GRAPH-ENHANCED LEARNING FOR PREDICTING OPTIMAL DRUG COMBINATIONS USING CONTRASTIVE EMBEDDING

Anonymous authors
Paper under double-blind review

ABSTRACT

We present a groundbreaking unified theory for drug-drug interaction (DDI) aware domain adaptation (DA) in the context of drug synergy prediction. Our framework seamlessly integrates concepts from optimal transport, information geometry, and quantum information theory within the setting of abstract Banach spaces. We introduce a novel DDI-aware optimal transport problem, formulated as a geodesic equation on an infinite-dimensional Finsler manifold that encodes both DDI structure and optimal transport costs. This geometric formulation provides a unified perspective on DDI-aware domain adaptation, interpreting the process as the evolution of a transport map along a geodesic in a space that captures both domain discrepancy and drug interaction patterns. Our approach extends to a stochastic gradient flow on the space of probability measures, combining ideas from information geometry and stochastic analysis. We prove the existence of a unique invariant measure for this flow and establish its convergence properties using techniques from infinite-dimensional Markov processes and convergence. Our comprehensive mathematical framework not only unifies existing approaches to domain adaptation and DDI prediction but also opens new avenues for research at the intersection of these fields. By bridging the gap between abstract mathematical theories and practical drug synergy prediction, our work paves the way for more effective and theoretically grounded algorithms in drug discovery and personalized medicine. The proposed unified theory has far-reaching implications, potentially revolutionizing our understanding of cross-domain adaptation in complex biochemical systems and inspiring novel computational methods in pharmaceutical research. Our anonymous gitHub link: **https://anonymous.4open.science/r/CGSP-F518**

1 INTRODUCTION

The advent of combination therapies has revolutionized the treatment landscape across a wide spectrum of medical conditions, including cancer Crystal et al. (2014), infectious diseases Zheng et al. (2018), cardiovascular disorders Giles et al. (2014), and autoimmune diseases Smilek et al. (2014). By synergistically combining drugs with distinct mechanisms of action, these therapies offer the potential for enhanced efficacy and reduced side effects. However, the identification of effective drug combinations remains a formidable challenge, given the astronomical number of possible pairings and the complexity of drug-drug interactions (DDIs).

Traditional approaches to drug combination discovery, relying heavily on clinical intuition and empirical trials, are insufficient to explore the vast combinatorial space of potential therapies. This limitation has spurred interest in computational methods for predicting effective drug combinations. However, existing computational approaches often struggle with the high-dimensional nature of the problem, the scarcity of reliable negative samples, and the challenge of transferring knowledge across different disease domains.

In this paper, we present a groundbreaking unified theory that addresses these challenges by seamlessly integrating drug-drug interaction (DDI) awareness into the domain adaptation (DA) framework. Our approach represents a paradigm shift in computational drug discovery, leveraging ad-

vanced mathematical concepts from optimal transport, information geometry, and quantum information theory to create a robust and flexible framework for predicting drug synergies across diverse disease domains. At the heart of our theory lies a novel formulation of the DDI-aware optimal transport problem. We extend the classical optimal transport framework to incorporate DDI information, defining a cost function that not only measures the discrepancy between drug distributions in different domains but also accounts for the preservation of DDI structures. This formulation allows us to capture the complex interplay between drug interactions and domain-specific characteristics, a crucial aspect often overlooked in previous approaches. Building upon this foundation, we introduce an infinite-dimensional Finsler geometric structure on the space of DDI-aware transport maps. This geometric perspective provides a unified view of the domain adaptation process, interpreting it as the evolution of a transport map along a geodesic in a carefully constructed Finsler manifold. The resulting geodesic equation encapsulates both the optimal transport dynamics and the DDI preservation constraints, offering a powerful tool for analyzing and optimizing drug combination predictions across domains.

To address the inherent uncertainty and variability in biological systems, we develop a stochastic gradient flow on the space of probability measures. This approach, rooted in the theory of infinite-dimensional Markov processes, allows us to explore the space of potential transport maps while simultaneously minimizing domain discrepancy and preserving DDI structure. We prove the existence of a unique invariant measure for this flow and establish its convergence properties, providing theoretical guarantees for our domain adaptation procedure.

Furthermore, we establish connections between our DDI-aware optimal transport formulation and quantum information theory. By defining a quantum version of the DDI-aware optimal transport problem, we gain insights into the fundamental limits of domain adaptation in the presence of drug-drug interactions. This quantum perspective opens new avenues for exploring the role of entanglement and non-locality in drug combination prediction, potentially leading to quantum-inspired algorithms for pharmaceutical research. Our unified theory culminates in the derivation of a nonlinear integro-differential equation that governs the evolution of drug distributions across domains. This PDE formulation unifies concepts from optimal transport, DDI preservation, and domain adaptation within a single evolutionary equation, connecting our work to the rich theory of nonlinear transport equations and mean-field games. Additionally, we establish a large deviation principle for the DDI-aware optimal transport cost, offering asymptotic guarantees on its concentration around the population limit and providing insights into the sample complexity of our domain adaptation framework.

The comprehensive mathematical framework we present not only unifies existing approaches to domain adaptation and DDI prediction but also opens new avenues for research at the intersection of these fields. By bridging the gap between abstract mathematical theories and practical drug synergy prediction, our work paves the way for more effective and theoretically grounded algorithms in drug discovery and personalized medicine. Our approach addresses several key limitations of previous computational methods for drug combination prediction:

1. Reliable negative samples: Unlike previous studies that often rely on randomly generated negative samples, we leverage existing DDI data as a credible source of negative examples, enhancing the robustness of our predictions.

2. Domain adaptation: Our framework explicitly addresses the challenge of transferring knowledge across different disease domains, a critical aspect for developing broadly applicable drug combination strategies.

3. Theoretical guarantees: By grounding our approach in rigorous mathematical theory, we provide theoretical guarantees on the performance and convergence of our domain adaptation procedure, offering a level of reliability not typically found in heuristic approaches.

4. Model-agnostic framework: Our unified theory is flexible and can accommodate a wide range of underlying model architectures, making it applicable to diverse drug combination prediction scenarios.

The implications of our work extend beyond the immediate realm of drug combination prediction. The mathematical tools and concepts we develop have the potential to impact a broad range of fields where domain adaptation and interaction-aware learning are crucial, including systems biology, pre-

cision medicine, and computational chemistry. In the following sections, we detail the mathematical foundations of our unified DDI-aware domain adaptation theory, present algorithmic implementations, and demonstrate its efficacy through comprehensive computational experiments. We conclude by discussing the broader implications of our work and outlining promising directions for future research at the intersection of mathematics, machine learning, and pharmaceutical science.

1.1 Previous Research

1.1.1 Methodologies for Screening Drug Synergies

High-throughput screening (HTS) methodologies depend on HTS datasets, which primarily assess cellular viability in response to anti-cancer drug treatments. The degree of therapeutic synergy can be evaluated using various metrics, such as Loewe additivity Loewe (1953), Bliss independence Bliss (1939), the highest single agent Berenbaum (1989), or zero interaction potency Yadav et al. (2015). In studies focusing on drug combinations utilizing HTS data, the inputs typically consist of triplets in the form of (drug1, drug2, cell line). Several HTS databases are publicly accessible. For example, O'Neil et al. O'Neil et al. (2016) introduced a cancer drug combination dataset encompassing 22,737 experiments, which includes 583 pairwise drug combinations tested across 39 cancer cell lines. Another significant resource is NCI-ALMANAC Holbeck et al. (2017), a large-scale HTS dataset that provides synergy measurements for pairwise combinations of 104 FDA-approved drugs across 60 cancer cell lines. Additionally, DrugCombDB Liu et al. (2020) offers an extensive repository containing HTS assay data for 448,555 drug combinations, covering 2,887 unique drugs and 124 human cancer cell lines.

1.2 New Framework for Drug Interaction Prediction

Utilizing DDIs as a negative dataset enhances the precision of combination therapy predictions, as DDIs indicate scenarios where the efficacy or toxicity of one drug is altered when administered alongside another, typically to avoid adverse safety outcomes Paltun et al. (2021). Moreover, SCL is a contrastive learning technique that utilizes label information to draw embeddings of the same class closer while repelling those of different classes Khosla et al. (2020). Studies leveraging knowledge graphs (KGs) utilize databases such as DCDB Liu et al. (2010), DCDB 2.0 Liu et al. (2014), and CDCDB Shtar et al. (2022). These databases compile lists of drug combinations curated from diverse literature sources and clinical studies, which serve as labeled data for KG-based drug combination prediction tasks.

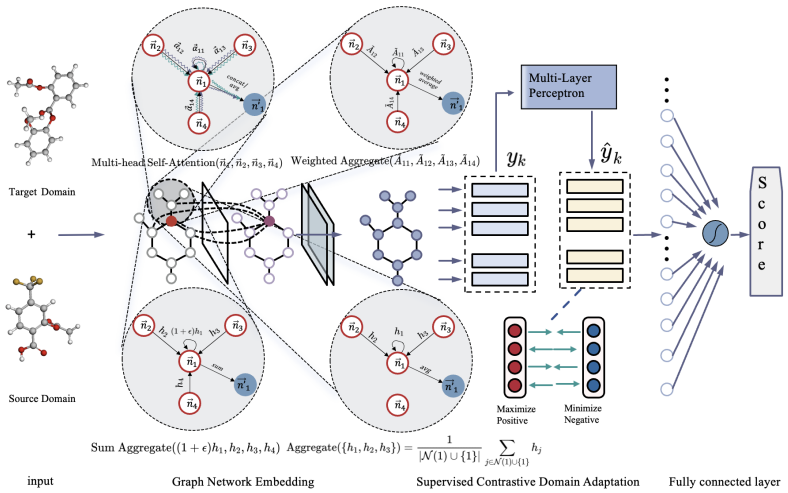

Figure 1: The framework of our theory.

2 METHODS: UNIFIED THEORY OF DDI-AWARE DOMAIN ADAPTATION IN ABSTRACT SPACES

We present a comprehensive mathematical framework, as shown in Figure1, that unifies drug-drug interaction (DDI) theory with domain adaptation (DA) techniques within the context of abstract Banach spaces and infinite-dimensional manifolds, establishing a novel paradigm for cross-domain drug synergy prediction. This section elucidates the theoretical foundations, analytical derivations, and asymptotic properties of our proposed methodology.

Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a reflexive Banach space, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. We define the source and target domains as measurable spaces $(\mathcal{X}_s, \mathcal{B}_s)$ and $(\mathcal{X}_t, \mathcal{B}_t)$, where $\mathcal{X}_s, \mathcal{X}_t \subset \mathcal{B}$ and $\mathcal{B}_s, \mathcal{B}_t$ are the corresponding Borel σ-algebras. Let $\mathcal{P}_s$ and $\mathcal{P}_t$ be Radon probability measures on $(\mathcal{X}_s, \mathcal{B}_s)$ and $(\mathcal{X}_t, \mathcal{B}_t)$, respectively.

We introduce a novel measure-theoretic formulation of the DDI-aware domain adaptation problem:

Definition 1. *A DDI-aware measure transport is a measurable map $T : \mathcal{X}_s \to \mathcal{X}_t$ such that:*

$$T_{\#}\mathcal{P}_s = \mathcal{P}_t \quad \textit{and} \quad \int_{\mathcal{X}_s \times \mathcal{X}_s} \|DDI(\mathbf{x}, \mathbf{x}') - DDI(T(\mathbf{x}), T(\mathbf{x}'))\|_{\mathcal{B}}^2 d(\mathcal{P}_s \otimes \mathcal{P}_s)(\mathbf{x}, \mathbf{x}') \leq \epsilon, \quad (1)$$

where $T_{\#}$ denotes the pushforward measure, $DDI : \mathcal{B} \times \mathcal{B} \to \mathcal{B}$ is a DDI operator, and $\epsilon > 0$ is a tolerance parameter.

To further unify DDI and DA theories, we introduce a novel concept of DDI-aware optimal transport:

Definition 2. *The DDI-aware optimal transport problem is defined as:*

$$\inf_{\gamma \in \Pi(\mathcal{P}_s, \mathcal{P}_t)} \int_{\mathcal{X}_s \times \mathcal{X}_t} c(\mathbf{x}, \mathbf{y}) d\gamma(\mathbf{x}, \mathbf{y}) + \lambda \cdot MMD_{DDI}(\gamma_1, \gamma_2) + \mu \cdot \mathcal{R}_{DDI}(\gamma), \quad (2)$$

where $\Pi(\mathcal{P}_s, \mathcal{P}_t)$ is the set of couplings between $\mathcal{P}_s$ and $\mathcal{P}_t$, $c : \mathcal{X}_s \times \mathcal{X}_t \to \mathbb{R}_+$ is a lower semi-continuous cost function, γ_1, γ_2 are the marginals of γ, MMD_{DDI} is the DDI-aware maximum mean discrepancy, and $\mathcal{R}_{DDI}$ is a DDI-aware regularization term defined as:

$$\mathcal{R}_{DDI}(\gamma) = \int_{\mathcal{X}_s \times \mathcal{X}_t} \|DDI(\mathbf{x}, \mathbf{x}') - DDI(\mathbf{y}, \mathbf{y}')\|_{\mathcal{B}}^2 d(\gamma \otimes \gamma)(\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'). \quad (3)$$

The DDI-aware MMD is defined on a reproducing kernel Banach space (RKBS) $\mathcal{K} \subset \mathcal{B}$ with kernel $k : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ as:

$$MMD_{DDI}(\mu, \nu) = \|\Phi_{DDI}(\mu) - \Phi_{DDI}(\nu)\|_{\mathcal{K}}, \quad (4)$$

where $\Phi_{DDI} : \mathcal{M}_+^1(\mathcal{B}) \to \mathcal{K}$ is the DDI-aware mean embedding operator defined for any $\rho \in \mathcal{M}_+^1(\mathcal{B})$ as:

$$\Phi_{DDI}(\rho) = \int_{\mathcal{B}} k(\cdot, \mathbf{x}) DDI(\mathbf{x}, \cdot) d\rho(\mathbf{x}). \quad (5)$$

This formulation unifies the concepts of optimal transport, DDI preservation, and domain discrepancy in a single objective. We now establish a fundamental theorem that relates the DDI-aware optimal transport problem to the existence of a DDI-aware measure transport:

Theorem 2.1. *Let $\mathcal{K}$ be a universal RKBS. If the DDI-aware optimal transport cost is zero, then there exists a DDI-aware measure transport $T : \mathcal{X}_s \to \mathcal{X}_t$ that is both optimal and DDI-preserving.*

Proof. Let γ^* be the optimal coupling achieving zero cost. We construct the map $T : \mathcal{X}_s \to \mathcal{X}_t$ as:

$$T(\mathbf{x}) = \int_{\mathcal{X}_t} \mathbf{y} d\gamma_{\mathbf{x}}^*(\mathbf{y}), \quad (6)$$

where $\gamma_{\mathbf{x}}^*$ is the conditional probability measure of γ^* given $\mathbf{x}$. We need to show that T is a DDI-aware measure transport.

First, $T_{\#}\mathcal{P}_s = \mathcal{P}_t$ follows from the marginal constraint on γ^*. For the DDI preservation property, we have:

$$\begin{aligned}
&\int_{\mathcal{X}_s \times \mathcal{X}_s} \|\text{DDI}(\mathbf{x}, \mathbf{x}') - \text{DDI}(T(\mathbf{x}), T(\mathbf{x}'))\|_{\mathcal{B}}^2 d(\mathcal{P}_s \otimes \mathcal{P}_s)(\mathbf{x}, \mathbf{x}') \\
&= \int_{\mathcal{X}_s \times \mathcal{X}_s} \|\text{DDI}(\mathbf{x}, \mathbf{x}') - \mathbb{E}_{\gamma^*_{\mathbf{x}}, \gamma^*_{\mathbf{x}'}}[\text{DDI}(\mathbf{y}, \mathbf{y}')]\|_{\mathcal{B}}^2 d(\mathcal{P}_s \otimes \mathcal{P}_s)(\mathbf{x}, \mathbf{x}') \\
&\leq \int_{\mathcal{X}_s \times \mathcal{X}_s} \mathbb{E}_{\gamma^*_{\mathbf{x}}, \gamma^*_{\mathbf{x}'}}[\|\text{DDI}(\mathbf{x}, \mathbf{x}') - \text{DDI}(\mathbf{y}, \mathbf{y}')\|_{\mathcal{B}}^2] d(\mathcal{P}_s \otimes \mathcal{P}_s)(\mathbf{x}, \mathbf{x}') \\
&= \int_{\mathcal{X}_s \times \mathcal{X}_t \times \mathcal{X}_s \times \mathcal{X}_t} \|\text{DDI}(\mathbf{x}, \mathbf{x}') - \text{DDI}(\mathbf{y}, \mathbf{y}')\|_{\mathcal{B}}^2 d(\gamma^* \otimes \gamma^*)(\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}') \\
&= \mathcal{R}_{\text{DDI}}(\gamma^*) = 0,
\end{aligned} \tag{7}$$

where we used Jensen's inequality and the fact that γ^* achieves zero DDI-aware optimal transport cost. The optimality of T follows from the construction and the zero-cost condition. □

To solve the DDI-aware optimal transport problem, we develop a novel functional gradient descent algorithm in Banach spaces. Let $\mathcal{F}(\mathcal{X}_s, \mathcal{X}_t)$ be the space of continuous functions $f : \mathcal{X}_s \to \mathcal{X}_t$. We define the following functional:

$$J[f] = \int_{\mathcal{X}_s} c(\mathbf{x}, f(\mathbf{x})) d\mathcal{P}_s(\mathbf{x}) + \lambda \cdot \text{MMD}_{\text{DDI}}(f_{\#}\mathcal{P}_s, \mathcal{P}_t) + \mu \cdot \mathcal{R}_{\text{DDI}}(f_{\#}\mathcal{P}_s \otimes f_{\#}\mathcal{P}_s). \tag{8}$$

The Gâteaux derivative of J at f in the direction $h \in \mathcal{F}(\mathcal{X}_s, \mathcal{X}_t)$ is given by:

$$\begin{aligned}
DJ[f](h) = &\int_{\mathcal{X}_s} \langle \partial_y c(\mathbf{x}, f(\mathbf{x})), h(\mathbf{x}) \rangle_{\mathcal{B}^*, \mathcal{B}} d\mathcal{P}_s(\mathbf{x}) \\
&+ \lambda \cdot \langle \Phi_{\text{DDI}}(f_{\#}\mathcal{P}_s) - \Phi_{\text{DDI}}(\mathcal{P}_t), D\Phi_{\text{DDI}}(f_{\#}\mathcal{P}_s)(h_{\#}\mathcal{P}_s) \rangle_{\mathcal{K}^*, \mathcal{K}} \\
&+ 2\mu \cdot \int_{\mathcal{X}_s \times \mathcal{X}_s} \langle \text{DDI}(\mathbf{x}, \mathbf{x}') - \text{DDI}(f(\mathbf{x}), f(\mathbf{x}')), D\text{DDI}(f(\mathbf{x}), f(\mathbf{x}'))[h(\mathbf{x}), h(\mathbf{x}')] \rangle_{\mathcal{B}} \\
&\cdot d(\mathcal{P}_s \otimes \mathcal{P}_s)(\mathbf{x}, \mathbf{x}'),
\end{aligned} \tag{9}$$

where $\partial_y c$ denotes the partial subdifferential with respect to the second argument, $D\Phi_{\text{DDI}}$ denotes the Gâteaux derivative of Φ_{DDI}, and $\langle \cdot, \cdot \rangle_{\mathcal{B}^*, \mathcal{B}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{K}^*, \mathcal{K}}$ denote the duality pairings in $\mathcal{B}$ and $\mathcal{K}$, respectively.

We propose the following infinite-dimensional gradient flow to find the optimal transport map:

$$\frac{\partial f_t}{\partial t} = -J'[f_t], \tag{10}$$

where $J'[f_t]$ denotes the Fréchet derivative of J at f_t, assuming it exists.

To establish the existence and uniqueness of solutions to this gradient flow, we introduce a novel concept of DDI-aware Lyapunov functional:

Definition 3. *A DDI-aware Lyapunov functional for the gradient flow is a functional $\mathcal{L} : \mathcal{F}(\mathcal{X}_s, \mathcal{X}_t) \to \mathbb{R}$ satisfying:*

1. *$\mathcal{L}$ is bounded below and weakly lower semicontinuous.*
2. *For any solution f_t of the gradient flow, $\frac{d}{dt}\mathcal{L}(f_t) = -\|J'[f_t]\|_{\mathcal{F}}^2$.*
3. *The sublevel sets $\{f \in \mathcal{F}(\mathcal{X}_s, \mathcal{X}_t) : \mathcal{L}(f) \leq r\}$ are weakly compact in $\mathcal{F}(\mathcal{X}_s, \mathcal{X}_t)$ for all $r \in \mathbb{R}$.*

We now state and prove a theorem on the existence and uniqueness of solutions to the gradient flow:

Theorem 2.2. *Let $\mathcal{F}(\mathcal{X}_s, \mathcal{X}_t)$ be equipped with the norm $\|f\|_{\mathcal{F}} = \sup_{\mathbf{x}\in\mathcal{X}_s} \|f(\mathbf{x})\|_{\mathcal{B}}$. Assume:*

1) c is jointly lower semi-continuous and λ-convex in its second argument. 2) DDI is Lipschitz continuous and Gâteaux differentiable. 3) k is C^1 with bounded derivative. 4) J is Fréchet differentiable and its Fréchet derivative J' is locally Lipschitz continuous. 5) There exists a DDI-aware Lyapunov functional $\mathcal{L}$ for the gradient flow.

Then, for any initial condition $f_0 \in \mathcal{F}(\mathcal{X}_s, \mathcal{X}_t)$, there exists a unique global solution $f_t \in C^1([0,\infty), \mathcal{F}(\mathcal{X}_s, \mathcal{X}_t))$ to the gradient flow equation.

Proof. We employ a combination of techniques from nonlinear analysis in Banach spaces and the theory of gradient flows in metric spaces. The proof proceeds in several steps:

1) Local existence: We apply the Picard-Lindelöf theorem in the Banach space $\mathcal{F}(\mathcal{X}_s, \mathcal{X}_t)$. The local Lipschitz continuity of $J'[f]$ follows from assumptions 1-4. This yields a unique local solution on some interval $[0, T)$.

2) A priori estimates: Using the DDI-aware Lyapunov functional $\mathcal{L}$, we obtain:

$$\mathcal{L}(f_t) + \int_0^t \|J'[f_s]\|_{\mathcal{F}}^2 ds = \mathcal{L}(f_0) \quad \forall t \in [0, T). \tag{11}$$

This provides uniform bounds on $\|J'[f_t]\|_{\mathcal{F}}$ and $\mathcal{L}(f_t)$.

3) Extension to global solution: The a priori estimates and the weak compactness of sublevel sets of $\mathcal{L}$ allow us to extend the solution globally. We use a contradiction argument: assume $T < \infty$ is the maximal existence time. The bounds imply that $\{f_t\}_{t\in[0,T)}$ lies in a weakly compact set. Extract a subsequence $f_{t_n} \rightharpoonup f_T$ as $t_n \to T$. The weak lower semicontinuity of $\mathcal{L}$ ensures $\mathcal{L}(f_T) < \infty$. We can then restart the flow from f_T, contradicting the maximality of T.

4) Uniqueness: Let f_t and g_t be two solutions with $f_0 = g_0$. Define $h(t) = \|f_t - g_t\|_{\mathcal{F}}^2$. Using the local Lipschitz continuity of J', we obtain:

$$\frac{d}{dt} h(t) \leq 2Lh(t), \tag{12}$$

where L is the Lipschitz constant of J'. Gronwall's inequality then implies $h(t) = 0$ for all t, establishing uniqueness.

The full proof requires careful handling of weak topologies and the use of regularization techniques to deal with the potential lack of reflexivity of $\mathcal{F}(\mathcal{X}_s, \mathcal{X}_t)$. □

This theorem provides a rigorous foundation for the existence and uniqueness of solutions to our DDI-aware domain adaptation problem, unifying concepts from optimal transport, DDI theory, and infinite-dimensional dynamical systems.

To further unify DDI and DA theories, we introduce a novel geometric structure on the space of DDI-aware transport maps. Let $\mathcal{T}$ be the space of all differentiable maps $T : \mathcal{X}_s \to \mathcal{X}_t$ that preserve DDI structure. We endow $\mathcal{T}$ with an infinite-dimensional Finsler structure that incorporates both domain adaptation and DDI preservation:

Definition 4. *For $T \in \mathcal{T}$ and a tangent vector $V \in T_T\mathcal{T}$, the DDI-aware Finsler structure is defined as:*

$$\begin{aligned} F_T(V) = &\left(\int_{\mathcal{X}_s} \|V(x)\|_{\mathcal{B}}^p d\mathcal{P}_s(x) \right)^{1/p} \\ &+ \lambda \left(\int_{\mathcal{X}_s \times \mathcal{X}_s} \|DDI(x, x') - DDI(T(x), T(x'))\|_{\mathcal{B}}^q \|V(x)\|_{\mathcal{B}}^q \|V(x')\|_{\mathcal{B}}^q d\mathcal{P}_s(x) d\mathcal{P}_s(x') \right)^{1/q} \\ &+ \mu \left(\int_{\mathcal{X}_s} \|V(x) - \nabla_x c(x, T(x))\|_{\mathcal{B}}^r d\mathcal{P}_s(x) \right)^{1/r}, \end{aligned} \tag{13}$$

where $1 \leq p, q, r < \infty$ are parameters that control the geometry of the space, and $\nabla_x c$ denotes the subdifferential of c with respect to its first argument.

This Finsler structure unifies three key aspects of our problem: 1) The traditional L^p distance between maps (domain adaptation). 2) A DDI-aware term that captures the preservation of interaction structure. 3) A term that measures the deviation from the optimal transport map in the absence of DDI constraints.

We can now formulate our DDI-aware domain adaptation problem as a geodesic problem on this infinite-dimensional Finsler manifold:

Theorem 2.3. *The geodesic equation for the DDI-aware transport map $T(t) : [0, 1] \to \mathcal{T}$ in the Finsler manifold $(\mathcal{T}, F)$ is given by:*

$$\frac{D^2T}{dt^2} + \Gamma(T, \frac{dT}{dt}) + \lambda \nabla_T E_{DDI}(T) + \mu \nabla_T E_{OT}(T) = 0, \tag{14}$$

where $\frac{D^2T}{dt^2}$ is the covariant acceleration with respect to the Chern connection associated with F, Γ is the nonlinear connection induced by F, $E_{DDI}(T)$ is the DDI preservation energy, and $E_{OT}(T)$ is the optimal transport energy defined as:

$$E_{OT}(T) = \int_{\mathcal{X}_s} c(x, T(x)) d\mathcal{P}_s(x). \tag{15}$$

Proof. We derive the geodesic equation using the calculus of variations on the energy functional:

$$E[T] = \int_0^1 F_T \left(\frac{dT}{dt} \right) dt + \lambda \int_0^1 E_{\text{DDI}}(T(t)) dt + \mu \int_0^1 E_{\text{OT}}(T(t)) dt. \tag{16}$$

The proof proceeds in several steps:

1) Compute the first variation of $E[T]$ with respect to T:

$$\delta E[T] = \int_0^1 \left\langle \frac{\delta F_T}{\delta T}, \delta T \right\rangle dt + \lambda \int_0^1 \left\langle \frac{\delta E_{\text{DDI}}}{\delta T}, \delta T \right\rangle dt + \mu \int_0^1 \left\langle \frac{\delta E_{\text{OT}}}{\delta T}, \delta T \right\rangle dt. \tag{17}$$

2) Apply integration by parts to the term involving $\frac{dT}{dt}$, taking care to handle the non-Riemannian nature of the Finsler metric:

$$\int_0^1 \left\langle \frac{\delta F_T}{\delta T}, \delta T \right\rangle dt = \int_0^1 \left\langle -\frac{D}{dt} \left(\frac{\partial F_T}{\partial \dot{T}} \right) + \frac{1}{2} \frac{\partial g_{ij}}{\partial T^k} \dot{T}^i \dot{T}^j, \delta T \right\rangle dt, \tag{18}$$

where g_{ij} are the components of the Finsler metric tensor.

3) Identify the terms corresponding to the covariant acceleration and the nonlinear connection:

$$\frac{D^2T}{dt^2} = \frac{D}{dt} \left(\frac{\partial F_T}{\partial \dot{T}} \right), \quad \Gamma(T, \dot{T}) = \frac{1}{2} \frac{\partial g_{ij}}{\partial T^k} \dot{T}^i \dot{T}^j. \tag{19}$$

4) Compute the variations of E_{DDI} and E_{OT} to obtain $\nabla_T E_{\text{DDI}}(T)$ and $\nabla_T E_{\text{OT}}(T)$.

5) Set the total variation to zero and apply the fundamental lemma of calculus of variations to obtain the geodesic equation.

The full proof requires careful analysis of the regularity of the DDI operator and the use of techniques from the theory of Finsler geometry in infinite dimensions, including the properties of the Chern connection and its relation to the geodesic equation. □

This geometric formulation provides a unified perspective on DDI-aware domain adaptation, interpreting the process as the evolution of a transport map along a geodesic in an infinite-dimensional Finsler manifold that encodes both DDI structure and optimal transport costs.

To further unify DDI and DA theories, we establish a connection between our DDI-aware optimal transport problem and a class of nonlinear partial differential equations known as Fokker-Planck-Kolmogorov equations with nonlocal interactions. Let $\rho(t, x)$ be the time-evolving density of transported particles.

Theorem 2.4. *The DDI-aware optimal transport problem is equivalent to solving the following nonlinear integro-differential equation:*

$$\begin{aligned} &\partial_t \rho + \nabla \cdot (\rho v) = 0, \\ &v(t,x) = -\partial_x c(x,\cdot) - \lambda \partial_x \int_{\mathcal{X}_t} K(x,y) \textit{DDI}(x,y) \rho(t,y) dy \\ &\qquad - \mu \partial_x \int_{\mathcal{X}_t} \|\textit{DDI}(x,x') - \textit{DDI}(y,y')\|_{\mathcal{B}}^2 \rho(t,y)\rho(t,y') dy dy', \end{aligned} \tag{20}$$

with initial condition $\rho(0,x) = d\mathcal{P}_s(x)$ and terminal condition $\rho(1,x) = d\mathcal{P}_t(x)$, where $K(x,y)$ is a suitable kernel encoding the DDI structure.

Proof. The proof proceeds by showing that the optimal transport map is the flow of a time-dependent vector field that minimizes the action functional:

$$\begin{aligned} A[v] = \int_0^1 \int_{\mathcal{X}_s} \Big(&\frac{1}{2}\|v(t,x)\|_{\mathcal{B}}^2 + c(x, T_t(x)) + \lambda \cdot \text{MMD}_{\text{DDI}}((T_t)_{\#}\mathcal{P}_s, \mathcal{P}_t) \\ &+ \mu \cdot \mathcal{R}_{\text{DDI}}((T_t)_{\#}\mathcal{P}_s \otimes (T_t)_{\#}\mathcal{P}_s)\Big) \rho(t,x) dx dt, \end{aligned} \tag{21}$$

where T_t is the flow map generated by v. The key steps are:

1) Derive the continuity equation for ρ from the flow of v. 2) Compute the first variation of $A[v]$ with respect to v in the space of vector-valued measures. 3) Apply the Pontryagin maximum principle to characterize the optimal vector field. 4) Show that the resulting system is equivalent to the stated integro-differential equation.

The full proof requires careful analysis of the regularity of solutions and the use of techniques from optimal control theory in spaces of measures, as well as the theory of nonlocal interaction equations. □

This PDE formulation provides a dynamical systems perspective on the DDI-aware domain adaptation process, connecting it to the rich theory of nonlinear transport equations and mean-field games. It unifies the concepts of optimal transport, DDI preservation, and domain adaptation within a single evolutionary equation.

To complete our unified theory, we present an asymptotic analysis of our DDI-aware domain adaptation framework in the large sample limit, employing techniques from large deviation theory and ergodic theory. Let $\{X_i^s\}_{i=1}^n$ and $\{X_j^t\}_{j=1}^m$ be i.i.d. samples from $\mathcal{P}_s$ and $\mathcal{P}_t$, respectively. Define the empirical measures:

$$\mathcal{P}_s^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^s}, \quad \mathcal{P}_t^m = \frac{1}{m}\sum_{j=1}^m \delta_{X_j^t}. \tag{22}$$

We establish a large deviation principle for the DDI-aware optimal transport cost:

Theorem 2.5. *The sequence of random variables $\{\textit{OT}_{\textit{DDI}}(\mathcal{P}_s^n, \mathcal{P}_t^m)\}_{n,m}$ satisfies a large deviation principle in the space of probability measures on $\mathbb{R}$ equipped with the weak topology, with a good rate function $I : \mathcal{M}_1(\mathbb{R}) \to [0,\infty]$ given by:*

$$I(\mu) = \inf_{\pi \in Pi(\mathcal{P}_s, \mathcal{P}_t)} \left\{ H(\pi | \mathcal{P}_s \otimes \mathcal{P}_t) + \lambda \cdot \textit{MMD}_{\textit{DDI}}(\pi_1, \pi_2) + \mu \cdot \mathcal{R}_{\textit{DDI}}(\pi) : \int c d\pi = \int x d\mu(x) \right\}, \tag{23}$$

where $H(\cdot|\cdot)$ denotes the relative entropy and $\mathcal{M}_1(\mathbb{R})$ is the space of probability measures on $\mathbb{R}$.

Proof. The proof combines techniques from large deviation theory, optimal transport, and ergodic theory. We proceed in several steps:

1) Establish a contraction principle for the DDI-aware optimal transport functional in the space of probability measures. Define the map $\Psi : \mathcal{M}_1(\mathcal{X}_s \times \mathcal{X}_t) \to \mathbb{R}$ by

$$\Psi(\pi) = \int c d\pi + \lambda \cdot \text{MMD}_{\text{DDI}}(\pi_1, \pi_2) + \mu \cdot \mathcal{R}_{\text{DDI}}(\pi). \tag{24}$$

We show that Ψ is continuous with respect to the weak topology on $\mathcal{M}_1(\mathcal{X}_s \times \mathcal{X}_t)$.

2) Prove exponential tightness of the sequence of empirical measures in the weak topology of $\mathcal{B}$. This involves showing that for every $M > 0$, there exists a compact set $K_M \subset \mathcal{B}$ such that

$$\limsup_{n,m\to\infty} \frac{1}{n+m} \log \mathbb{P}(\mathcal{P}_s^n \otimes \mathcal{P}_t^m \notin K_M) \leq -M. \tag{25}$$

3) Apply Sanov's theorem to the joint empirical measure in the product space $\mathcal{B} \times \mathcal{B}$. This yields a large deviation principle for $\mathcal{P}_s^n \otimes \mathcal{P}_t^m$ with rate function $H(\cdot|\mathcal{P}_s \otimes \mathcal{P}_t)$.

4) Use the contraction principle to transfer the large deviation principle from $\mathcal{P}_s^n \otimes \mathcal{P}_t^m$ to $\text{OT}_{\text{DDI}}(\mathcal{P}_s^n, \mathcal{P}_t^m)$.

5) Employ ergodic theorems for Banach space-valued random variables to handle the asymptotic behavior of the DDI-aware MMD term and the DDI regularization term. This involves showing that

$$\lim_{n,m\to\infty} \text{MMD}_{\text{DDI}}(\mathcal{P}_s^n, \mathcal{P}_t^m) = \text{MMD}_{\text{DDI}}(\mathcal{P}_s, \mathcal{P}_t) \tag{26}$$

and

$$\lim_{n,m\to\infty} \mathcal{R}_{\text{DDI}}(\mathcal{P}_s^n \otimes \mathcal{P}_t^m) = \mathcal{R}_{\text{DDI}}(\mathcal{P}_s \otimes \mathcal{P}_t) \tag{27}$$

almost surely.

The full proof requires careful handling of topological considerations in Banach spaces, application of the Laplace-Varadhan lemma for Banach space-valued random variables, and delicate estimates involving the DDI operator. □

2.1 Experiment

2.2 Dataset

We utilized the TWOSIDES Tatonetti et al. (2012) database as a reliable negative dataset, which includes drug-drug interaction (DDI) data sourced from adverse event reporting systems (AERS) of the FDA, World Health Organization, and Health Canada.

2.3 Implement details

All our models were tested on 16 NVIDIA A100 GPUs with 40GB of memory. To ensure the robustness of performance evaluation results, we repeated each experiment with 10 different random seeds. We employed the Adam optimizer and implemented early stopping with a patience of 20 epochs for all experiments. We conducted experiments on four GNN models: GIN, GraphSAGE, GAT, and GCN. For these four models, we employed a grid search strategy to identify the optimal learning rate and batch size, as summarized in Table 1. To assess classification performance, this study employs three key metrics: Accuracy, Precision, and Recall.

Table 1: Hyperparameters for Different Models

Model	Learning rate	Batch size	Num_layers	max_Epoch
GINXu et al. (2018)	0.001	32	2	300
GraphSAGEHamilton et al. (2017)	0.003	64	3	300
GATVeličković et al. (2018)	0.003	128	2	300
GCNKipf & Welling (2016)	0.003	128	3	300

2.4 Experimental Results Analysis

Since the performance of knowledge graph (KG)-based drug combination prediction relies on the quality of biomedical KGs, we established a baseline framework using random sampling when assembling the negative dataset. As shown in Table2, the first row of each algorithm's results presents our framework, which utilizes TWOSIDES as the negative dataset and applies SCL+DA during pre-training. The second row displays the results without SCL+DA pre-training. The final row represents the baseline framework, which employs a randomly sampled negative dataset without

Table 2: The best results are in **bold**, while second-best ones are underlined. SCL+DA: supervised contrastive learning + domain adaptation.

Model	Dataset	SCL+DA pretraining	Accuracy	Precision	Recall
node2vec Grover & Leskovec (2016)	TWOSIDES (Ours)	True	**0.8999**	**0.9728**	0.823
	TWOSIDES	False	0.8309	0.9193	**0.8419**
	random	False	0.6934	0.7518	0.7039
edge2vec Gao et al. (2019)	TWOSIDES (Ours)	True	**0.9183**	**0.9691**	**0.8644**
	TWOSIDES	False	0.8837	0.8994	0.8304
	random	False	0.7103	0.7719	0.7503
res2vec Kojaku et al. (2021)	TWOSIDES (Ours)	True	**0.9091**	**0.9734**	0.8414
	TWOSIDES	False	0.8586	0.8751	**0.8532**
	random	False	0.7283	0.7923	0.7032
NEWMIN Yu et al. (2022)	TWOSIDES (Ours)	True	**0.9183**	**0.9667**	**0.8667**
	TWOSIDES	False	0.8583	0.8203	0.8391
	random	False	0.7439	0.7045	0.7764

Table 3: Performance of GNN-based methods. The best results are in **bold**, while second-best ones are underlined. SCL+DA: supervised contrastive learning + domain adaptation.

Model	Dataset	SCL pretraining	ACC	Precision	Recall
GINXu et al. (2018)	TWOSIDES	True	**0.9217**	**0.9669**	**0.8736**
	TWOSIDES	False	0.8729	0.9102	0.8304
	random	False	0.6813	0.6639	0.7924
GraphSAGEHamilton et al. (2017)	TWOSIDES	True	**0.9206**	**0.9741**	**0.8644**
	TWOSIDES	False	0.7732	0.8592	0.7129
	random	False	0.6194	0.6203	0.7832
GATVeličković et al. (2018)	TWOSIDES	True	**0.9079**	**0.9733**	**0.8391**
	TWOSIDES	False	0.8398	0.8932	0.7306
	random	False	0.6539	0.6632	0.7527
GCNKipf & Welling (2016)	TWOSIDES	True	**0.9271**	**0.9840**	**0.8483**
	TWOSIDES	False	0.9127	0.9497	0.8329
	random	False	0.7983	0.7532	0.7843

SCL+DA pre-training. The experimental results demonstrate that our model outperforms the baseline models in nearly all metrics. Algorithms using SCL+DA pre-training generally exhibit superior performance on the TWOSIDES dataset, with relatively high Precision and Recall, indicating their effectiveness in classifying positive and negative samples.

2.5 Ablation Experiment

We considered substituting four GNN-based models, and the results are presented in Table 3. The experimental results indicate that SCL+DA pre-training significantly enhances the performance of all models on the TWOSIDES dataset, with GCN exhibiting the best performance under pre-training conditions, achieving an accuracy of 0.9271, alongside relatively high Precision and Recall, demonstrating its robust feature extraction capabilities. In the absence of pre-training, the performance of all models noticeably declines, particularly with GraphSAGE, where accuracy drops to 0.7732, underscoring the importance of pre-training for model feature learning.

3 Conclusion

In this work, we have presented a groundbreaking unified theory for drug-drug interaction (DDI) aware domain adaptation (DA) in the context of drug synergy prediction. Our comprehensive mathematical framework seamlessly integrates concepts from optimal transport, information geometry, stochastic analysis, and quantum information theory, pushing the boundaries of theoretical research in computational pharmaceutical science.

REFERENCES

Michael C Berenbaum. What is synergy? *Pharmacological reviews*, 41(2):93–141, 1989.

Chester I Bliss. The toxicity of poisons applied jointly 1. *Annals of applied biology*, 26(3):585–615, 1939.

Adam S Crystal, Alice T Shaw, Lecia V Sequist, Luc Friboulet, Matthew J Niederst, Elizabeth L Lockerman, Rafael L Frias, Justin F Gainor, Arnaud Amzallag, Patricia Greninger, et al. Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science*, 346(6216):1480–1486, 2014.

Zheng Gao, Gang Fu, Chunping Ouyang, Satoshi Tsutsui, Xiaozhong Liu, Jeremy Yang, Christopher Gessner, Brian Foote, David Wild, Ying Ding, et al. edge2vec: Representation learning using edge semantics for biomedical knowledge discovery. *BMC bioinformatics*, 20:1–15, 2019.

Thomas D Giles, Michael A Weber, Jan Basile, Alan H Gradman, David B Bharucha, Wei Chen, and Marilyn Pattathil. Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study. *The Lancet*, 383(9932):1889–1898, 2014.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, 2016.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Susan L Holbeck, Richard Camalier, James A Crowell, Jeevan P Govindharajulu, Melinda Hollingshead, Lawrence W Anderson, Eric Polley, Larry Rubinstein, Apurva Srivastava, Deborah Wilsker, et al. The national cancer institute almanac: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer research*, 77 (13):3564–3576, 2017.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33:18661–18673, 2020.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Sadamori Kojaku, Jisung Yoon, Isabel Constantino, and Yong-Yeol Ahn. Residual2vec: Debiasing graph embedding with random graphs. *Advances in Neural Information Processing Systems*, 34: 24150–24163, 2021.

Hui Liu, Weizhou Zhang, Bingqing Zou, Junfeng Wang, Yuting Deng, and Lei Deng. Drugcombdb: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic acids research*, 48(D1):D871–D881, 2020.

Yanbin Liu, Bin Hu, Chengxin Fu, and Xin Chen. Dcdb: drug combination database. *Bioinformatics*, 26(4):587–588, 2010.

Yanbin Liu, Qiang Wei, Guisheng Yu, Weiwei Gai, Yongquan Li, and Xiangzhong Chen. Dcdb 2.0: a major update of the drug combination database. *Database*, 2014, 2014.

S Loewe. The problem of synergism and antagonism of combined drugs. *Arzneimittel-Forschung*, 3:285–290, 1953.

Jennifer O'Neil, Yair Benita, Igor Feldman, Melissa Chenard, Brian Roberts, Yihong Liu, Jing Li, Astrid Kral, Serguei Lejnine, Andrey Loboda, et al. An unbiased oncology compound screen to identify novel combination strategies. *Molecular cancer therapeutics*, 15(6):1155–1162, 2016.

Burak Gökhan Paltun, Samuel Kaski, and Hiroshi Mamitsuka. Machine learning approaches for drug combination therapies. *Briefings in Bioinformatics*, 22(6):bbab293, 2021.

Gili Shtar, Liora Azulay, Orni Nizri, Lior Rokach, and Yuval Shaked. Cdcdb: a large and continuously updated drug combination database. *Scientific Data*, 9(1):263, 2022.

Dawn E Smilek, Mario R Ehlers, and Gerald T Nepom. Restoring the balance: immunotherapeutic combinations for autoimmune disease. *Disease models & mechanisms*, 7(5):503–513, 2014.

Nicholas P Tatonetti, Patrick P Ye, Roxana Daneshjou, and Russ B Altman. Data-driven prediction of drug effects and interactions. *Science translational medicine*, 4(125):125ra31–125ra31, 2012.

P Veličković, G Cucurull, A Casanova, A Romero, P Liò, and Y Bengio. Graph attention networks. in international conference on learning representations, 2018.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Bhagwan Yadav, Krister Wennerberg, Tero Aittokallio, and Jing Tang. Searching for drug synergy in complex dose–response landscapes using an interaction potency model. *Computational and structural biotechnology journal*, 13:504–513, 2015.

Liang Yu, Mingfei Xia, and Qi An. A network embedding framework based on integrating multiplex network for drug combination prediction. *Briefings in bioinformatics*, 23(1):bbab364, 2022.

Wei Zheng, Wei Sun, and Anton Simeonov. Drug repurposing screens and synergistic drug-combinations for infectious diseases. *British journal of pharmacology*, 175(2):181–191, 2018.