Equation 2, $c_i \succ c_j$ indicates that c_i is ranked higher than c_j in ground truth in B_k based on their IC_{50} values, $\tilde{s}_i^k \leq \tilde{s}_j^k$ indicates that compound c_i is predicted as being ranked lower than c_j (i.e., c_i's predicted score $\tilde{s}_i^k$ is smaller than that of c_j; dCPPP ranks compounds with larger scores higher).

Essentially, `nCI` represents the fraction of mis-ordered compound pairs by a certain compound ranking method. A lower `nCI` value indicates better ranking performance. Therefore, activity prioritization seeks a scoring function that can produce lower `nCI`, and thus we use `nCI` over the predicted ranking in B_k as the loss (denoted as $\mathcal{L}_c^k$) for activity prioritization in the dCPPP objective, that is,

$$\mathcal{L}_c^k = \texttt{nCI}(\{\tilde{s}_i^k\}, C_k). \tag{4}$$

4.3 Bi-directional Selectivity Push with Power

To favor selective compounds in compound prioritization, two key questions need to be addressed: 1). how to enforce the selective compounds to go beyond the ranking structures

of ordinary activity prioritization and get better ranked; and 2). how much the enforcement should be and how to decide that. To address the first question, we develop the bi-directional powered push scheme, which, for a target t, pushes t's selective compounds higher, and pushes t's x-selective compounds lower in compound ranking. To address the second question, we develop a scheme to determine push powers by comparing ranking difference of a compound in multiple bioassays.

4.3.1 Pushing up Selective Compounds

To push up selective compounds, dCPPP measures the ranking positions of selective compounds and optimizes such positions. Specifically, the reverse height of a selective compound[32] is used to quantitatively represent such ranking positions.

Reverse height of a selective compound is the number of non-selective compounds that are ranked higher than the selective compound, that is,

$$R_i^{k+} = R(c_i^{k+}) = \sum_{c_j \in A_k} \mathbb{I}(\tilde{s}_i^{k+} \leq \tilde{s}_j^{k-}), \tag{5}$$

where R_i^{k+} is the reverse height of selective compound c_i^{k+} in B_k, A_k is the set of non-selective compounds in B_k, and $\mathbb{I}(\cdot)$ is the indicator function (Equation 3). Thus, to enforce higher ranking of selective compounds, it is to minimize the reverse heights of the selective compounds. In Equation 5, the predicted ranking scores are used to indicate that the reverse height of a selective compound is produced from a ranking model.

Push-up power for a selective compound decides how strongly a selective compound c_i^{k+} should be pushed up in B_k, which depends on 1). how c_i is ranked in B_k; and 2). how c_i is ranked in other bioassays B_l's which c_i is also involved in. Intuitively, if c_i is ranked higher in B_l (i.e., c_i is very active to t_l but not selective to t_l), c_i should be pushed much higher in B_k and much lower in B_l. This is because c_i is very specific to t_k, and if c_i is selected for B_l (t_l), it will introduce low efficacy or side effects.

Based on the above intuition, we define the push-up power for a selective compound c_i^{k+}:

$$\begin{aligned} g_i^{k+} &= g(c_i^{k+}, B_k, \{B_l\}|\theta^{\uparrow}, \xi^{\uparrow}) \\ &= \exp\{\theta^{\uparrow}[(1-\bar{r}_i^{k+}) + \max_{c_i \in A_l} \phi(\bar{r}_i^{l-}, \bar{r}_i^{k+}|\xi^{\uparrow})]\}, \end{aligned} \tag{6}$$

where $\theta^{\uparrow}$ is a parameter, and $\phi(x, y|\xi)$ is a thresholding function:

$$\phi(x, y|\xi) = (x - y + \xi)_+ = \begin{cases} x - y + \xi, & \text{if } x - y + \xi \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

In Equation 6, $\bar{r}_i^{k+}$ is the predicted percentile ranking of c_i from B_k's baseline activity prioritization model, $\bar{r}_i^{l-}$ is the predicted percentile ranking of c_i from B_l's baseline activity prioritization model, and $\xi^{\uparrow}$ is a thresholding parameter. Essentially, the push-up power in Equation 6 considers whether c_i^{k+} has been ranked high enough in B_k (i.e., $1 - \bar{r}_i^{k+}$) and how differentially it is ranked in other bioassays (i.e., $\phi(\bar{r}_i^{l-}, \bar{r}_i^{k+}|\xi^{\uparrow})$). If the ranking positions of c_i^{k+} in B_k and other bioassays are not sufficiently different, the push-up power is exponentially large.

Selectivity Loss with Powered Push-up To differentially push selective compounds up, we take the average reverse heights of selective compounds enhanced by respective push-up powers in the dCPPP learning objective, that is, the push-up loss $\mathcal{L}_s^{k+}$ is defined as

$$\mathcal{L}_s^{k+} = \frac{1}{|S_k|} \sum_{c_i \in S_k} R_i^{k+} \cdot g_i^{k+}, \tag{8}$$

where S_k is the set of selective compounds in B_k, $|S_k|$ is the size of S_k.

4.3.2 Pushing down x-Selective Compounds

To push down x-selective compounds, dCPPP measures the ranking positions of such x-selective compounds and optimize such positions. Specifically, the height [32] of an x-selective compound is used to quantitatively measure its ranking position.

Height of an x-selective compound is the number of compounds that are ranked below the x-selective compound c_j^{k-} (i.e., c_j is non-selective in B_k but selective in a different bioassay), that is,

$$H_j^{k-} = H(c_j^{k-}) = \sum_{c_i \in C_k} \mathbb{I}(\tilde{s}_i^k \leq \tilde{s}_j^{k-}) \tag{9}$$

where H_j^{k-} is the height of x-selective compound c_j^{k-} in B_k, C_k is the set of compounds in B_k, $\mathbb{I}(\cdot)$ is the indicator function (Equation 3).

Push-down power for an x-selective compound determines how strongly the x-selective compound should be pushed down in a bioassay. We define the push-down power for an x-selective compound in bioassay B_k as follows,

$$\begin{aligned} h_j^{k-} &= h(c_j^{k-}, B_k, B_l | \theta^{\downarrow}, \xi^{\downarrow}) \\ &= \exp\{\theta^{\downarrow}[\bar{r}_j^{k-} + \phi(\bar{r}_j^{k-}, \bar{r}_j^{l+} | \xi^{\downarrow})]\} \end{aligned} \tag{10}$$

where $\theta^{\downarrow}$ is a parameter, $\bar{r}_j^{k-}$ is the predicted percentile ranking of c_j in B_k's baseline activity prioritization model, $\bar{r}_j^{l+}$ is the predicted percentile ranking of c_j in B_l ($c_j \in S_l$) from B_l's baseline activity prioritization model, $\phi(\bar{r}_j^{k-}, \bar{r}_j^{l+} | \xi^{\downarrow})$ is thresholding function as defined in Equation 7, and $\xi^{\downarrow}$ is the thresholding parameter. Thus, the push-down power h_j^{k-} considers the difference of percentile rankings of c_j in B_k ($c_j \in S_k^x$) and B_l ($c_j \in S_l$). If $\bar{r}_j^{l+}$ is not significantly higher than $\bar{r}_j^{k-}$, the push-down power is large. Please note that a compound can appear in multiple bioassays, but can be selective in only one bioassay. Therefore, we only consider the bioassay B_l in which c_j is selective when we push down c_j in B_k.

x-Selectivity Loss with Powered Push-down To differentially push x-selective compounds down, we take the average heights of x-selective compounds enhanced by their push-down powers in the dCPPP learning objective, that is, the push-down loss $\mathcal{L}_x^{k-}$ is defined as

$$\mathcal{L}_x^{k-} = \frac{1}{|S_k^x|} \sum_{c_j \in S_k^x} H_j^{k-} \cdot h_j^{k-}. \tag{11}$$

4.4 dCPPP Optimization Problem and Solutions

The overall optimization problem of dCPPP to learn a selectivity prioritization model (i.e., the scoring function as in Equation 1, parameterized by $\boldsymbol{w_k}$), which ranks selective compounds higher and x-selective compounds lower, is formulated as follows,

$$\min_{\boldsymbol{w_k}} \mathcal{L}^k = (1-\alpha-\beta)\mathcal{L}_c^k + \alpha\mathcal{L}_s^{k+} + \beta\mathcal{L}_x^{k-}, \tag{12}$$

where α and β are two weighting parameters ($\alpha \in [0,1], \beta \in [0,1], \alpha+\beta \in [0,1]$). Thus, the dCPPP objective is a weighted combination of the loss on activity prioritization ($\mathcal{L}_c^k$), the loss on pushing up selective compounds ($\mathcal{L}_s^{k+}$), and the loss on pushing down x-selective compounds ($\mathcal{L}_x^{k-}$).

Since the indicator function in Equation 3 is not continuous or smooth, we use the logistic loss as the surrogate function:[35]

$$\mathbb{I}(x \leq y) \approx \log[1+\exp(-(x-y))] = -\log\sigma(x-y), \tag{13}$$

where $\sigma(x)$ is a sigmoid function:

$$\sigma(x) = \frac{1}{1+\exp(-x)}. \tag{14}$$

The surrogate function is continuous, smooth and differentiable. Thus, the loss $\mathcal{L}^k$ in Equation 12 with the surrogate function is differentiable, and thus we can use gradient descent[36]

to solve the optimization problem.

4.4.1 Gradient of Powered Push

The gradient of the loss function in Equation 12 is composed of the gradients on the loss of compound ranking, the loss on push-up and the loss on push-down, that is,

$$\nabla_{\boldsymbol{w_k}}\mathcal{L}^k = (1-\alpha-\beta)\nabla_{\boldsymbol{w_k}}\mathcal{L}_c^k + \alpha\nabla_{\boldsymbol{w_k}}\mathcal{L}_s^{k+} + \beta\nabla_{\boldsymbol{w_k}}\mathcal{L}_x^{k-}, \tag{15}$$

where

$$\nabla_{\boldsymbol{w_k}}\mathcal{L}_c^k = \frac{1}{|\{s_i^k > s_j^k\}|}\sum_{\{s_i^k > s_j^k\}} \nabla_{\boldsymbol{w_k}}\mathbb{I}(\tilde{s}_i^k \leq \tilde{s}_j^k), \tag{16}$$

$$\begin{aligned}\nabla_{\boldsymbol{w_k}}\mathcal{L}_s^{k+} &= \frac{1}{|S_k|}\sum_{c_i^{k+}\in S_k}\{g_i^{k+}\cdot\nabla_{\boldsymbol{w_k}}R_i^{k+}\} \\ &= \frac{1}{|S_k|}\sum_{c_i^{k+}\in S_k}\{g_i^{k+}\cdot\sum_{c_j\in A_k}\nabla_{\boldsymbol{w_k}}\mathbb{I}(\tilde{s}_i^{k+} \leq \tilde{s}_j^{k-})\},\end{aligned} \tag{17}$$

and

$$\begin{aligned}\nabla_{\boldsymbol{w_k}}\mathcal{L}_x^{k-} &= \frac{1}{|S_k^x|}\sum_{c_j^{k-}\in S_k^x}\{h_j^{k-}\cdot\nabla_{\boldsymbol{w_k}}H_j^{k-}\} \\ &= \frac{1}{|S_k^x|}\sum_{c_j^{k-}\in S_k^x}\{h_j^{k-}\cdot\sum_{c_i\in C_k}\nabla_{\boldsymbol{w_k}}\mathbb{I}(\tilde{s}_i^{k} \leq \tilde{s}_j^{k-})\}.\end{aligned} \tag{18}$$

In Equation 16 to Equation 18, $\nabla_{\boldsymbol{w_k}}\mathbb{I}(\tilde{s}_i^k \leq \tilde{s}_j^k)$ will be approximated by the gradient over logistic loss function (Equation 13). The variable $\boldsymbol{w_k}$ is updated via the following rule:

$$\boldsymbol{w}_k \leftarrow \boldsymbol{w}_k - \lambda\nabla_{\boldsymbol{w_k}}\mathcal{L}^k \tag{19}$$

where λ is the learning rate.

Algorithm 1: Iterative Optimization for dCPPP

Input: a set of training bioassays $\{B_k\}$;
parameters α, $\xi^{\uparrow}$, $\theta^{\uparrow}$, β, $\xi^{\downarrow}$, $\theta^{\downarrow}$;
learning rate λ; max number of iterations $niters$
Output: ranking models $\{\text{dCPPP}_k^*\}$.
for $t = 1, \cdots, niters$ **do**
  for *each bioassay* B_k **do**
    if $t == 1$ **then**
      $\text{dCPPP}_k^{\circ}(t) = \text{dCPPP}_k^{\circ}$
    else
      $\text{dCPPP}_k^{\circ}(t) = \text{dCPPP}_k^{*}(t-1)$
    end
    while *not converged* **do**
      Update $\text{dCPPP}_k^{*}(t)$ upon $\text{dCPPP}_k^{\circ}(t)$ using gradient descent (Equation 19)
    end
  end
end
return $\{\text{dCPPP}_k^*\}$

4.5 System Equilibrium from Powered Push

It is possible that after one iteration of the powered push among all related bioassays, the ranking models are still not optimal due to the change of ranking structures of other updated models. Thus, multiple iterations of systematically powered push should be conducted until an equilibrium is achieved among all the bioassays. When multiple iterations of dCPPP pushes are conducted, the optimal model from the previous iteration serves as the baseline model for the next iteration.

The initial baseline model for the first iteration corresponds to dCPPP at $(\alpha = 0, \beta = 0)$, that is, the standard ranking model without any push. This baseline model is denoted as dCPPP°. If each bioassay uses its own optimal α and β values (i.e., the α and β value that together give the optimal performance for each bioassay), the corresponding optimal model is denoted as dCPPP^{*}. Thus, dCPPP^{*} from the previous iteration is the baseline for the next iteration. The models trained in the t-th iteration are denoted by having (t) (e.g., $\text{dCPPP}^{*}(1)$, $\text{dCPPP}^{\circ}(2)$) and thus $\text{dCPPP}^{*}(t-1) = \text{dCPPP}^{\circ}(t)$. Algorithm 1 presents the overall iterative algorithm for dCPPP optimization.

5 Materials

In this section, we present the details on dataset generation, experimental protocol and evaluation metrics. All the datasets and source code are available online and on our website[¶]

5.1 Dataset Generation

The dataset for the experimental evaluation is very critical, and therefore we present the dataset construction in detail here. We constructed a set of bioassays from ChEMBL[‖] in accordance with the protocols in Section 5.1.1 and Section 5.1.2 in order to 1). have a sufficiently large number of bioassays to study; and 2). have a sufficiently large number of active and selective compounds in each bioassay to reliably learn models.

5.1.1 Initial Bioassay Selection

We first selected a set of bioassays which are enriched with selective compounds, and meanwhile, the compound selectivity in these bioassays can be largely defined by other selected bioassays. This set of bioassays provides a closed space from which a subset of bioassays will be further constructed (Section 5.1.2) for the experiments. We constructed this initial set of bioassays as follows:

1. Identify all "binding" bioassays with one "single protein" target;
2. From such single-target binding bioassays, find all the bioassays that use IC_{50} to measure compound activities, and keep the compounds in such bioassays that have exact IC_{50} values (i.e., discard from each bioassay the compounds with IC_{50} ranges, for example, $IC_{50} \geq 0.0001\mu M$; also discard compounds whose measurement cannot be converted to IC_{50} values);
3. Combine bioassays of a same target into one bioassay;

[¶] http://cs.iupui.edu/~liujunf/projects/selRank_2017/
[‖] https://www.ebi.ac.uk/chembl/, v.22_1, accessed on 12/08/2016)

4. Clean the combined bioassays as follows:

   (a) If a compound appears multiple times with a same IC_{50} value in one bioassay, keep the compound with the unique IC_{50} in the bioassay;

   (b) If a compound appears multiple times with different IC_{50} values in one bioassay, remove the compound and all its activities from the bioassay. This is to avoid the complication to resolve conflicts of inconsistent activity values;

   (c) If a compound has an invalid IC_{50} value (e.g., negative or zero IC_{50}), remove the compound from the bioassay.

5. Select the cleaned bioassays that have at least 20 active compounds.

After the above process, we identified 1,033 bioassays in total. Among these 1,033 bioassays (denoted as $\mathcal{B}_s^0$), 594 bioassays have selective compounds that are defined within these 1,033 bioassays. Among these 594 bioassays, 553 bioassays have selective compounds that are defined within these 594 bioassays. Among these 553 bioassays, 227 bioassays have more than 10 selective compounds, and these selective compounds are involved in 529 out of the 553 bioassays. This set of 529 bioassays represents the initial closed set of selectivity-enriched bioassays.

5.1.2 Initial Bioassay Pruning

Among the initial closed set of 529 selectivity-enriched bioassays, we defined selectivity for the compounds in each bioassay with respect to the rest 528 bioassays. These 529 bioassays are further pruned according to the following protocol in order to have reasonable number of compounds for dCPPP learning:

1. If a bioassay has less than 100 compounds, keep the bioassay as it is;

2. If a bioassay has more than 100 compounds, identify all its selective compounds and x-selective compounds:

(a) If such identified selective and x-selective compounds are more than 100, keep all such compounds and discard all the other compounds;

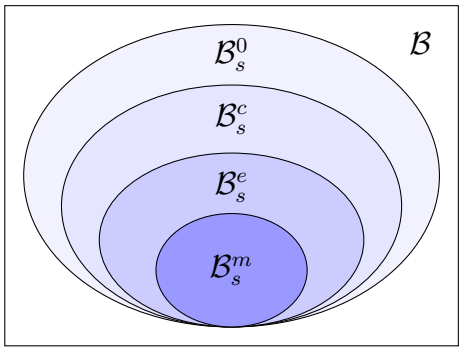
(b) If such identified compounds are less than 100, randomly select active compounds in this bioassay until the total number of selected compounds reaches 100.

The above pruning process retains all the selectivity related information in the original closed space of selectivity-enriched bioassays. All the remaining bioassays and their compounds are used as the final dataset in our experiments. This set of 529 pruned bioassays is denoted as $\mathcal{B}_s^c$. In $\mathcal{B}_s^c$, 408 bioassays have at least one selective compound. This set of 408 bioassays with selective compounds is denoted as $\mathcal{B}_s^e$. The rest of 121 bioassays (i.e., $\mathcal{B}_s^c \setminus \mathcal{B}_s^e$) do not have selective compounds, but they contain x-selective compounds (i.e., selective compounds in other bioassays).

5.1.3 Dataset Description

We use $\mathcal{B}_s^c$ in our experiments. Models with powered-push will be built for the bioassays in $\mathcal{B}_s^c$. In $\mathcal{B}_s^c$, 155 bioassays have $10 \sim 50$ selective compounds and less than 500 compounds. In this manuscript, we only report experimental results on these 155 bioassays, denoted as $\mathcal{B}_s^m$, because they have on average more selective compounds. Additional experimental results on $\mathcal{B}_s^c$ are available in the Supporting Information. Note that if a bioassay in the final dataset has more than 100 compounds, these compounds have to be either selective compounds or x-selective compounds, based on the protocol in Section 5.1.2.

Figure 2 presents the relations among all bioassay sets generated during the dataset construction process. Table 2 (the "before split" row) presents the data description for $\mathcal{B}_s^c$ and $\mathcal{B}_s^m$. Figure 3 presents the size of bioassays in $\mathcal{B}_s^c$. Figure 4 presents the size of bioassays in $\mathcal{B}_s^m$.



$\mathcal{B}$: the entire bioassay space
$\mathcal{B}_s^0$: 1,033 activity-enriched bioassays
$\mathcal{B}_s^c$: 529 selectivity-closed bioassays
$\mathcal{B}_s^e$: 408 bioassays with selective compounds
$\mathcal{B}_s^m$: 155 bioassays with $10 \sim 50$ selective compounds

Bi-directional push strategy will be applied to bioassays in $\mathcal{B}_s^m$ within the context of $\mathcal{B}_s^c$.

Figure 2: Relations among bioassay sets

Table 2: Dataset Description

	dataset	$\lvert\{B\}\rvert$	$\lvert\{c_i\}\rvert$	$\lvert C_k\rvert$	$\lvert A_k\rvert$	$\lvert S_k\rvert$	$\lvert S_k^x\rvert$
before split	$\mathcal{B}_s^c$	529	35,226	104.50	80.24	24.26	31.12
	$\mathcal{B}_s^m$	155	14,568	102.27	80.67	21.60	36.56
after split	$\mathcal{B}_s^m$	155	14,568	102.27	84.18	18.09	18.61

The column "$\lvert\{B\}\rvert$" has the number of bioassays in the dataset. The column "$\lvert\{c_i\}\rvert$" has the total number of unique compounds in the dataset. The column "$\lvert C_k\rvert$" has the average number of compounds in each bioassay. The column "$\lvert A_k\rvert$" has the average number of non-selective compounds in each bioassay. The column "$\lvert S_k\rvert$" has the average number of selective compounds in each bioassay. The column "$\lvert S^x\rvert$" has the average number of x-selective compounds in each bioassay.

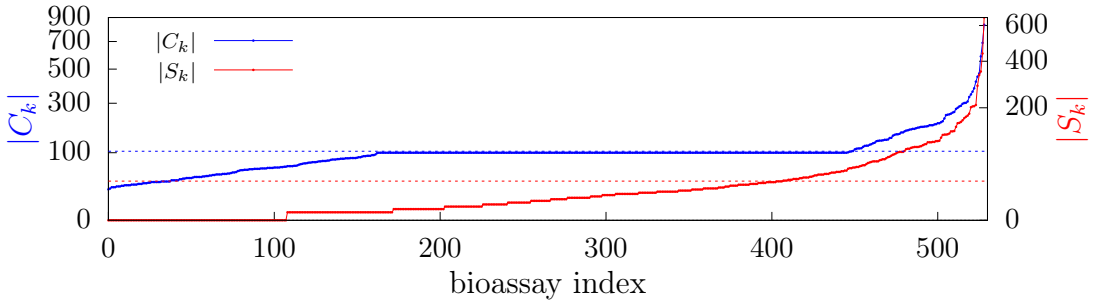


Figure 3: Bioassay size in $\mathcal{B}_s^c$ (before split)

5.2 Compound Feature Generation

We used AFGen** to generate binary compound fingerprints from the compound structures provided by ChEMBL. Each dimension of the fingerprints represents a compound substructure, and the binary value at each dimension represents whether the corresponding substruc-

**http://glaros.dtc.umn.edu/gkhome/afgen/overview

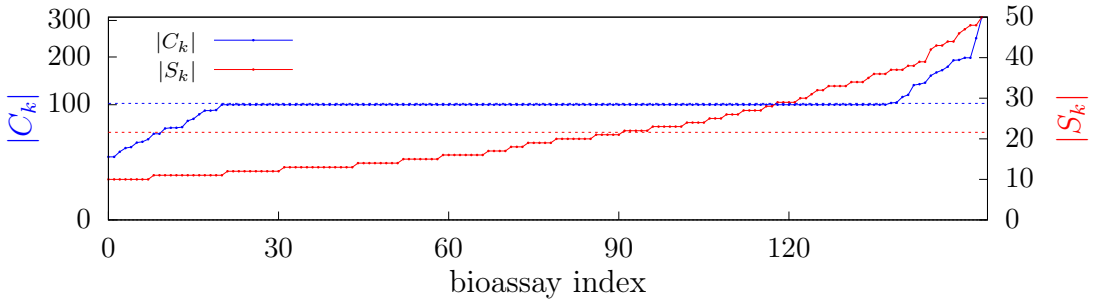


Figure 4: Bioassay size in $\mathcal{B}_s^m$ (before split)

ture is present in the corresponding compound or not. Previous research[37] demonstrates that such compound fingerprints are superior to others in compound classification.

For each bioassay, we calculated the pairwise Tanimoto similarity[38] of all the compounds in the bioassay, and used each row of the Tanimoto matrix as the feature representation of the corresponding compound. Intuitively, the features of a compound c_i represent the similarities between c_i and all training compounds in the same bioassay. This feature representation scheme is inspired by the idea in Que and Belkin.[39] Therefore, a same compound will have different features in different bioassays, and the different compound information that may induce different ranking structures is also encoded in the bioassay-specific compound features. This compound feature representation is unique compared to the existing compound fingerprint representations, and it is generated in a way that is dependent of computational tasks.

In our experiments, the bioassay-specific compound feature representation achieves best `CI` (will be discussed later in Section 5.4.5) 0.717 in dCPPP° on $\mathcal{B}_s^m$, compared to the best `CI` 0.734 using AFGen features in dCPPP°, and the best `CI` 0.748 using Tanimoto on AFGen features as a kernel in SVMRank.[34] Although AFGen feature with SVMRank achieves better results, it is significantly slower (i.e., 640 seconds on average to train a model) than bioassay-specific compound feature with dCPPP° (i.e., 79 seconds on average). Similarly, AFGen feature in dCPPP° is also significantly slower (i.e., 310 seconds on average to train a model) than bioassay-specific compound feature in dCPPP° (i.e., 79 seconds on average). Thus,

the bioassay-specific compound feature representation together with dCPPP° gives the best performance in terms of the combination of run time and the ranking results, and will be used in the experiments.

5.3 Experimental Protocol

We randomly split each bioassay into five folds and make sure that selective compounds are evenly split into the five folds. We conducted five-fold cross validation over the splits to evaluate the dCPPP performance. Note that once the data are split, the selectivity for any training compounds needs to be re-defined with respect to only the training (i.e., known) compounds of the bioassays. This is because that testing compounds are hold out as unknown compounds, and thus cannot be used to define selectivity. Similarly, the selectivity of the testing compounds (i.e., the ground-truth for performance evaluation) is also re-defined with respect to training data. In principle, the selectivity re-defined after data split will be different from that before data split. However, due to the fact that the data are split randomly and independently for selective (defined before data split) and active compounds, it is expected the selective (defined after data split) and active compounds are still evenly distributed across folds. Table 2 (the "after split" row) presents the data description after the split. After the data split, all the 155 bioassays in $\mathcal{B}_s^m$ have selective compounds in at least one testing fold. The evaluation metrics are only calculated and averaged over testing folds which have selective compounds.

5.4 Evaluation Metrics

We define and use the following metrics to evaluate the performance of dCPPP.

5.4.1 Average Precision at k (ap@k)

The average precision at k (ap@k)[††] is a popular metric used in LETOR. It considers the ranking positions of selective compounds among the top k positions of the ranking list. Average precision at k is defined as:

$$\texttt{ap@k} = \sum_{i=1}^{k} \frac{P(i)}{min(|S_k|, k)}, \tag{20}$$

where $P(i)$ is the precision[‡‡] among the top-i items in the ranking list. Higher ap@k values indicate that selective compounds are ranked higher.

5.4.2 Reciprocal Selectivity Position Index (RSPI)

Absolute ranking position is an important metric in compound prioritization. This is because in real applications, typically, the top few compounds in a ranking list will be of primary interest. Thus, we define a reciprocal selectivity position index, denoted as RSPI, to measure the average absolute reciprocal ranking positions of selective compounds in a ranking list:

$$\texttt{RSPI}(C_k) = \frac{1}{|S_k|} \sum_{c_i \in S_k} \frac{1}{\tilde{p}_i^k}, \tag{21}$$

where $\tilde{p}_i^k$ is the ranking position of a selective compound c_i in bioassay B_k predicted by a ranking model. The reciprocals are used to favor highly ranked compounds by up weighting the contribution of highly ranked selective compounds, and down weighting the contribution of lowly ranked selective compounds. Higher RSPI values indicate higher average absolute ranking positions for selective compounds and thus better performance of the ranking model.

[††] https://www.kaggle.com/wiki/MeanAveragePrecision

[‡‡] https://en.wikipedia.org/wiki/Information_retrieval#Precision

5.4.3 Normalized Reciprocal Selectivity Position Index (NRSPI)

A normalized version of RSPI, denoted as NRSPI, is defined via the inclusion of reciprocal ranking positions of all the compounds in a bioassay, so as to also measure the relative ranking positions of selective compounds in the ranking list:

$$\texttt{NRSPI}(C_k) = \sum_{c_i \in S_k} \frac{1}{\tilde{p}_i^k} \Bigg/ \sum_{c_j \in C_k} \frac{1}{\tilde{p}_j^k}. \tag{22}$$

Higher NRSPI values indicate higher average relative reciprocal ranking positions of selective compounds. Both RSPI and NRSPI are similar to ap@k, a popular metric for ranking performance, but RSPI and NRSPI consider the ranking structures among selective/active compounds.

5.4.4 Normalized Selectivity Position Index (NSPI)

We also define a normalized selectivity position index, denoted as NSPI, which measures the average percentile rankings of selective compounds:

$$\texttt{NSPI}(C_k) = \frac{1}{|C_k| \times |S_k|} \sum_{c_i \in S_k} \tilde{p}_i^k, \tag{23}$$

where $\tilde{p}_i^k$ is the ranking position of a selective compound c_i in bioassay B_k predicted by a ranking model. NSPI is normalized by the size of bioassays. Lower NSPI values indicate higher ranking positions for selective compounds on average.

5.4.5 Concordance Index (CI)

Concordance Index (CI) is a popular metric that is used to evaluate the performance of ranking algorithms.[40] CI measures the fraction of correctly ordered pairs among all possible

pairs and thus it is complementary to the nCI defined in Equation 2, that is,

$$\texttt{CI}(C_k) = 1 - \texttt{nCI}(C_k). \tag{24}$$

Higher CI values indicate better prediction overall (i.e., more concordant pairs are predicted correctly).

In our experiments, we measure the CI values over all compounds C_k in a bioassay B_k. We also measure the CI values among only selective compounds S_k, and among only non-selective compounds A_k in B_k, respectively. In this case, the CI values are specifically denoted as sCI and aCI, respectively.

6 Conclusions

We have developed the differential compound prioritization via bi-directional push with power method dCPPP. In dCPPP, activity ranking and selectivity prioritization are both tackled within one differential optimization model that leverages collaborative information from multiple bioassays. A bi-directional powered push strategy is implemented in dCPPP, which pushes selective compounds up and x-selective compounds down in ranking. We have also conducted a comprehensive set of experiments and analysis on the ranking performance of dCPPP. Our experiments demonstrate that dCPPP is very effective in prioritizing selective compounds while maintaining a good activity ranking.

Overall, dCPPP achieves significant improvement in compound selectivity prioritization. In specific, dCPPP* outperforms dCPPP$^\circ$ in selective compound prioritization in terms of ap@5 at 47.0%, and in terms of RSPI at 26.1%, with statistical significance. Meanwhile, dCPPP still preserves a good overall activity ranking among all compounds. Specifically, dCPPP* maintains a similar performance in CI (even slightly better by 1.2%) as dCPPP$^\circ$. The overall experimental results on all evaluation metrics are available in Section 7.1, and dCPPP needs only two iterations in order to achieve its optimality.

The experimental results show that, after the first iteration, the performance of dCPPP increases significantly in terms of all evaluation metrics related to selective compounds prioritization, and slightly decreases in compound activity ranking (e.g., in `CI`). Specifically, the performance of dCPPP* in terms of `ap`@5 and `RSPI` increases from 0.558 and 0.411 to 0.687 and 0.490 over dCPPP°, respectively. However, the compound activity ranking performance, in terms of `CI`, decreases from 0.635 to 0.631 in the first iteration. In the second iteration, dCPPP is still able to improve compound selectivity prioritization but the improvement is not as significant as that from the first iteration. This indicates that the system quickly converges to a stable state, and the selectivity prioritization has been updated toward optimal conditions. Specifically, the performance in terms of `ap`@5 and `RSPI` is increased from 0.687 and 0.490 to 0.702 and 0.499, respectively, which is relatively marginal compared to that in the first iteration. On the other hand, dCPPP tries to fix the compound activity ranking in the second iteration that has been altered in the first iteration, and thus the `CI` performance increases from 0.631 to 0.636 in the second iteration. Detailed results on compound ranking and selective compound prioritization over the two iterations and over the hyperparameters are available and discussed in Section 7.2 and 7.3.

In terms of top-N ranking performance, dCPPP has siginificantly better performance in retaining top-N compounds of ground truth, in ranking selective compounds among top, and in retaining selective compounds from top-N compounds of ground truth. In specific, in terms of retaining top-N compounds, dCPPP* has better performance (on average 2.40/6.59 top-5/10 compounds retained among top5/10 rankings, respectively) compared to that of dCPPP° (on average 2.37/6.51 top-5/10 compounds retained among top5/10 rankings, respectively). In terms of ranking selective compounds among top, dCPPP* significantly outperforms dCPPP°. On average, dCPPP* ranks 2.52/3.21 selective compounds among top-5/10 rankings, but dCPPP° ranks only 2.25/3.04 selective compounds among top-5/10 rankings. Moreover, among the average 1.98 selective compounds among top-5 compounds of each bioassay in the ground truth, dCPPP* is able to retain 1.51 of them on average,

while dCPPP° is able to only retain 1.38. Among the average 1.01 selective compounds in top-6 to top-10 compounds of ground truth, dCPPP* is able to push 0.66 of them into top 5, while dCPPP° has 0.56 such compounds in top 5. Detailed results and analysis on top-N performance are presented in Section 7.4.

Overall, our experiments demonstrate that dCPPP is very effective in compound selectivity prioritization and competent in compound activity ranking. Detailed result analysis will be thoroughly discussed in Section 7.

7 Experimental Results

In the results presented in this section, we used parameters $\theta^{\uparrow} = 0.5$ and $\theta^{\downarrow} = 0.5$. We tested combinations of various $\theta^{\uparrow}$ and $\theta^{\downarrow}$ values, and found that $\theta^{\uparrow} = 0.5$ and $\theta^{\downarrow} = 0.5$ give the best performance over all the evaluation metrics overall. Based on our experiments, only two iterations will lead to systematic convergence. Therefore, we only report the results from the two iterations.

7.1 Overall Performance

Table 3 presents overall performance comparison between the dCPPP° and the optimal dCPPP* models. Note that for each bioassay, its optimal dCPPP* is the model that introduces the best `RSPI` value, and thus the performance of dCPPP* in terms of other metrics (e.g., `ap`@5; the dCPPP*(t) rows in Table 3) does not necessarily correspond to the optimal in those metrics. The optimal performance in each respective metric is calculated as the "b-imprv (%)" values, and therefore, the performance in "b-imprv (%)" does not necessarily correspond to a same set of parameters. The "diff (%)" values in Table 3 are calculated as percentage difference of average dCPPP* performance over average dCPPP° performance, where the average performance is calculated as the average over all the bioassays in respective metrics. The "imprv (%)" values in Table 3 are calculated as the average of bioassay-wise

Table 3: Overall Performance Comparison

iter	method	ap@5	ap@10	RSPI	NRSPI	NSPI	CI	aCI	sCI
1	dCPPP°(1)	0.558	0.613	0.411	0.383	0.268	0.635	0.599	0.506
	dCPPP*(1)	0.687	0.733	0.490	0.439	0.218	0.631	0.590	0.462
	diff (%)	23.118	19.576	19.221	14.621	18.657	-0.630	-1.503	-8.696
	imprv (%)	43.193	28.402	23.761	22.813	17.214	0.263	0.169	1.665
	p-value	6.68e-24	1.06e-25	7.96e-23	2.58e-28	9.46e-18	4.36e-1	2.12e-1	1.40e-3
	b-imprv (%)	46.765	30.704	23.761	23.367	22.468	17.161	22.474	68.473
	p-value	6.43e-49	7.03e-47	7.96e-23	7.90e-47	7.21e-34	2.87e-7	3.81e-6	1.77e-23
2	dCPPP°(2)	0.687	0.733	0.490	0.439	0.218	0.631	0.590	0.462
	dCPPP*(2)	0.702	0.746	0.499	0.445	0.213	0.636	0.596	0.467
	diff (%)	2.183	1.774	1.837	1.367	2.294	0.792	1.017	1.082
	imprv (%)	2.726	2.160	1.784	1.785	1.749	1.110	1.382	1.680
	p-value	2.47e-7	4.31e-9	9.39e-9	6.57e-11	3.20e-3	1.99e-2	7.68e-2	4.79e-1
	b-imprv (%)	4.562	3.322	1.784	2.019	6.457	17.119	23.077	72.756
	p-value	1.97e-17	2.73e-16	9.39e-9	3.84e-12	6.58e-17	7.74e-32	6.37e-36	1.16e-34
overall	dCPPP°(1)	0.558	0.613	0.411	0.383	0.268	0.635	0.599	0.506
	dCPPP*(2)	0.702	0.746	0.499	0.445	0.213	0.636	0.596	0.467
	diff (%)	25.806	21.697	21.411	16.188	20.522	0.157	-0.501	-7.708
	imprv (%)	47.003	31.269	26.096	25.157	18.952	1.203	1.320	1.813
	p-value	1.76e-26	1.68e-28	1.61e-24	4.33e-31	9.17e-20	8.15e-1	6.96e-1	3.90e-3
	b-imprv (%)	49.181	32.732	26.096	25.437	22.734	16.478	21.520	64.846
	p-value	1.33e-30	3.36e-31	1.61e-24	7.48e-32	4.38e-25	2.67e-27	8.78e-29	5.92e-24

The columns "ap@5", "ap@10", "RSPI", "NRSPI", "NSPI", "CI", "aCI" and "sCI" have the average ap@5, ap@10, RSPI, NRSPI, NSPI, CI, aCI and sCI values of dCPPP over the testing sets. The row "dCPPP°(t)" has the average model performance in each respective metric from dCPPP° in iteration t. The row "dCPPP*(t)" has the average model performance in each respective metric from dCPPP* in iteration t (dCPPP* corresponds to the model of best RSPI value). The row "diff (%)" has the percentage difference of average performance in each respective metric of dCPPP°(t) and dCPPP*(t). The row "imprv (%)" has the average of bioassay-wise improvement from dCPPP*(t) over dCPPP°(t) in each respective metric. The row "b-imprv (%)" has the average of each bioassay's best improvement in each respective metric. The row "p-value" has the p-values for "imprv (%)"/"b-imprv (%)".

performance improvement from dCPPP° over dCPPP*.

In dCPPP iteration 1 (i.e., the row block where "iter" has "1" in Table 3), the average performance of dCPPP* is significantly better (i.e., "imprv (%)") than that of dCPPP° in terms of ap@5, ap@10, RSPI, NRSPI and NSPI (p-values 6.68e-24, 1.06e-25, 7.96e-23, 2.58e-28 and 9.46e-18, respectively). In terms of CI and aCI, dCPPP* is not significantly different

(p-values 4.36e-1 and 2.12e-1, respectively) from dCPPP$^\circ$ on the average performance (i.e., "imprv (%)"). This demonstrates that dCPPP is able to better prioritize selective compounds while retaining the overall ranking structures of active compounds. In terms of `sCI`, it turns out that dCPPP* is still significantly better (p-value 1.40e-3) than dCPPP$^\circ$ on the average performance (i.e., "imprv (%)"). This indicates that for a significant amount of bioassays, differential push could also help activity ranking. In terms of the best performance with respective to each metric (i.e., "b-imprv (%)"), dCPPP* significantly outperforms dCPPP$^\circ$ on all the metrics including `CI`, `aCI` and `sCI`. This indicates that by pushing compounds differently, it may also help better rank all the compounds overall.

In dCPPP iteration 2 (i.e., the row block where "iter" has "2" in Table 3), the average ranking performance (i.e., "imprv") of dCPPP* is still significantly better than that of dCPPP$^\circ$ in all the metrics (except in `aCI` and `sCI` where the improvement is not significant). However, the performance improvement is not as great as that in iteration 1, and the smaller improvement also applies in the best performance with respect to each metric (i.e., "b-imprv (%)"). This indicates that the iterative learning process starts to converge in iteration 2. In particular, the dCPPP* performance of ranking both active and selective compounds (i.e., in terms of `CI`) is improved significantly from dCPPP$^\circ$. The performance in terms of `aCI` and `sCI` is also improved in iteration 2 (i.e., positive "diff (%)" in iteration 2 compared to the negative value in iteration 1). This indicates that in iteration 2, the learning process tends to fix the broken ranking structures among both selective and active compounds and thus converge to a systematically stable state. The results from the two iterations show that the dCPPP method is able to continuously push the selective/x-selective compounds over iterations, and meanwhile, it tends to maintain good ranking structures among both selective and active compounds.

Over these two iterations (i.e., the row block where "iter" has "overall" in Table 3), dCPPP* significantly outperforms dCPPP$^\circ$ in all the evaluation metrics (except in `CI` and `aCI`, in which the improvement is not significant). In particular, dCPPP is able to improve

selectivity prioritization in terms of ap@5 at 47.003%, and in terms of RSPI at 26.096%, both with statistical significance. These results demonstrate the superiority of the dCPPP in prioritizing selective compounds.

7.2 Selective Compound Prioritization

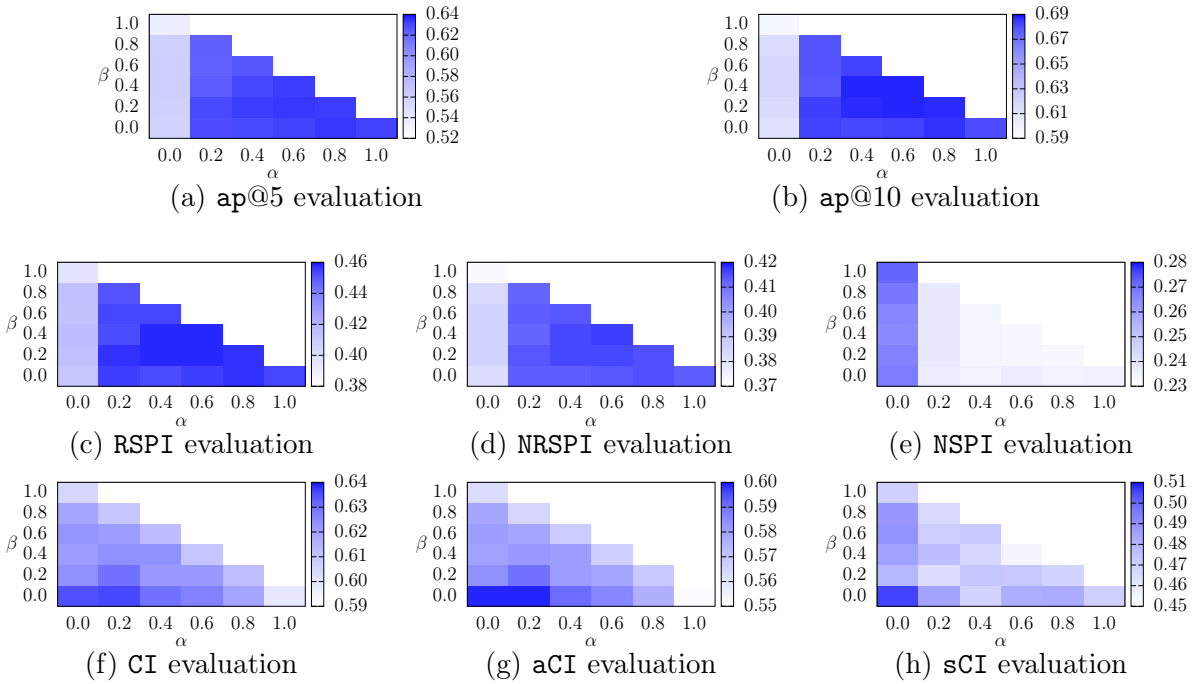


Figure 5: Evaluation of dCPPP(1) on $\mathcal{B}_s^m$

Figure 5a, 5b, 5c, 5d and 5e present the results of dCPPP(1) in terms of ap@5, ap@10, RSPI, NRSPI and NSPI, respectively, over various α and β values (i.e., the parameters to weight the push-up and push-down terms, respectively, in dCPPP Equation 12). The values in these figures are the average performance in respective evaluation metrics over all the bioassays in which both push-up for selective compounds and push-down for x-selective compounds can be applied (i.e., bioassays in dataset $\mathcal{B}_s^m$). Correspondingly, Figure 6a, 6b, 6c, 6d and 6e show performance in terms of ap@5, ap@10, RSPI, NRSPI and NSPI of dCPPP(2) over different α and β settings.

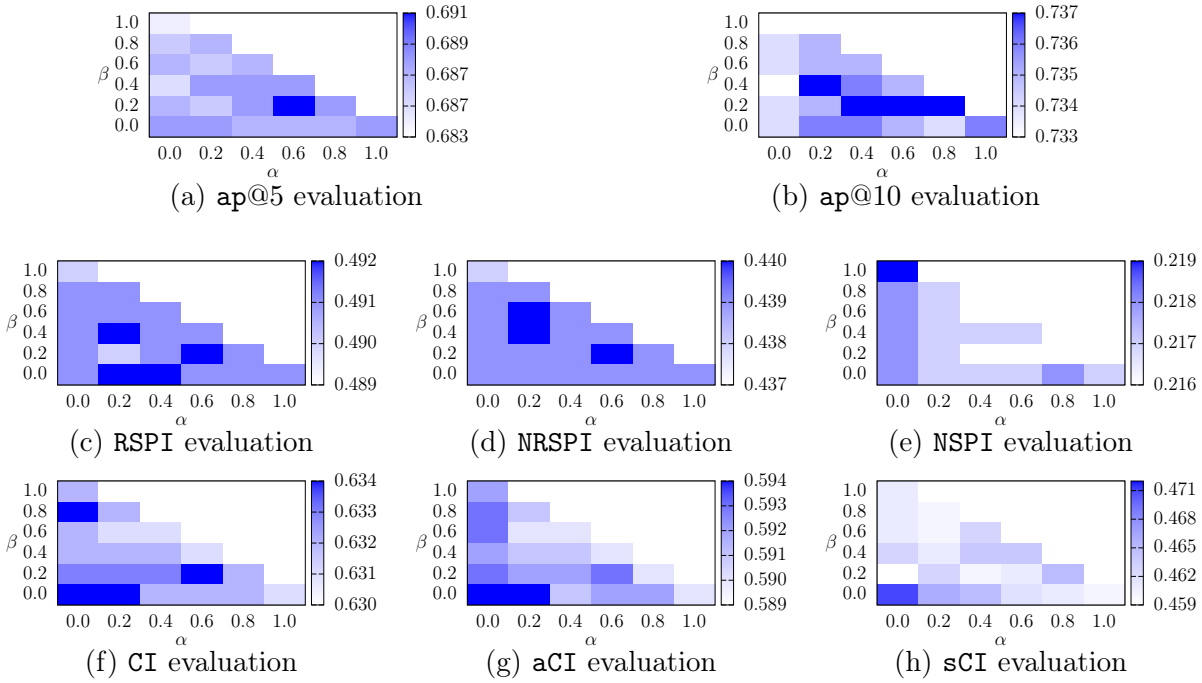


Figure 6: Evaluation of dCPPP(2) on $\mathcal{B}_s^m$

7.2.1 dCPPP(1) **Performance**

Figure 5a and 5b show that in the first iteration, dCPPP has the optimal ap@5 performance (ap@5 $= 0.634$) at ($\alpha = 0.6, \beta = 0.2$), and the optimal ap@10 performance (ap@10 $= 0.688$) when $\alpha = 0.6$ and $\beta \in [0.2, 0.4]$. The optimal results demonstrate that, when push-up weight is large ($\alpha \geq 0.6$) and push-down is also applied, the selective compounds are preferably pushed into top-5/10 of the ranking lists.

In Figure 5a, there is a notable gap between the ap@5 values when $\alpha = 0$ and $\alpha > 0$. Specifically, when the push-up starts to take effect (i.e., α is increased from 0), the ap@5 values are increased significantly. A similar gap also exists in the ap@10 performance between $\alpha = 0$ and $\alpha > 0$ in Figure 5b. This indicates that even a slight push-up could alter the ranking structure significantly and push the selective compounds up into the top of the ranking lists. However, the full-power push-up (i.e., $\alpha = 1.0$) without considering the activity ranking performance among compounds (i.e., considering only the $\mathcal{L}_s^{k+}$ term and neglecting the $\mathcal{L}_c^k$ and $\mathcal{L}_x^{k-}$ terms in Equation 12) does not lead to the optimal solu-

tion in terms of both ap@5 and ap@10. This indicates that the prioritization of selective compounds over non-selective compounds is structurally constrained by the ordering among both selective and non-selective compounds together, and leveraging the information from non-selective compounds and their ordering structures is beneficial in improving selective compound prioritization in top-5/10 of the ranking.

On the other hand, push-down over the x-selective compounds also benefits the selective compounds prioritization. For example, ap@10 is increased from 0.682 at ($\alpha = 0.4, \beta = 0.0$), to 0.687 at ($\alpha = 0.4, \beta = 0.2$) in Figure 5b. This may be due to the fact that the push-down exerts extra force on altering the overall ranking structures of each bioassay and thus better separates selective compounds from x-selective compounds. However, an over push-down does not benefit selective prioritization any more. For example, ap@10 is decreased from 0.688 at ($\alpha = 0.4, \beta = 0.4$) to 0.683 at ($\alpha = 0.4, \beta = 0.6$) in Figure 5b. The reason could be that an overemphasis on x-selective becomes detrimental to the overall ranking structures among both selective and non-selective compounds.

Figure 5c presents the performance in terms of RSPI of all the $\mathcal{B}_s^m$ bioassays in the first iteration. In terms of RSPI (i.e., the average reciprocal positions of selective compounds), the best performance of dCPPP (RSPI $= 0.458$) is achieved at the parameter region $\alpha \in [0.4, 0.6], \beta \in [0.2, 0.4]$, that is, when both the push-up and push-down are applied, the selective compounds are most effectively to be ranked higher in the bioassays.

The trend of performance in RSPI is similar to that in ap@k, that is, 1) when α is increased from 0 (i.e., the push-up starts to take place), the RSPI values are significantly increased; 2) the full-power push-up does not lead to optimal performance; 3) push-down over the x-selective compounds also has effects on better ranking selective compounds; and 4) an over push-down (e.g., $\beta \geq 0.6$ with $\alpha = 0.4$) does not benefit selectivity prioritization; etc.

Figure 5d and 5e demonstrate concordant trend of NRSPI and NSPI with that of ap@5, ap@10 and RSPI, that is, the best performance in terms of NRSPI and NSPI, respectively,

happens with non-zero α and β values. NRSPI (Equation 22) is a very similar metric to RSPI (Equation 21), which considers all the compounds, instead of only selective compounds as in RSPI, in evaluating ranking positions of selective compounds. High RSPI and NRSPI values associated with non-zero α and β values indicate that selective compounds are ranked both higher in their average absolute positions and higher in their average relative positions among all the compounds. NSPI measures the average percentile ranking of selective compounds. On average, the selective compounds are ranked at 77 percentile at best ($\alpha = 0.6, \beta = 0.2$ in Figure 5e), while in the baseline dCPPP$^\circ$ the average ranking percentile is 73.

7.2.2 dCPPP(2) Performance

Figure 6a and 6b present the performance in terms of ap@5 and ap@10 in the second iteration, respectively. The dCPPP method has optimal average ap@5 value (ap@5 $= 0.691$) at ($\alpha = 0.6, \beta = 0.2$), and optimal average ap@10 value (ap@10 $= 0.737$) at ($\alpha = 0.2, \beta = 0.4$), and ($\alpha \in [0.4, 0.8], \beta = 0.2$). Both of ap@5 and ap@10 in the second iteration are significantly improved from that in the first iteration (8.99% and 7.12%, respectively). This demonstrates that as in dCPPP(2), more selective compounds are pushed into top-5/10 as the push-up and push-down powers are applied ($\alpha > 0$, $\beta > 0$). Please note that in Table 3, the best ap@5 and ap@10 values are calculated according to dCPPP* that is defined with respect to optimal RSPI values, but in Figure 6a and 6b, the ap@5 and ap@10 values are the average values over all the bioassays under certain α and β values.

In the second iteration, the change of the ap@5 and ap@10 over α and β values is generally smooth. However, there are still some minor irregular trends. For example, ap@5 values first decrease from 0.687 at ($\alpha = 0.0, \beta = 0.2$) to 0.686 at ($\alpha = 0.2, \beta = 0.2$), then increase to 0.688 at ($\alpha = 0.4, \beta = 0.2$), although the changes are very small. This may indicate that in the second iteration, the ranking structures become more sensitive to push powers, since they are close to optimal. Also, in the second iteration, both ap@5 and ap@10 results fall into a much smaller range over various α and β values (i.e., ap@5 $\in [0.684, 0.691]$

and ap@10 $\in [0.733, 0.737]$) compared to that of the first iteration (i.e., ap@5 $\in [0.536, 0.634]$ and ap@10 $\in [0.596, 0.688]$). The best results of ap@5 and ap@10 are only 1.02% and 0.55% better than their worst results in the second iteration. Actually, this is a common trend among all the evaluation metrics in the second iteration, which indicates that the system is becoming stabilized in terms of ap@k performance.

In the second iteration, as shown in Figure 6c, the best RSPI (RSPI $= 0.492$) is still at ($\alpha = 0.6, \beta = 0.2$) as that in the first iteration. The best RSPI performance from the second iteration is improved by 7.42% from that in the first iteration (RSPI $= 0.458$). However, some other α and β settings (i.e., ($\alpha = 0.2, \beta = 0.0$), ($\alpha = 0.2, \beta = 0.4$), ($\alpha = 0.2, \beta = 0.4$)) also result in similar optimal RSPI performance. This indicates that the system is becoming stabilized and more sensitive to push powers. The RSPI performance results from the second iteration also show that when push-up power is applied (i.e., $\alpha > 0$), the results are better than that without push-up power (i.e., $\alpha = 0$). However, too large push-up power (i.e., $\alpha > 0.6$) does not yield optimal results. This is a similar trend as in the first iteration. Similarly, a full push-down also breaks the overall ranking structures among selective and non-selective compounds, and thus, a non-optimal result (RSPI $= 0.490$) is expected when $\beta = 1.0$.

Figure 6d and Figure 6e present the performance in terms of NRSPI and NSPI of $\mathcal{B}_s^m$ bioassays in the second iteration, respectively. In Figure 6d and Figure 6e, NRSPI and NSPI also have similar trend with that of RSPI in Figure 6c. That is, when moderate push-up and push-down powers are applied, the optimal results are achieved. Specifically, in terms of NRSPI, the optimal result (NRSPI $= 0.440$) is achieved at ($\alpha = 0.2, \beta = 0.4$), ($\alpha = 0.2, \beta = 0.6$), and ($\alpha = 0.6, \beta = 0.2$). In terms of NSPI, the optimal result (NSPI $= 0.216$) is achieved at ($\alpha \in [0.4, 0.2]$, $\beta = 0.2$).

7.2.3 Overall Performance for Selective Compound Prioritization

For all the bioassays, we compared their ap@5, ap@10, RSPI, NRSPI and NSPI values

Table 4: Percentage Improvement of dCPPP($\alpha = 0.6, \beta = 0.2$) vs. dCPPP°

iter	method	ap@5	ap@10	RSPI	NRSPI	NSPI
1	dCPPP°(1)	0.558	0.613	0.411	0.383	0.268
	dCPPP(1)	0.634	0.688	0.458	0.416	0.233
	diff (%)	13.620	12.235	11.436	8.616	13.060
	imprv (%)	29.197	19.268	15.049	15.246	10.874
	p-value	4.83e-13	2.41e-15	3.12e-13	5.24e-14	4.66e-11
	b-imprv (%)	30.598	20.139	15.049	15.582	14.804
	p-value	9.27e-39	2.63e-38	3.12e-13	5.87e-38	1.09e-09
2	dCPPP°(2)	0.687	0.733	0.490	0.439	0.218
	dCPPP(2)	0.691	0.737	0.492	0.440	0.216
	diff (%)	0.582	0.546	0.408	0.228	0.917
	imprv (%)	0.874	0.813	0.240	0.366	0.821
	p-value	6.92e-02	3.10e-03	3.65e-02	2.37e-02	3.54e-02
	b-imprv (%)	1.239	1.093	0.240	0.457	2.075
	p-value	4.70e-03	5.33e-05	3.65e-02	9.80e-03	3.49e-05

The columns "ap@5", "ap@10", "RSPI", "NRSPI", and "NSPI" have the average RSPI, NRSPI, NSPI, CI, aCI and sCI values of dCPPP over the testing sets. The row "dCPPP°(t)" has the average model performance in each respective metric from dCPPP° in iteration t. The row "dCPPP (t)" has the average model performance in each respective metric from dCPPP($\alpha = 0.6, \beta = 0.2$) in iteration t (dCPPP($\alpha = 0.6, \beta = 0.2$) corresponds to the model of best RSPI value with various learning rate). The row "diff (%)" has the percentage difference of average performance in each respective metric. The row "imprv (%)" has the average of bioassay-wise improvement from dCPPP (t) over dCPPP°(t) in each respective metric. The row "b-imprv (%)" has the average of each bioassay's best improvement in each respective metric. The row "p-value" has the p-value for "imprv (%)"/"b-imprv (%)".

of dCPPP at ($\alpha = 0.6, \beta = 0.2$) with the respective values of dCPPP° in both iteration 1 and 2 in Table 4. The paired t-tests demonstrate the significance of the improvement from dCPPP on dCPPP° in iteration 1. However, in iteration 2, the improvement is relatively less significant (though mostly still significant at 5% confidence level). This is expected as the ranking starting to converge to a systematically stable state. Additionally, the small difference among performances with various push-up and push-down powers also indicates that the system is approaching an equilibrium.

7.3 Compound Ranking

Figure 5f, 5g and 5h present the CI values among all compounds, aCI among non-selective compounds and sCI among selective compounds over all the bioassays in the first iteration, respectively. Correspondingly, Figure 6f, 6g and 6h present the respective values over all the bioassays in the second iteration. In Figure 5f, as α and β increase, the CI values over all the bioassays decrease in general. This is anticipated as increasing α and β values will induce less emphasis on overall ranking structures as in Equation 12 and thus decreased CI values. However, dCPPP at ($\alpha = 0.2, \beta = 0.0$) slightly increases CI (CI = 0.636) from dCPPP$^\circ$ (CI = 0.635). This may be due to the fact that pushing up selective compounds may affect the ranking on other non-selective compounds and thus increase CI. Figure 5g shows the similar trend over aCI as that of CI, because the majority of compounds are non-selective compounds in the bioassays.

In iteration 2, Figure 6f and 6g show the similar trend that higher α and β values will lead to lower CI and aCI values. Also, dCPPP achieves both optimal CI and aCI at ($\alpha = 0.0, \beta = 0.0$) (CI = 0.634 and aCI = 0.594, respectively). This is because that, without any emphasis on selectivity, dCPPP is only interested in the ranking structure among all compounds by their activities. However, dCPPP also achieves optimal CI at ($\alpha = 0.0, \beta = 0.8$) and ($\alpha = 0.6, \beta = 0.2$). This indicates that in this iteration, dCPPP tends to repair the skewed active compound ranking structures even during selective compound prioritization.

In Figure 5h, the ranking performance in terms of sCI among only selective compounds changes relatively irregularly. Specifically, with $\alpha \in [0.4, 0.6], \beta \in [0.2, 0.4]$ (i.e., the optimal parameter region in which RSPI achieves the best), sCI is even below 0.5 (i.e., random ranking). This is because the selective compounds may be pushed into discordant orders compared to the ground truth. Note that the push-up power (Equation 6) is defined based on the difference of percentile rankings of a compound in multiple bioassays. Therefore, different selective compounds may receive different push powers within a bioassay due to

their ranking positions among others bioassays. Together with the combinatorial influence from multiple x-selective compounds pushed-down at same time in the same bioassay, it is less likely that the selective compounds are pushed up but still in their original orders as before the push.

In Figure 6h, dCPPP also achieves optimal `sCI` (`sCI` = 0.471) among selective compounds iteration 2 at ($\alpha = 0.0, \beta = 0.0$). The reason is similar to that of Figure 6f and 6g, that is, a full emphasis on the compound activity prioritization without any selectivity push (i.e., α=0 and β=0) will introduce a better overall ranking structure based on compound activities, and therefore, the selective compounds are also prioritized based on their activities. As α and β increase, `sCI` starts to vary irregularly. This is still because that different selective/x-selective compounds will receive different push-up/-down powers, depending on the compounds' ranking percentile differences among bioassays, and thus pushed into disconcordant pairs compared to the ground truth. Similar to the `ap`@5, `ap`@10, `RSPI`, `NRSPI` and `NSPI` values, which fluctuate in a very small range in iteration 2 (Section 7.2), `CI`, `aCI` and `sCI` also become more stable in iteration 2 than in iteration 1. This also indicates that the overall ranking is converging to a systematically equilibrium state in the second iteration.

7.4 Top-N Performance

In this section, we evaluate the top-N performance of dCPPP.

7.4.1 Compound Ranking

Table 5 presents the top-N (N = 5 and 10) performance of dCPPP compared to dCPPP$^\circ$ in ranking compounds (both selective and non-selective). Since $\alpha = 0.6$ and $\beta = 0.2$ represent a reasonably good set of parameters for all the bioassays overall as indicated in Section 7.2, we compare dCPPP at ($\alpha = 0.6, \beta = 0.2$) in top-$N$ performance evaluation. Please note that dCPPP* corresponds to the model which achieves optimal performance in terms of `RSPI` for each individual bioassay using their respective optimal α and β values, and the baseline

Table 5: Top-N Performance on Compound Ranking (Compound Counts)

iter	N	dCPPP$^\circ$	dCPPP$(0.6, 0.2)$	dCPPP*
1	5	2.37	2.31 (7.59×10^{-2})	2.36 (9.40×10^{-1})
	10	6.51	6.42 (2.47×10^{-2})	6.50 (7.74×10^{-1})
2	5	2.36	2.39 (9.18×10^{-2})	2.40 (4.02×10^{-2})
	10	6.50	6.56 (1.45×10^{-2})	6.59 (2.80×10^{-3})

The column "N" has the numbers of compounds on top of the ranking results that are considered. The columns "dCPPP$^\circ$", "dCPPP$(0.6, 0.2)$" and "dCPPP*" have the number of compounds from the top-N compounds in the ground truth that are still ranked among top N by dCPPP$^\circ$, by dCPPP at $(\alpha = 0.6, \beta = 0.2)$ and by dCPPP*, respectively. The numbers in parentheses in "dCPPP$(0.6, 0.2)$" and "dCPPP*" columns are the p-values comparing the results of dCPPP and dCPPP* with those of dCPPP$^\circ$, respectively.

model in iteration 2 dCPPP$^\circ(2)$ that dCPPP at $(\alpha = 0.6, \beta = 0.2)$ and dCPPP* improve from is dCPPP$^*(1)$.

In the first iteration, among the top 5/10 of the ranking results, dCPPP at $(\alpha = 0.6, \beta = 0.2)$ rank fewer compounds (i.e., 2.31/6.42 compounds, respectively) that are among top 5/10 in the ground truth than dCPPP$^\circ$ (i.e., 2.37/6.51 compounds, respectively) and the difference is close to statistical significance (p-value $7.59 \times 10^{-2}/2.47 \times 10^{-2}$). The optimal dCPPP* ranks about same ground-truth top-5/top-10 compounds (i.e., 2.36/6.50) compared to dCPPP$^\circ$ (the difference is statistically insignificant). This indicates that in terms of top-N ranking of ground-truth compounds (both selective and non-selective), dCPPP is very similar to dCPPP$^\circ$. In the second iteration, dCPPP at $(\alpha = 0.6, \beta = 0.2)$ is able to rank among top 5/10 more compounds (i.e., 2.39/6.56 compounds, respectively) that are among top 5/10 in the ground truth than dCPPP$^\circ$, and the difference is very close to statistical significance (p-value $9.18 \times 10^{-2}/1.45 \times 10^{-2}$). Moreover, dCPPP* in iteration 2 also has better performance in terms of ranking the top5/10 compounds from ground truth (i.e., 2.40/6.59 compounds, respectively) than dCPPP$^\circ$ with statistical significance. The optimal dCPPP* outperforms dCPPP at $(\alpha = 0.6, \beta = 0.2)$ in iteration 2 as well. Overall, the performance in iteration 2 is better than that of iteration 1, in term of both top-5 and top-10 ranking of both selective and active compounds. Particularly, in the first iteration, both dCPPP at $(\alpha = 0.6, \beta = 0.2)$ and dCPPP* do not outperform dCPPP$^\circ$. However, in the

second iteration, they outperform dCPPP° with reasonable significance. This indicates that dCPPP is able to improve the ranking at the top of the ranking lists over iterations.

Table 6: Top-N Performance on Compound Ranking (Bioassay Counts)

iter	N	method	0	1	2	3	4	5	6	7	8	9	10
1	5	dCPPP°	14	27	38	42	28	4	-	-	-	-	-
		dCPPP(0.6, 0.2)	13	29	42	44	24	4	-	-	-	-	-
		dCPPP*	11	30	40	44	26	4	-	-	-	-	-
	10	dCPPP°	0	0	1	3	11	26	34	33	29	15	2
		dCPPP(0.6, 0.2)	0	0	1	4	9	32	35	31	28	13	2
		dCPPP*	0	0	1	5	9	26	38	34	28	14	2
2	5	dCPPP°	11	30	40	44	26	4	-	-	-	-	-
		dCPPP(0.6, 0.2)	11	27	42	44	27	4	-	-	-	-	-
		dCPPP*	10	29	40	45	27	4	-	-	-	-	-
	10	dCPPP°	0	0	1	5	9	26	38	34	28	14	2
		dCPPP(0.6, 0.2)	0	0	1	3	11	22	36	36	29	14	2
		dCPPP*	0	0	1	4	11	20	34	37	31	15	2

The column "N" has the numbers of compounds on top of the ranking results that are considered. The column "method" has all the methods in comparison. The columns corresponding to number $0, 1, \cdots, k, \cdots, 10$ represent the number of bioassays that retrain k out of the top-N (N = 5, 10) most active compounds in the ground truth in top-N compound rankings by the various methods, respectively.

In Table 6, we compare the number of bioassays in which sufficient amount of top-N compounds in the ground truth are retained still among top-N rankings by the various methods. Note that here only the activity ranking is considered. Table 6 shows that in iteration 1, dCPPP° enables more bioassays to retain more true top-N compounds. For example, 28/4 bioassays retain 4/5 of the top-5 most active compounds in their top-5 rankings, respectively. Thus, cumulatively 32 bioassays retain at least 4 of the 5 most active compounds in their top-5 rankings, compared to 28 bioassays from dCPPP at ($\alpha = 0.6, \beta = 0.2$) and 30 bioassays from dCPPP*, respectively. Similarly for top-10 rankings, dCPPP° enables 46 bioassays to retain at least 8 compounds out of the 10 most active compounds, compared to 43 bioassays from dCPPP at ($\alpha = 0.6, \beta = 0.2$) and 44 bioassays from dCPPP*, respectively. The performance is expected, because dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP* push selective compounds higher than they should be as if solely activity is considered, and as a result lower some activity compounds from the top of the ranking lists. Even though,

the performance of dCPPP$^\circ$ and dCPPP are very comparable, indicating that dCPPP is able to achieve the overall compound ranking structures similarly as dCPPP$^\circ$. Table 6 also shows that in the second iteration, dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP* enable 31 bioassays to retain at least 4 out of 5 most active compounds among top 5 rankings, and 45 and 48 bioassays top retain at least 8 out of 10 most active compounds among top 10 rankings, respectively, which is better than dCPPP$^\circ$. Note that dCPPP$^\circ$ in iteration 2 is dCPPP* from iteration 1, and thus Table 6 shows that in iteration 2, the performance of dCPPP in terms of retaining top active compounds start to get better. This indicates that in the second iteration, dCPPP tends to fix the altered ranking lists from the first iteration, similarly as indicated in Table 5.

7.4.2 Compound Selectivity Ranking

Table 7: Top-N Performance on Selectivity Ranking (Compound Counts)

iter	N	dCPPP$^\circ$	dCPPP$(0.6, 0.2)$	dCPPP*
1	5	2.25	2.40 (1.47×10^{-8})	2.49 (3.76×10^{-17})
	10	3.04	3.18 (9.14×10^{-9})	3.19 (1.07×10^{-10})
2	5	2.49	2.50 (3.93×10^{-2})	2.52 (3.80×10^{-3})
	10	3.19	3.21 (2.90×10^{-2})	3.21 (4.98×10^{-2})

The column "N" has the numbers of compounds on top of the ranking results that are considered. The columns "dCPPP$^\circ$", "dCPPP$(0.6, 0.2)$" and "dCPPP*" have the number of selective compounds that are ranked among top N by dCPPP$^\circ$, by dCPPP at ($\alpha = 0.6, \beta = 0.2$) and by dCPPP*, respectively. The numbers in parentheses in "dCPPP$(0.6, 0.2)$" and "dCPPP*" columns are the p-values comparing the results of dCPPP and dCPPP* with those of dCPPP$^\circ$, respectively.

Table 7 presents the top-N ($N = 5$ and 10) performance of dCPPP compared to dCPPP$^\circ$ in prioritizing selective compounds. The comparison is in terms of the number of selective compounds that are ranked on top by dCPPP, regardless whether these selective compounds are ranked on top or not in the ground truth. Among the top 5/10 of the ranking results from iteration 1, dCPPP at ($\alpha = 0.6, \beta = 0.2$) consistently ranking more selective compounds (i.e., 2.40/3.18 selective compounds, respectively) compared to dCPPP$^\circ$ (i.e., 2.25/3.04 selective compounds, respectively), with statistical significance. Please note that these top ranked

selective compounds could be either among top N in the ground truth or below top N in the ground truth. If each bioassay uses its own optimal (in terms of RSPI) α and β parameters, dCPPP* also ranks more selective compounds (i.e., 2.49/3.19) than both dCPPP$^\circ$ with statistical significance and dCPPP at $(\alpha = 0.6, \beta = 0.2)$. In the second iteration, dCPPP at $(\alpha = 0.6, \beta = 0.2)$ also outperforms dCPPP$^\circ$ in ranking selective compounds among top 5/10. Specifically, dCPPP at $(\alpha = 0.6, \beta = 0.2)$ is able to rank 2.50/3.21 selective compounds in top 5/10 of the ranking list, while dCPPP$^\circ$ could rank 2.49/3.19 selective compounds, and the difference is statistically significant (p-value $3.93 \times 10^{-2}/2.90 \times 10^{-2}$). In addition, dCPPP* outperforms dCPPP$^\circ$ in iteration 2 as well and is able to rank 2.52/3.21 selective compounds in top5/10 with statistical significance. Also, dCPPP* outperforms dCPPP at $(\alpha = 0.6, \beta = 0.2)$ in ranking more selective compounds in top 5/10. The results in Table 7 demonstrates that over the two iterations, dCPPP is able to consistently push more selective compounds onto top.

Table 8: Top-N Performance on Selectivity Ranking (Bioassay Counts)

iter	N	method	0	1	2	3	4	5	6	7	8	9	10
1	5	dCPPP$^\circ$	11	35	45	37	22	6	-	-	-	-	-
		dCPPP$(0.6, 0.2)$	6	31	45	46	19	7	-	-	-	-	-
		dCPPP*	4	28	49	44	21	9	-	-	-	-	-
	10	dCPPP$^\circ$	5	21	42	34	25	14	9	4	1	0	0
		dCPPP$(0.6, 0.2)$	3	18	40	38	24	15	10	5	2	0	0
		dCPPP*	3	18	40	38	24	16	9	4	3	0	0
2	5	dCPPP$^\circ$	4	28	49	44	21	9	-	-	-	-	-
		dCPPP$(0.6, 0.2)$	4	28	48	44	22	9	-	-	-	-	-
		dCPPP*	4	28	46	45	23	9	-	-	-	-	-
	10	dCPPP$^\circ$	3	18	40	38	24	16	9	4	3	0	0
		dCPPP$(0.6, 0.2)$	3	18	40	39	23	16	9	5	3	0	0
		dCPPP*	3	17	40	38	23	16	10	5	2	0	0

The column "N" has the numbers of compounds on top of the ranking results that are considered. The column "method" has all the methods in comparison. The columns corresponding to number $0, 1, \cdots, k, \cdots, 10$ represent the number of bioassays that rank k selective compounds in top-N (N = 5, 10) compound rankings by the various methods, respectively.

Table 8 presents the number of bioassays that rank selective compounds on top. In this comparison, dCPPP is significantly better than dCPPP$^\circ$. For example, dCPPP at

($\alpha = 0.6, \beta = 0.2$) and dCPPP* enable 72 and 74 bioassays, respectively, to rank at least 3 selective compounds among top-5 rankings in iteration 1, compared to 65 bioassays from dCPPP$^\circ$. In terms of top 10 rankings, dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP* enable 32 and 32 bioassays, respectively, to rank at least 5 selective compounds among top-10 rankings in iteration 1, compared to 28 bioassays from dCPPP$^\circ$. In the second iteration, dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP* enables even more bioassays to rank more selective compounds. For example, for top-5 rankings, dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP* enable 75 and 77 bioassays to rank at least 3 selective compounds among top-5 rankings, respectively, compared to 74 bioassays from dCPPP$^\circ$. Note that in iteration 2, dCPPP$^\circ$ is the dCPPP* from iteration 1. Thus, compared to the best performance from iteration 1, dCPPP further improves selectivity ranking among top 5 in iteration 2. Similar conclusions can be drawn for top-10 rankings.

7.4.3 Compound Selectivity Push

Table 9: Top-N Performance on Selectivity Push (Compound Counts)

iter	N	gt	dCPPP$^\circ$	dCPPP$(0.6, 0.2)$	dCPPP*
1	1-5	1.98	1.38	$1.44(7.10 \times 10^{-3})$	$1.49(1.63 \times 10^{-6})$
	6-10	1.01	$\uparrow$0.56	$\uparrow 0.61(1.60 \times 10^{-3})$	$\uparrow 0.65(6.21 \times 10^{-8})$
2	1-5	1.98	1.49	$1.50(1.95 \times 10^{-1})$	$1.51(7.36 \times 10^{-2})$
	6-10	1.01	$\uparrow$0.65	$\uparrow 0.65(4.92 \times 10^{-1})$	$\uparrow 0.66(2.28 \times 10^{-1})$

The column "N" has the numbers of compounds on top of the ranking results that are considered. The column "gt" has the average number of selective compounds in top-N compounds in the ground truth. The columns "dCPPP$^\circ$", "dCPPP$(0.6, 0.2)$" and "dCPPP*" have results for dCPPP$^\circ$, dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP*, respectively. The numbers in parentheses in "dCPPP$(0.6, 0.2)$" and "dCPPP*" columns are the p-values comparing the results of dCPPP and dCPPP* with those of dCPPP$^\circ$, respectively. The first row corresponds to the number of selective compounds among top 5 in the ground truth that are still ranked in top 5 by the different methods. The second row corresponds to the number of selective compounds that are among top 10 to top 6 in the ground truth and ranked into top 5 (denoted by $\uparrow$) by the different methods.

Table 9 presents the performance of dCPPP in pushing ground-truth top-N selective compounds on top. The comparison is in terms of the number of selective compounds that are ranked among top N in the ground truth and have also been retained among top N by

dCPPP. In the first iteration, among the average 1.98 selective compounds among top 5 in the ground truth, dCPPP$^\circ$ is able to retain 1.38 of such selective compounds still among top 5, but dCPPP at ($\alpha = 0.6, \beta = 0.2$) is able to retain 1.44 and dCPPP* is able to retain 1.49, both with statistical significance compared to dCPPP$^\circ$. In addition, among 1.01 selective compounds that are among top 10 to top 6 in the ground truth, dCPPP$^\circ$ is able to push on average 0.56 selective compounds into its top-5 ranking compounds, while dCPPP at ($\alpha = 0.6, \beta = 0.2$) is able to push 0.61 and dCPPP* is able to push 0.65, both with statistical significance.

In the second iteration, among the 1.98 selective compounds among top 5 in the ground truth, dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP* are able to retain 1.50 and 1.51 such selective compounds still among top 5, respectively, while dCPPP$^\circ$ could retain 1.49 such selective compounds. Among the 1.01 selective compounds that are ranked in top 10 upto top 6 in the ground truth, dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP* could push 0.65 and 0.66 such selective compounds into top 5 of their ranking lists, while dCPPP$^\circ$ could push 0.65. The results in Table 9 demonstrate that dCPPP is able to retain most of the selective compounds on top, and push lower ranked selective compounds onto top. In addition, Table 9 shows that in the first iteration, in total there are 2.14 (i.e., $1.49 + 0.65$) selective compounds that are ranked on top 5 by dCPPP*, and those 2.14 selective compounds are ranked among top 10 in the ground truth. On the other hand, Table 7 shows that in the first iteration, dCPPP* ranks 2.49 (more than 2.14) selective compounds among top 5. This indicates that dCPPP* even pushes selective compounds that are ranked below top 10 in the ground truth onto top 5.

Table 10 compares the number of bioassays that retain a certain portion of selective compounds that are among top-N active compounds in the ground truth and still keep such selective compounds in top-N rankings. Table 10 shows that from dCPPP at ($\alpha = 0.6, \beta = 0.2$) and dCPPP*, more bioassays have a larger portion of top-5 selective compounds ($\geq 60\%$) still retained among top-5 rankings than from dCPPP$^\circ$ after the first iteration, and more

Table 10: Top-N Performance on Selectivity Push (Bioassay Counts)

iter	N	method	(%)	$[0,20)$	$[20,40)$	$[40,60)$	$[60,80)$	$[80,100)$	$[100,100]$	NA
1	5	dCPPP$^\circ$		20	8	20	20	2	67	18
		dCPPP$(0.6,0.2)$		18	5	18	22	2	72	18
		dCPPP*		13	5	19	20	3	77	18
	↑10	dCPPP$^\circ$		27	5	14	6	0	41	63
		dCPPP$(0.6,0.2)$		20	6	14	6	0	47	63
		dCPPP*		16	5	15	7	0	49	63
2	5	dCPPP$^\circ$		13	5	19	20	3	77	18
		dCPPP$(0.6,0.2)$		12	5	18	22	2	78	18
		dCPPP*		12	4	19	21	2	79	18
	↑10	dCPPP$^\circ$		16	5	15	7	0	49	63
		dCPPP$(0.6,0.2)$		16	5	15	7	0	50	63
		dCPPP*		16	5	14	7	0	51	63

The column "N" has the numbers of compounds on top of the ranking results that are considered. The column "method" has all the methods in comparison. The columns corresponding to number (%) "$[a,b)$" represent the portion (in percentage) of selective compounds are retained or pushed. The row blocks corresponding to "N=5" represents the number of bioassays which retain the corresponding portions of selective compounds among the top-5 compounds in the ground truth. The row blocks corresponding to "N=↑10" represent the number of bioassays which push corresponding portions of selective compounds from top-6 to top-10 active compounds in the ground truth into top-5 rankings.

bioassays have their top-6 to top-10 selective compounds pushed onto top-5 by dCPPP at $(\alpha = 0.6, \beta = 0.2)$ and dCPPP*. In the second iteration, even more bioassays have their top-5 selective compounds retained still among top-5 by dCPPP at $(\alpha = 0.6, \beta = 0.2)$ and dCPPP*, and more top-6 to top-10 selective compounds pushed up. This demonstrates that dCPPP is effective in prioritizing selective compounds.

7.5 Percentile Ranking Change

Figure 7 presents the difference of percentile rankings introduced by dCPPP(1) among the training selective compounds. The difference of percentile rankings of a compound c_i is defined as $\tilde{r}_i^{k+} - \max_{B_l} \tilde{r}_i^{l-}$, where $\tilde{r}_i^{k+}$ and $\tilde{r}_i^{l-}$ are the estimated percentile ranking of c_i in bioassay B_k as a selective compound, and in bioassay B_l as a non-selective compound, respectively. A positive/negative difference indicates that c_i is ranked higher/lower in B_k as a selective compound than in any/some other bioassays B_l as a non-selective compound.

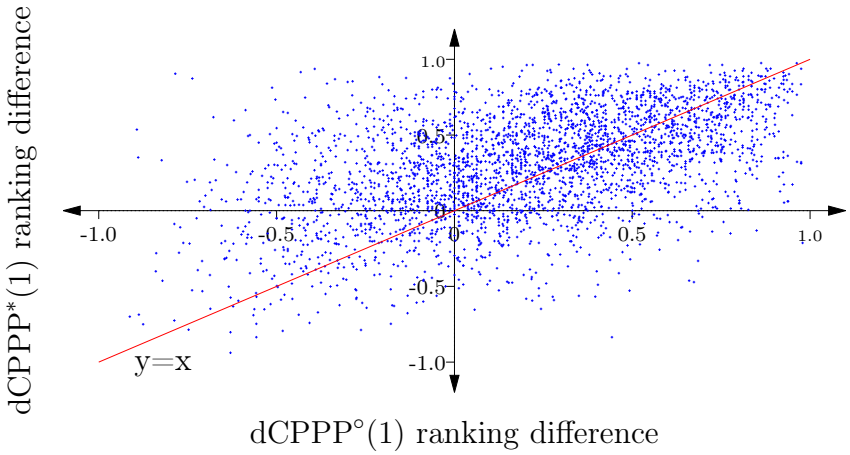


Figure 7: Ranking Difference among Selective Compounds in Iteration 1

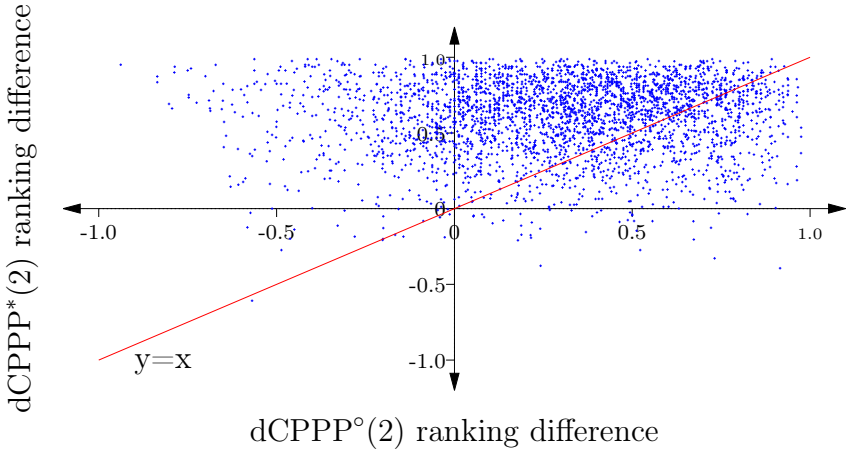


Figure 8: Ranking Difference among Selective Compounds in Iteration 2

Figure 7 shows that for dCPPP*(1), the majority of percentile ranking difference is positive (i.e., along y-axis, more data points above the line $x = 0$). This indicates that dCPPP is able to push selective compounds on top effectively. In addition, the average percentile ranking difference from dCPPP*(1) is larger than that from dCPPP°(1) (i.e., more data points above the line $y = x$ in Figure 7). This indicates that dCPPP is able to further distinguish selective compounds from non-selective compounds by pushing selective compounds on top. Specifically, in dCPPP°(1), selective compounds are ranked 20 percentage higher on average in the bioassays in which they are selective than in the bioassays in which they are non-selective. In dCPPP*(1), selective compounds are ranked 30 percentage higher on average. The difference between the ranking percentile difference in dCPPP*(1) and in dCPPP°(1) is

statistically significant (p-value=2.18×10^{-50}).

Figure 8 presents the difference of percentile rankings among the training selective compounds introduced by dCPPP(2). In dCPPP*(2), selective compounds are ranked on average 61 percentage higher in the bioassays in which they are selective than in bioassays in which they are non-selective (i.e., along y-axis in Figure 8). The difference between the ranking percentiles in dCPPP*(2) and in dCPPP°(2) is statistically significant (i.e., more data points above the line $y = x$; p-value=2.12×10^{-306}). The increase in the percentile ranking difference of training selective compounds indicates that dCPPP is powerful to further push up the selective compounds and push down the x-selective compounds in iteration 2. Also, the significant difference between the ranking difference introduced by dCPPP°(2) and that introduced by dCPPP*(2) shows that, after iteration 2, the selective compounds have been ranked significantly higher in the bioassays in which they are selective and in other bioassays in which the compounds are non-selective.

7.6 Push Power Change

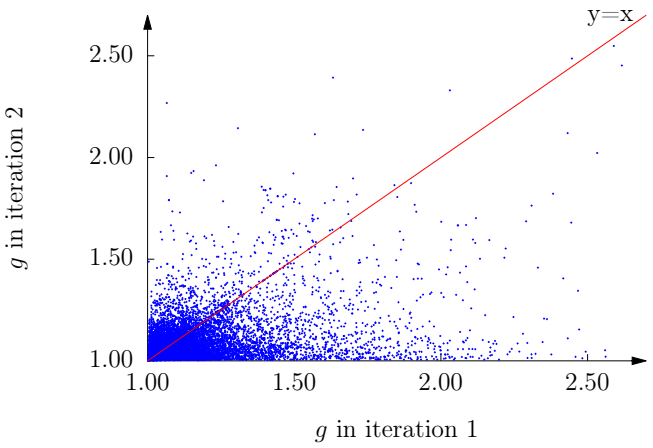


Figure 9: Push-up Weight Change among Selective Compounds

Figure 9 and Figure 10 present the change of push-up/push-down powers (i.e., g in Equation 6 and h in Equation 10) on the training selective compounds between the two iterations, respectively. The average push-up power in iteration 1 and 2 is $\bar{g}_1 = 1.16$ and

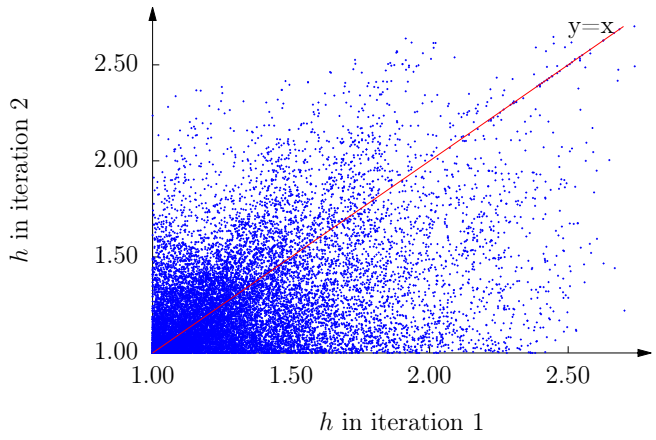


Figure 10: Push-down Weight Change among x-Selective Compounds

$\bar{g}_2 = 1.09$, respectively. The average push-down power in iteration 1 and 2 is $\bar{h}_1 = 1.34$ and $\bar{h}_2 = 1.27$, respectively. The difference between the average push-up powers in iteration 1 and the average push-up power in iteration 2 is statistically significant with p-value 2.47×10^{-322}. The difference between the average push-down powers is also significant with p-value 2.20×10^{-163}. The decrease of the push powers in iteration 2 indicates that when the selective compounds are pushed higher after iteration 1, the ranking difference of selective compounds in the bioassay in which they are selective and in other bioassays in which they are non-selective is increased (Equation 6 and 10).

8 Discussions

8.1 Push Relation Among Bioassays

Figure 11 presents a subset of push relations among all the bioassays in the first iteration of dCPPP as a weighted directed network. Each node in the network represents one bioassay. Since each bioassay has one unique target, the gene name of the target is used to represent each bioassay on the corresponding node. An edge from bioassay B_l to bioassay B_k represents that there is a compound shared by B_l and B_k, and the compound in B_k is pushed with a power determined by the its ranking difference in B_k and B_l (i.e., B_l helps to push the

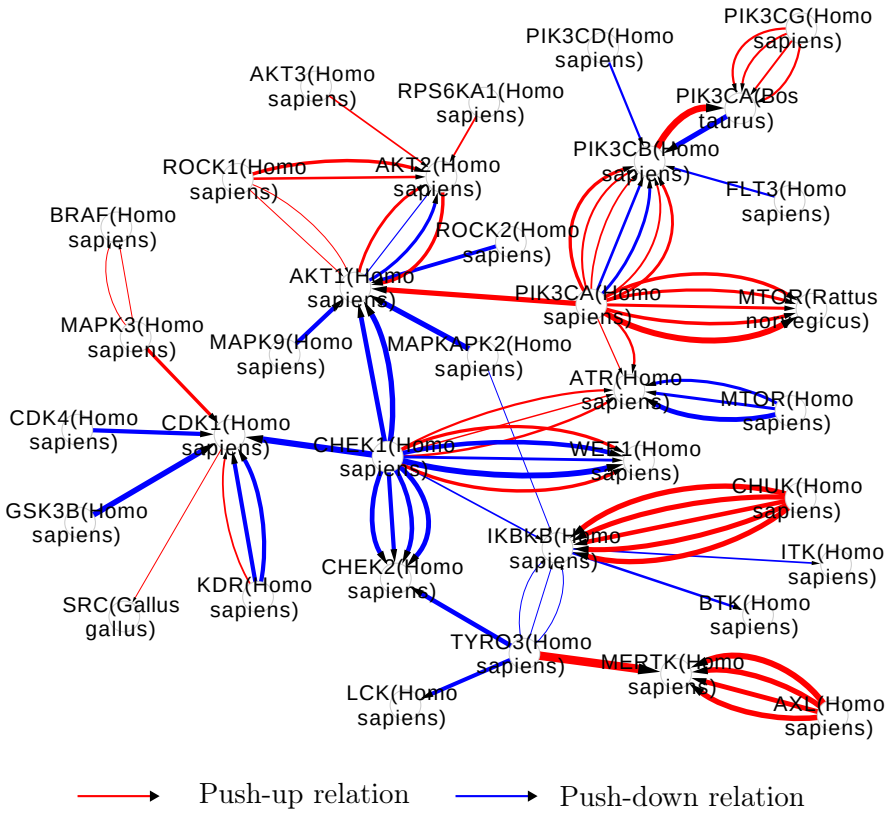


Figure 11: Push relation among bioassays

compound in B_k). A red edge from B_l to B_k represents that the corresponding pushed (up) compound is selective in B_k. A blue edge from B_l to B_k represents that the corresponding pushed (down) compound is x-selective in B_k. The weight (width) of an edge represents the corresponding push-up/down power. Figure 11 shows that there are many edges among genes of a same family (e.g., PIK3CA, PIK3CB, PIK3CD, PIK3CG; SSTR1, SSTR2, SSTR3, SSTR4, SSTR5). This well conforms to the Chemogenomics principle[41,42] that targets of a same family tend to bind to similar compounds. The full set of relations is available in Figure S1 in the Supporting Information.

The weighted directed networks are constructed based on push-up/down powers that are collectively determined by multiple compound prioritization models. Compared to conventional compound-sharing based networks[43,44] that are typically undirected and/or unweighted, such model-based weighted directed networks may exhibit interesting signals that

could inform novel drug development approaches. Further research may be oriented along this direction via better exploring the structures of the weighted directed networks.

8.2 Bioassay-Specific Compound Features

In dCPPP, the vector of Tanimoto similarities of c_i compared to other training compounds in a bioassay B is used as the compound features for c_i in B_k. Therefore, the compound features are task specific. This compound feature representation follows the idea of the very recent trend of learning task-specific compound features using deep learning[45–47] for various compound prediction problems. Thus, we will explore better compound feature learning for compound prioritization purposes.

8.3 Differential Promisucous Compound Prioritization

The x-selective compounds that are pushed down in dCPPP represent a certain type of promiscuous compounds, which are the promiscuous compounds that show multi-fold difference in their activitivities against an off-target and the target of interest (based on the definition of "selectivity" as in Section 3). This type of promiscuous compounds is much less preferable for the target of interest, compared to the other promiscuous compounds, which are active against multiple targets, but not very differentiably. In this work, we focus on pushing x-selective compounds down but not explicitly other promiscuous compounds. However, other promiscuous compounds should also be properly considered for pushing down as well. We will tackle this aspect in the future work.

Supporting Information Availability

Supporting Information Available: Assay information, push relation and additional experimental results are available in the Supporting Information. Detailed method description and results can be found at `https://cs.iupui.edu/~liujunf/projects/selRank_2017/`.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant Number IIS-1566219 and and IIS-1622526. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author Information

Corresponding Author

*E-mail: xning@iupui.edu. Phone: +1 317-278-3784.

ORCID

Xia Ning: 0000-0002-6842-1165

Notes

The authors declare no competing financial interest.

List of Journal Abbreviations

Journal	*Abbreviation*
ACS Central Science	*ACS Cent. Sci.*
BMC Bioinformatics	*BMC Bioinf.*
British Journal of Pharmacology	*Br. J. Pharmacol.*
Chemical Biology & Drug Design	*Chem. Biol. Drug Des.*
ChemMedChem	*ChemMedChem*
Computational and Structural Biotechnology Journal	*Comput. Struct. Biotechnol. J.*
Current Opinion in Chemical Biology	*Curr. Opin. Chem. Biol.*
Drug Discovery Today	*Drug Discovery Today*
Journal of Chemical Information and Computer Sciences	*J. Chem. Inf. Model.*
Journal of Chemical Information and Modeling	*J. Chem. Inf. Model.*
Journal of Health Economics	*J. Health Econ.*
Journal of Machine Learning Research	*J. Mach. Learn. Res.*
Knowledge and Information Systems	*Knowl. Inf. Syst.*
Nature	*Nature*
Nature Biotechnology	*Nat. Biotechnol.*
Nature Reviews Genetics	*Nat. Rev. Genet.*
Statistics in Medicine	*Stat. Med.*
Trends in Pharmacological Sciences	*Trends Pharmacol. Sci.*

References

(1) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The Price of Innovation: New Estimates of Drug Development Costs. *J. Health Econ.* **2003**, *22*, 151 – 185.

(2) Hansch, C.; Maolney, P. P.; Fujita, T.; Muir, R. M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178–180.

(3) Ashley, E. A. Towards Precision Medicine. *Nat. Rev. Genet.* **2016**, *17*, 507–522.

(4) Deng, X.; Nakamura, Y. Cancer Precision Medicine: From Cancer Screening to Drug Selection and Personalized Immunotherapy. *Trends Pharmacol. Sci.* **2016**, *38(1)*, 15–24.

(5) Geppert, H.; Vogt, M.; Bajorath, J. Current Trends in Ligand-Based Virtual Screening: Molecular Representations, Data Mining Methods, New Application Areas, and Performance Evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205–216, PMID: 20088575.

(6) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A Quantitative Analysis of Kinase Inhibitor Selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.

(7) Hu, Y.; Gupta-Ostermann, D.; Bajorath, J. Exploring Compound Promiscuity Patterns and Multi-Target Activity Spaces. *Comput. Struct. Biotechnol. J.* **2014**, *9*, 1–11.

(8) Peltason, L.; Hu, Y.; Bajorath, J. From Structure-Activity to Structure-Selectivity Relationships: Quantitative Assessment, Selectivity Cliffs, and Key Compounds. *ChemMedChem* **2009**, *4*, 1864–1873.

(9) Wassermann, A.; Geppert, H.; Bajorath, J. In *Chemoinformatics and Computational Chemical Biology*; Bajorath, J., Ed.; Application of Support Vector Machine-Based Ranking Strategies to Search for Target-Selective Compounds; Humana Press, **2011**; Vol. 672; pp 517–530.

(10) Lindström, A.; Pettersson, F.; Almqvist, F.; Berglund, A.; Kihlberg, J.; Linusson, A. Hierarchical PLS Modeling for Predicting the Binding of a Comprehensive Set of Structurally Diverse Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2006**, *46*, 1154–1167, PMID: 16711735.

(11) Weill, N.; Rognan, D. Development and Validation of a Novel Protein-Ligand Fingerprint to Mine Chemogenomic Space: Application to G Protein-Coupled Receptors and Their Ligands. *J. Chem. Inf. Model.* **2009**, *49*, 1049–1062, PMID: 19301874.

(12) Gönen, M.; Kaski, S. Kernelized Bayesian Matrix Factorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2047–2060.

(13) Nigsch, F.; Bender, A.; Jenkins, J. L.; Mitchell, J. B. O. Ligand-Target Prediction Using Winnow and Naive Bayesian Algorithms and the Implications of Overall Performance Statistics. *J. Chem. Inf. Model.* **2008**, *48*, 2313–2325.

(14) Ning, X.; Rangwala, H.; Karypis, G. Multi-Assay-Based Structure-Activity Relationship Models: Improving Structure-Activity Relationship Models by Incorporating Activity Information from Related Targets. *J. Chem. Inf. Model.* **2009**, *49*, 2444–2456.

(15) Chapelle, O., Schölkopf, B., Zien, A., Eds. *Semi-Supervised Learning*; MIT Press: Cambridge, MA, **2006**.

(16) Caruana, R. Multitask Learning. *Machine Learning* **1997**, *28*, 41–75.

(17) Kuncheva, L. I.; Whitaker, C. J. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning* **2003**, *51*, 181–207.

(18) Liu, J.; Ning, X. Multi-Assay-Based Compound Prioritization via Assistance Utilization: a Machine Learning Framework. *J. Chem. Inf. Model.* **2017**, *57*, 484–498.

(19) Dixon, S. L.; Villar, H. O. Bioactive Diversity and Screening Library Selection via Affinity Fingerprinting. *J. Chem. Inf. Model.* **1998**, *38*, 1192–1203.

(20) Bender, A.; Jenkins, J. L.; Glick, M.; Deng, Z.; Nettles, J. H.; Davies, J. W. "Bayes Affinity Fingerprints" Improve Retrieval Rates in Virtual Screening and Define Orthogonal Bioactivity Space: When are Multitarget Drugs a Feasible Concept? *J. Chem. Inf. Model.* **2006**, *46*, 2445–2456.

(21) Lessel, U. F.; Briem, H. Flexsim-X: a Method for the Detection of Molecules with Similar Biological Activity. *J. Chem. Inf. Model.* **2000**, *40*, 246–253.

(22) Stumpfe, D.; Geppert, H.; Bajorath, J. Methods for Computer-Aided Chemical Biology. Part 3: Analysis of Structure-Selectivity Relationships through Single-or Dual-Step Selectivity Searching and Bayesian Classification. *Chem. Biol. Drug Des.* **2008**, *71*, 518–528.

(23) Wassermann, A. M.; Geppert, H.; Bajorath, J. Searching for Target-Selective Compounds Using Different Combinations of Multiclass Support Vector Machine Ranking Methods, Kernel Functions, and Fingerprint Descriptors. *J. Chem. Inf. Model.* **2009**, *49*, 582–592, PMID: 19249858.

(24) Vogt, I.; Stumpfe, D.; Ahmed, H. E. A.; Bajorath, J. Methods for Computer-Aided Chemical Biology. Part 2: Evaluation of Compound Selectivity using 2D Molecular Fingerprints. *Chem. Biol. Drug Des.* **2007**, *70*, 195–205.

(25) Ning, X.; Walters, M.; Karypis, G. Improved Machine Learning Models for Predicting Selective Compounds. *J. Chem. Inf. Model.* **2012**, *52*, 38–50, PMID: 22107358.

(26) Li, H. *Learning to Rank for Information Retrieval and Natural Language Processing*; Synthesis Lectures on Human Language Technologies; Morgan & Claypool Publishers, **2011**.

(27) Cao, Z.; Qin, T.; Liu, T.-Y.; Tsai, M.-F.; Li, H. Learning to Rank: from Pairwise Approach to Listwise approach. *Proceedings of the 24th international conference on Machine learning*; **2007**; pp 129–136.

(28) Burges, C. J.; Ragno, R.; Le, Q. V. Learning to Rank with Nonsmooth Cost Functions. *Advances in Neural Information Processing Systems (NIPS)*. **2006**; pp 193–200.

(29) Lebanon, G.; Lafferty, J. Cranking: Combining Rankings using Conditional Probability Models on Permutations. *International Conference on Machine Learning*. **2002**; pp 363–370.

(30) Boyd, S.; Cortes, C.; Mohri, M.; Radovanovic, A. Accuracy at the Top. *Advances in Neural Information Processing Systems (NIPS)*. **2012**; pp 962–970.

(31) Agarwal, S.; The Infinite Push: A new Support Vector Ranking Algorithm that Directly Optimizes Accuracy at the Absolute Top of the List. *Proceedings of the 2011 SIAM International Conference on Data Mining*; **2011**; pp 839–850.

(32) Agarwal, S.; Dugar, D.; Sengupta, S. Ranking Chemical Structures for Drug Discovery: A New Machine Learning Approach. *J. Chem. Inf. Model.* **2010**, *50*, 716–731, PMID: 20387860.

(33) Jorissen, R. N.; Gilson, M. K. Virtual Screening of Molecular Databases Using a Support Vector Machine. *J. Chem. Inf. Model.* **2005**, *45*, 549–561, PMID: 15921445.

(34) Joachims, T. Optimizing Search Engines Using Clickthrough Data. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA, **2002**; pp 133–142.

(35) Rudin, C. The P-Norm Push: A Simple Convex Ranking Algorithm That Concentrates at the Top of the List. *J. Mach. Learn. Res.* **2009**, *10*, 2233–2271.

(36) Burges, C.; Shaked, T.; Renshaw, E.; Lazier, A.; Deeds, M.; Hamilton, N.; Hullender, G. Learning to Rank Using Gradient Descent. *Proceedings of the 22nd International Conference on Machine Learning*. New York, NY, USA, **2005**; pp 89–96.

(37) Wale, N.; Watson, I. A.; Karypis, G. Comparison of Descriptor Spaces for Chemical Compound Retrieval and Classification. *Knowl. Inf. Syst.* **2008**, *14*, 347–375.

(38) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Model.* **1998**, *38*, 983–996.

(39) Que, Q.; Belkin, M. Back to the Future: Radial Basis Function Networks Revisited. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. **2016**; pp 1375–1383.

(40) Harrell, F. E.; Lee, K. L.; Mark, D. B. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Stat. Med.* **1996**, *15*, 361–387.

(41) Caron, P. R.; Mullican, M. D.; Mashal, R. D.; Wilson, K. P.; Su, M. S.; Murcko, M. A. Chemogenomic Approaches to Drug Discovery. *Curr. Opin. Chem. Biol.* **2001**, *5*, 464–70.

(42) Klabunde, T. Chemogenomic Approaches to Drug Discovery: Similar Receptors Bind Similar Ligands. *Br. J. Pharmacol.* **2007**, *152*, 5–7.

(43) Chen, B.; Dong, X.; Jiao, D.; Wang, H.; Zhu, Q.; Ding, Y.; Wild, D. J. Chem2Bio2RDF: a Semantic Framework for Linking and Data Mining Chemogenomic and Systems Chemical Biology Data. *BMC Bioinf.* **2010**, *11*, 255.

(44) Hu, Y.; Bajorath, J. Compound Promiscuity: What Can We Learn from Current Data? *Drug Discovery Today* **2013**, *18*, 644 – 650.

(45) Altae-Tran, H.; Ramsundar, B.; Pappu, A. S.; Pande, V. Low Data Drug Discovery with One-Shot Learning. *ACS Cent. Sci.* **2017**, *3*, 283–293.

(46) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. Deep Neural Nets as a Method for Quantitative Structure-Activity Relationships. *J. Chem. Inf. Model.* **2015**, *55*, 263–274, PMID: 25635324.

(47) Coley, C. W.; Barzilay, R.; Green, W. H.; Jaakkola, T. S.; Jensen, K. F. Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction. *J. Chem. Inf. Model.* **2017**, *57*, 1757–1772, PMID: 28696688.

for Table of Contents use only