Wasserstein-Based Similarity Constrained Matrix Factorization for Drug-Drug Interaction Prediction

Sarah Malone[1], Mohammed Aburidi[1], and Roummel F. Marcia[1]

***Abstract*— We introduce a modification to Manifold Regularized Matrix Factorization for predicting drug-drug interactions, employing a Wasserstein similarity measure. Our findings demonstrate that this measure offers enhanced accuracy compared to previously tested similarity metrics. Notably, it exhibits resilience to variations in the latent feature vector size, denoted as k, enabling the projection of the drug-drug interaction matrix onto a significantly reduced dimensional space, thereby reducing computational costs. The proposed method holds significant implications for clinical practice by highlighting drug-drug combinations warranting further investigation for potential interactions.**

I. Introduction

As of August 2023, it was estimated that approximately 60% of the adult population in the United States had taken at least one prescription medication within the previous year, with 37% reporting the use of three or more in the same time period [1]. Given the large percentage of the population that takes prescription drugs, it is paramount for public health initiatives to understand potential drug interactions.

A significant hurdle in achieving this understanding lies in the timeline from a drug's conception to its approval, spanning nearly eight years on average from the initiation of clinical trials to final regulatory clearance, excluding the initial phases of drug development [2]. While clinical trials address some potential interactions, the focus tends to center on commonly co-prescribed medications [3]. Compounding this issue is the accelerated rate at which drugs are gaining approval. Between 2011 and 2018, the FDA's Center for Drug Evaluation and Research sanctioned an average of 26 drugs annually, compared to an average of 23 approvals per year from 2000 to 2010 [4]. This influx of medications makes it increasingly impractical to assess all possible drug combinations that a patient might encounter. Consequently, healthcare practitioners often only recognize interactions between frequently prescribed drugs or in cases where adverse effects are severe.

In response to these challenges, researchers have turned to data analytics and machine learning techniques to enhance the prediction of drug-drug interactions (DDIs). Leveraging graphical representations of drugs, these methods employ matrix completion algorithms to predict interactions based on drug features and similarities [5]–[8]. We aim to add to this class of understanding by proposing the Wasserstein distance as an alternate similarity measure to those already studied, as it allows us to look at features of a drug at the atomic level which ultimately provides greater insight into the drug itself. Specifically, we apply the Wasserstein distance to the Manifold Regularized Matrix Factorization method for drug interaction prediction, which has previously seen success at predicting DDIs.

*This work is supported by the Department of Defense Science, Mathematics, and Research for Transformation (SMART) Scholarship-for-Service Program and by NSF Grants IIS 1741490 and DMS 1840265.

[1]Sarah Malone, Mohammed Aburidi and Roummel Marcia are with the Department of Applied Mathematics, University of California Merced, Merced, California, USA.

(Corresponding author: Sarah Malone, email: sdowns@ucmerced.edu)

II. Methods

A. Drug molecules as probability measures

Graphs serve as a powerful depiction of small molecules, where atoms are nodes and chemical bonds are edges. This graph-based representation efficiently captures both the structural and chemical characteristics of these molecules. Our methodology of obtaining the similarity matrix unfolds in several stages: (1) converting each drug molecule into a set of node embeddings derived from its atomic properties; (2) measuring the Wasserstein distance between all pairs of graphs by solving the drug matching problem as an optimal transport problem; and (3) building the similarity matrix using the computed Wasserstein distance. By solving the graph matching of the chemical atomic-level and topological features of the drugs, we produce an informative similarity matrix that is used to enforce the drug similarity constraints on our matrix factorization optimization problem.

To formalize the attributed graph-matching problem, we consider undirected labeled graphs, denoted as tuples in the form: $\mathcal{G}(\mathcal{V}, \mathcal{E}, l_f)$. Here, $(\mathcal{V}, \mathcal{E})$ represents the set of vertices and edges within the graph. The function l_f assigns each vertex $v_i \in \mathcal{V}$ a feature vector $a_i = l_f(v_i)$ in a specific feature metric space. In our specific application, this feature vector captures atomic properties in a chemical compound. Following [9], [10], we enhance the graph by introducing a histogram to convey the relative significance of its vertices. In this enhancement, assuming the graph comprises N vertices, individual weights h_i are assigned to each vertex. Consequently, our graph takes the form $\mathcal{G}(\mathcal{V}, \mathcal{E}, l_f, h_\mathcal{G})$, where $h_\mathcal{G}$ is a function associating a weight with each vertex with $h_i = h_\mathcal{G}(v_i)$. This definition enables us to represent the graph as a probability measure with comprehensive support across the feature space. In scenarios where all weights are equal ($h_i = \frac{1}{N}$), every vertex holds an equivalent degree of relative importance.

Our objective is to define a matching distance metric between two graphs (i.e., drug molecules), labeled $\mathcal{G}_1$ and $\mathcal{G}_2$, each possessing N and K vertices (atoms), respectively.

2024 IEEE Workshop on Signal Processing Systems (SiPS) | 979-8-3503-7375-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/SiPS62058.2024.00017

2374-7390/24/$31.00 ©2024 IEEE
DOI 10.1109/SiPS62058.2024.00017

These graphs are distinguished by their corresponding probability measures, $h_{\mathcal{G}_1}$ and $h_{\mathcal{G}_2}$. Subsequently, we proceed to calculate the pairwise distances between the drugs using optimal transport. This begins with the computation of ground distances for each node pair.

B. Wasserstein distance on drug molecules

Optimal transport, as outlined in [11], offers a mathematical paradigm tailored to solving the challenge of finding the most efficient way of relocating objects, often probability distributions, from one configuration to another while minimizing costs. Efficiency, in this context, pertains to achieving such transfers at minimal expense.

To mathematically formulate the transportation problem in the context of graphs, consider two sets of nodes, denoted as $\mathcal{X} = \{x_j\}_{j=1}^{N}$ representing the source samples, which correspond to the nodes of the first graph (representing the molecules of one drug), and $\mathcal{Y} = \{y_i\}_{i=1}^{K}$, corresponding to the target nodes of the second graph (representing a second drug). We introduce two discretized distributions of interest, denoted as $\mathbf{p} \in \mathcal{H}_N$ and $\mathbf{q} \in \mathcal{H}_K$. Here, $\mathcal{H}_N$ and $\mathcal{H}_K$ signify histograms with N and K bins, respectively. The vector $\mathbf{p}$ serves to convey the relative significance of the vertices (atoms) within the first drug, and the vector $\mathbf{q}$ for the second drug. These distributions adhere to the following constraints: $\mathbf{p} \in \mathbb{R}_+^N$, indicating that each element of $\mathbf{p}$ is a non-negative real number, and the sum of its elements equals 1, i.e., $\sum_j p_j = 1$. Similarly, $\mathbf{q} \in \mathbb{R}_+^K$ with the constraint $\sum_i q_i = 1$.

Define $Q \in \mathbb{R}+^{K\times N}$ as the transportation plan or coupling matrix, with elements denoted as Q_{ij}. This matrix precisely quantifies the amount situated at the source x_j, having a total mass of p_j, which necessitates transportation to the target location y_i, possessing a total mass of q_i. From a linear algebra perspective, the entries in the j^{th} column of Q represent the amounts from the j^{th} source to be transported to the targets, while the entries in the i^{th} row signify the amounts from the sources transported to the i^{th} target. Mathematically,

$$\begin{aligned} \sum_{j=1}^{N} Q_{ij} &= q_i \qquad \text{for all} \quad i \in \{1,..,K\}, \\ \sum_{i=1}^{K} Q_{ij} &= p_j \qquad \text{for all} \quad j \in \{1,..,N\}. \end{aligned} \tag{1}$$

The transportation process involves a cost matrix C, where C_{ij} represents the geometric distance between x_i and y_j. The total cost related to a transport plan is defined as follows:

$$\langle C, Q \rangle_F = \sum_{i=1}^{K} \sum_{j=1}^{N} C_{ij} Q_{ij} \tag{2}$$

where $\langle \cdot, \cdot \rangle$ is the Frobenius dot-product of two matrices. The optimal transport is thus formulated as the following optimization problem:

$$\begin{aligned} \underset{Q\in\mathbb{R}^{K\times N}}{\text{minimize}} \quad & \langle C, Q \rangle_F \\ \text{subject to} \quad & \sum_{j=1}^{N} Q_{ij} = q_i \quad \text{for all} \quad i \in \{1,..,K\} \\ & \sum_{i=1}^{K} Q_{ij} = p_j \quad \text{for all} \quad j \in \{1,..,N\} \\ & Q_{ij} \geq 0 \quad \text{for all} \quad (i,j) \in \{1,..,K\} \times \{1,..,N\} \end{aligned} \tag{3}$$

In each column, the transportation process must precisely allocate the mass p_i to achieve a perfect match with the intended mass q_j. This optimization problem can be expressed more succinctly as follows:

$$\begin{aligned} \underset{Q\in\mathbb{R}^{K\times N}}{\text{minimize}} \quad & \langle C, Q \rangle_F \\ \text{subject to} \quad & Q\mathbf{1}_N = \mathbf{q}, \qquad Q^T\mathbf{1}_K = \mathbf{p}, \qquad Q \geq 0, \end{aligned} \tag{4}$$

where $\mathbf{1}_N$ and $\mathbf{1}_K$ represent vectors of ones in dimension N and K, respectively. We define the set of all admissible couplings or transport plans $\mathbf{Q}(\mathbf{p}, \mathbf{q})$ between histograms as follows:

$$\mathbf{Q}(\mathbf{p}, \mathbf{q}) = \{Q \in \mathbb{R}_+^{K\times N} \mid Q\mathbf{1}_N = \mathbf{q}, Q^T\mathbf{1}_K = \mathbf{p}\}.$$

In a specific scenario where the cost function C aligns with a distance matrix, the optimal transport problem leads to the Wasserstein distance on $\mathcal{H}_N \times \mathcal{H}_K$. The Wasserstein distance is defined as:

$$W(\mathbf{p}, \mathbf{q}) = \underset{Q\in\mathbf{Q}(\mathbf{p},\mathbf{q})}{\text{minimize}} \ \langle C, Q \rangle_F. \tag{5}$$

We employ the simplex method [12] to solve (4), which iteratively moves from one feasible solution to another along the edges of the feasible region until an optimal solution is attained. The method employs a tableau representation, with each iteration selecting a pivot column and row for tableau update. The algorithm iterates until optimality conditions are satisfied.

We solve (4) for each pair of drugs and obtain the distance matrix $D \in \mathbb{R}^{m\times m}$, where m is the number of drugs and where the entries are given by (5). We define our similarity matrix S as follows:

$$S = \frac{1}{\max(D)}\big(\max(D)\mathbf{1}_{m\times m} - D\big) \tag{6}$$

where $\max(D)$ is the maximum element in the matrix D and $\mathbf{1}_{m\times m}$ is the matrix of ones of size $m \times m$. Note that the entries s_{ij} of S are between 0 and 1. Also, the larger the distance, i.e., the entry in D, the closer the corresponding entry in S is to 0, and vice versa.

C. Similarity-Constrained Matrix Factorization

Our implementation follows the methodology introduced by Zhang [13]. We assume we have a DDI matrix, A, which is an $m \times m$ symmetric matrix with zeroes on the diagonal. The elements of A indicate whether or not an interaction is known between each drug, with a value of 1 indicating a

known interaction while 0 indicates either no interaction or no known interaction.

Once A and S are known, the goal is to decompose A into two low-rank matrices $X \in \mathbb{R}^{m\times k}$, $Y \in \mathbb{R}^{m\times k}$ such that $A = XY^T$, where k is the dimension of a latent feature vector. The objective function we consider in this work is

$$\begin{aligned} L(X,Y) &= \frac{1}{2}\|A - XY^T\|_F^2 + \frac{\lambda}{2}\left(\|X\|_F^2 + \|Y\|_F^2\right) \qquad (7)\\ &= \frac{1}{2}\sum_{i,j}\left(a_{ij} - \vec{x}_i\vec{y}_j^T\right)^2 + \frac{\lambda}{2}\sum_i\left(\|\vec{x}_i\|_2^2 + \|\vec{y}_j\|_2^2\right) \end{aligned}$$

where the first term in (7) is a fitting term and the second is a regularization term to make the problem well-posed. Here $\|\cdot\|_F$ is the Frobenius norm, $\|\cdot\|_2$ is the Euclidean norm, $\vec{x}_i$ is the ith row of X, $\vec{y}_j$ is the jth row of Y, and $\lambda > 0$ is the Tikhonov regularization parameter.

This method assumes that drug-drug similarities can be represented by manifolds in the drug feature space, meaning that each drug feature has a neighborhood that maps to the drug similarity. By making this assumption, we can utilize the fact that nearby points in the drug feature space will likely have a similar mapping to their drug similarity. Therefore, the objective function L can then be regularized over the manifolds by

$$L_{reg}^{row}(X) = \frac{1}{2}\sum_{i,j} s_{ij}\|\vec{x}_i - \vec{x}_j\|_2^2 \qquad (8)$$

$$L_{reg}^{col}(Y) = \frac{1}{2}\sum_{i,j} s_{ij}\|\vec{y}_i - \vec{y}_j\|_2^2 \qquad (9)$$

where s_{ij} is the similarity between the ith and jth drugs. Finally, we attain the Manifold Regularized Matrix Factorization (MRMF) [13] by combining (7) with (8) and (9) resulting in the following optimization problem:

$$\begin{aligned} &\underset{X,Y\in\mathbb{R}^{m\times k}}{\text{minimize}}\ L_{mix}(X,Y)\\ &\qquad = L(X,Y) + \mu\left(L_{reg}^{row}(X) + L_{reg}^{col}(Y)\right), \end{aligned} \qquad (10)$$

where $\mu > 0$ is a manifold regularization parameter.

To find the minimizer of (10), we utilize an alternating descent method in [13]. This method is equivalent to a gradient descent where $\vec{x}_i$ and $\vec{y}_j$ are successively fixed. We start by randomly initializing $\vec{x}_i$ and $\vec{y}_j$. Then, we use Newton's method to compute each update to $\vec{x}_i$ and $\vec{y}_j$:

$$\vec{x}_i \leftarrow \vec{x}_i - \nabla_{\vec{x}_i}L_{mix}\left(\nabla^2_{\vec{x}_i}L_{mix}\right)^{-1} \qquad (11)$$

$$\vec{y}_j \leftarrow \vec{y}_j - \nabla_{\vec{y}_j}L_{mix}\left(\nabla^2_{\vec{y}_j}L_{mix}\right)^{-1} \qquad (12)$$

where $\nabla_{\vec{x}_i}L_{mix}$ and $\nabla_{\vec{y}_j}L_{mix}$ are the gradients of L_{mix} with respect to $\vec{x}_i$ and $yvec_j$, respectively, and $\nabla^2_{\vec{x}_i}L_{mix}$ and $\nabla^2_{\vec{y}_j}L_{mix}$ are the Hessians of L_{mix} with respect to $\vec{x}_i$ and $yvec_j$, respectively. The updates (11) and (12) are continued until convergence.

III. Numerical Experiments

A. Data

As pharmaceutical data becomes increasingly accessible, researchers have turned to curated databases for comprehensive drug information. TWOSIDEs, for instance, aggregates reported drug interactions without attribution to a specific drug [14]. Complementary feature data, encompassing side effects, off-label effects, and chemical compositions, are drawn from established repositories such as SIDER [15], OFFSIDES [14], PubChem [16], and DrugBank [17]. By linking drugs with known interactions to those with detailed feature information, a comprehensive dataset is formed. This mapping, compiled by Zhang [18], identifies 548 drugs with both interaction data and detailed features. The dataset, including the drug-drug interaction matrix, A, is publicly accessible via Github at: `https://github.com/zw9977129/drug-drug-interaction/tree/master`.

The key feature utilized to construct our similarity matrix is the Simplified Molecular-Input Line-Entry System (SMILES) string associated with each drug. SMILES leverages molecular graph theory to provide atomic-level feature representation for drugs [19]. However, it's worth noting that four drugs in the original dataset lack SMILES indices, resulting in a final dataset comprising 544 drugs with a total of 47,537 known interactions.

To evaluate the accuracy of our method, 10% of the known interactions are randomly withheld. Subsequently, our objective is to accurately reconstruct the original drug-drug interaction matrix A.

B. Parameter Settings

The three parameters in the MRMF method that can be altered are (i) the latent feature vector size, k, (ii) the Tikhonov regularization parameter, λ, and (iii) the manifold regularization parameter, μ. Several experiments were run altering these parameters to find the optimal parameter setting. Zhang [13] found $k = 0.08 * m$, $\lambda = 2^2$, and $\mu = 2^2$ to produce the best result, so we initially started with this combination. In their study, this was found to be the best combination of parameters, regardless of the similarity metric used. Our results show that this is not an optimal parameter combination when using the Wasserstein distance metric, as it predicts that no drug interactions are present.

We then swept over the parameter space field to find the optimal settings. Our first sweep held $k = 0.1 * m$ while looping over $\lambda, \mu \in J = \{2^\kappa : \kappa \in -4, -3, -2, \cdots, 2, 3\}$. Next, we conducted four variations of an experiment where μ and λ are held constant while looping over $k = 0.01\beta * m$, where $\beta \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$. In the first of these we set $\mu = \lambda = 2^2$, in the second $\mu = \lambda = 2^{-2}$, in the third $\mu = \lambda = 2^{-4}$, and finally in the fourth $\mu = \lambda = 2^{-6}$.

Of these experiments, the greatest results were achieved when $k = 0.1 * m$ and $\mu = \lambda = 2^{-6}$. The results presented from here on will be based on these parameters.

Ultimately, these experiments found that the value for the latent feature vector, k, did not have as great of an impact on

Similarity	AUPRC	AUC	Recall	Precision	Accuracy	F-score
Jaccard	0.7958	0.9585	0.6812	0.7722	0.9546	0.7237
Cosine	0.7958	0.9584	0.6906	0.7729	0.9546	0.7237
Gauss	0.7959	0.9584	0.6806	0.7726	0.9546	0.7236
Wasserstein	0.9449	0.9690	0.7777	0.9292	0.9095	0.8467

TABLE I

ACCURACY METRICS FOR VARIOUS SIMILARITY METRICS. THE FIRST THREE ROWS COME FROM ZHANG [13] WITH $\mu = \lambda = 2^2$, $k = 0.8 * m$, WHILE THE LAST ROW COMES FROM OUR APPROACH (USING THE WASSERSTEIN DISTANCE) WITH $\mu = \lambda = 2^{-6}$, $k = 0.1 * m$.

the results as the regularization parameters μ and λ. When these regularization parameters are properly tuned, nearly identical results are achieved, regardless of k. This allows us to project a large DDI matrix into a much smaller space, saving computational cost. In our numerical experiments, we found that $k = 0.1 * m$ is sufficient.

IV. RESULTS AND DISCUSSION

A. Evaluation Metrics

To analyze the performance of our method, we compare recall, precision, accuracy, and F-score values:

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{13}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{15}$$

$$\text{F-Score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \tag{16}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Additionally, we utilize the area under the ROC Curve (AUC), which is a metric of the true positives vs false positive rates, and the area under the precision-recall curve (AUPRC), which measures the true positives to predicted positives.

Of these metrics, the AUPRC takes precedence in assessing method performance, as it directly measures the ratio of correctly predicted true positives to all predicted positives. This has been chosen as the leading performance metric as in clinical settings where this may be applied, the ability to correctly predict true positives is highly valued, and therefore it is acceptable to have a slight decrease in overall performance if this measure is being met. However, the remaining metrics do still provide supplementary insight into model performance.

B. Comparison with Other Similarity Measures

We conducted a comparative analysis between the results obtained using the Wasserstein distance to generate our similarity matrix and those reported by Zhang [13], as summarized in Table I. While the Wasserstein distance looks at features of a drug on an atomic level, the Jaccard similarity, Cosine similarity, and Gauss similarity each look at the drug features holistically. As such, these measures look to understand how similar two drugs may be in terms of their overall composition and effects but may miss key information that can only be seen when exploring the chemical structure of a drug. Each of these are common measures in bioinformatics problems and therefore serve as a basis to compare our proposed Wasserstein-based similarity measure. Zhang's study indicated that, under their optimal parameters, similar accuracy was achieved regardless of the similarity metric employed, suggesting that the accuracy of the Manifold Regularized Matrix Factorization (MRMF) method was independent of the similarity measure utilized. In contrast, our approach using the Wasserstein similarity measure yielded a significantly higher AUPRC of 0.9449. This outcome underscores the effectiveness of our metric in accurately identifying potential drug-drug interactions.

It's important to note that our measure's high AUPRC primarily reflects its ability to predict true positives, while having slightly lower performance with regards to predicting true negatives. We surmise that this is because the Wasserstein distance captures the similarity between features that drive drug interaction, while excluding the features that do not. This may result in a lower overall accuracy score, however, in practical application, the focus lies on identifying potential interactions for further investigation, thus prioritizing true and false positives over true negatives.

V. CONCLUSION

This study proposes an alternative similarity metric for similarity-constrained drug-drug interaction prediction. Our findings demonstrate that the Wasserstein similarity measure outperforms previously studied metrics when integrated with the MRMF method for DDI prediction, particularly in terms of true positive prediction rates.

REFERENCES

[1] A. Kirzinger, A. Montero, G. Sparks, I. Valdes, and L. H. Published, "Public Opinion on Prescription Drugs and Their Prices," Aug. 2023. [Online]. Available: https://www.kff.org/health-costs/poll-finding/public-opinion-on-prescription-drugs-and-their-prices/

[2] J. J. Darrow, J. Avorn, and A. S. Kesselheim, "FDA Approval and Regulation of Pharmaceuticals, 1983-2018," *JAMA*, vol. 323, no. 2, p. 164, Jan. 2020. [Online]. Available: https://jamanetwork.com/journals/jama/fullarticle/2758605

[3] B. Percha and R. B. Altman, "Informatics confronts drug–drug interactions," *Trends in Pharmacological Sciences*, vol. 34, no. 3, pp. 178–184, Mar. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165614713000072

[4] A. Batta, B. S. Kalra, and R. Khirasaria, "Trends in FDA drug approvals over last 2 decades: An observational study," *Journal of Family Medicine and Primary Care*, vol. 9, no. 1, p. 105, Jan. 2020. [Online]. Available: https://journals.lww.com/jfmpc/fulltext/2020/09010/trends_in_fda_drug_approvals_over_last_2_decades_.20.aspx

[5] A. Gottlieb, G. Y. Stein, Y. Oron, E. Ruppin, and R. Sharan, "INDI: a computational framework for inferring drug interactions and their associated recommendations," *Molecular Systems Biology*, vol. 8, p. 592, Jul. 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3421442/

[6] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, and C. Friedman, "Drug–drug interaction through molecular structure similarity analysis," *Journal of the American Medical Informatics Association : JAMIA*, vol. 19, no. 6, pp. 1066–1074, 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3534468/

[7] S. Vilar, E. Uriarte, L. Santana, N. P. Tatonetti, and C. Friedman, "Detection of Drug-Drug Interactions by Modeling Interaction Profile Fingerprints," *PLOS ONE*, vol. 8, no. 3, p. e58321, Mar. 2013, publisher: Public Library of Science. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058321

[8] P. Li, C. Huang, Y. Fu, J. Wang, Z. Wu, J. Ru, C. Zheng, Z. Guo, X. Chen, W. Zhou, W. Zhang, Y. Li, J. Chen, A. Lu, and Y. Wang, "Large-scale exploration and analysis of drug combinations," *Bioinformatics (Oxford, England)*, vol. 31, no. 12, pp. 2007–2016, Jun. 2015.

[9] M. Aburidi and R. Marcia, "Wasserstein Distance-Based Graph Kernel for Enhancing Drug Safety and Efficacy Prediction," in *2024 IEEE First IEEE International Conference on AI for Medicine, Health, and Care (AIMHC)*, 2024.

[10] ——, "Optimal Transport-Based Graph Kernels for Drug Property Prediction," in *Scientific Reports*, 2024.

[11] F. Santambrogio, *Optimal Transport for Applied Mathematicians*. Birkhäuser, 2015.

[12] M. K. Wood and G. B. Dantzig, "Programming of Interdependent Activities: I General Discussion," *Econometrica*, vol. 17, no. 3/4, pp. 193–199, 1949, publisher: [Wiley, Econometric Society]. [Online]. Available: https://www.jstor.org/stable/1905522

[13] W. Zhang, Y. Chen, D. Li, and X. Yue, "Manifold regularized matrix factorization for drug-drug interaction prediction," *Journal of Biomedical Informatics*, vol. 88, pp. 90–97, Dec. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046418302144

[14] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, "Data-Driven Prediction of Drug Effects and Interactions," *Science Translational Medicine*, vol. 4, no. 125, pp. 125ra31–125ra31, Mar. 2012, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/full/10.1126/scitranslmed.3003377

[15] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, "The SIDER database of drugs and side effects," *Nucleic Acids Research*, vol. 44, no. D1, pp. D1075–1079, Jan. 2016.

[16] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, "PubChem 2023 update," *Nucleic Acids Research*, vol. 51, no. D1, pp. D1373–D1380, Jan. 2023. [Online]. Available: https://doi.org/10.1093/nar/gkac956

[17] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, and M. Wilson, "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1074–D1082, Jan. 2018.

[18] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, and X. Li, "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data," *BMC Bioinformatics*, vol. 18, no. 1, p. 18, Jan. 2017. [Online]. Available: https://doi.org/10.1186/s12859-016-1415-9

[19] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *Journal of Chemical Information and Computer Sciences*, vol. 28, no. 1, pp. 31–36, Feb. 1988, publisher: American Chemical Society. [Online]. Available: https://doi.org/10.1021/ci00057a005