How Far Apart Is Distinguishable? Measuring the Expressivity of Topological Neural Networks through Weisfeiler–Leman and Co-Optimal Transport

Anonymous authors
Paper under double-blind review

Abstract

The Weisfeiler-Leman (WL) isomorphism test is a classical tool for analyzing the expressive power of message-passing graph neural networks (GNNs). Message-passing topological neural networks (TNNs) generalize message passing from graphs to higher-order combinatorial domains, such as simplicial and cell complexes. Recent work extends the WL test to attributed combinatorial complexes through the Combinatorial Complex Weisfeiler-Leman (CCWL) test, showing that CCWL upper-bounds the expressivity of a broad class of TNNs. Like the classical WL test, however, CCWL is binary, i.e., it decides whether two complexes can be distinguished, but not how different they are. For this reason, we turn CCWL into a quantitative distance by replacing its color refinement with a recursive distributional representation: for each cell, the messages passed through its boundary, coboundary, lower-adjacency, and upper-adjacency channels are represented as a distribution, and these channel-wise representations are recursively compared across refinement depths. We prove that the resulting distance is a pseudometric and that it vanishes exactly for pairs of attributed combinatorial complexes that CCWL cannot distinguish. We then connect it to co-optimal transport (Co-OT) between attributed combinatorial complexes through a convex relaxation of Co-OT. We show that this convex Co-OT discrepancy and the CCWL distance characterize the same zero set and correlate strongly. To the best of our knowledge, this is the first work to jointly connect TNNs, optimal transport, and CCWL expressivity, opening several future research directions.

1 Introduction

The structure of data plays a fundamental role in machine learning. Graph neural networks (GNNs) have been a dominant framework for learning from data with intrinsic relational structure, with applications spanning molecular science (65; 64), computational biology (77; 3) and social networks (43). Graphs provide a canonical representation of such relational data by encoding pairwise relationships as edges connecting vertices. On graphs, the most used learning models are message-passing GNNs, including graph convolutional networks(39). However, many real-world systems exhibit interactions that extend beyond pairwise relations. For example, atoms form and interact through higher-order chemical structures such as functional groups and rings, or individuals form and interact through groups and communities. Topological deep learning (TDL) (34; 48; 33; 50; 12) emerges as a framework for incorporating such higher-order structure into deep learning, allowing models to propagate and aggregate information not only between individual entities but also across groups and structures of different dimensions in a hierarchical way. Common topological domains include simplicial complexes (70; 16; 9; 6; 41), cellular complexes (14; 28), and hypergraphs. Combinatorial complexes (CCs) unify the combinatorial descriptions of all of these spaces by allowing flexible hierarchical relations among cells of different ranks (7). On CCs, the most used learning models are message-passing topological neural networks (TNNs) (49; 61).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The class of tasks a GNN can represent and solve has been shown to be closely tied to its isomorphism-distinguishing power, which has made characterizing model expressivity a central topic of study. As such, a prominent line of work builds on connecting GNNs to the Weisfeiler-Leman (WL) isomorphism test (66), a classical tool from graph theory for distinguishing non-isomorphic graphs. In particular, Xu et al. (67) showed that the expressive power of a broad class of first-order message-passing GNNs is upper-bounded by that of the WL test. Subsequent work has extended this perspective through WL variants, including subgraph WL (73) and approaches that incorporate spectral information (75). Beyond characterizing the limitations of existing architectures, these insights have also guided the development of higher-order and generalized GNNs designed to overcome the expressivity bottlenecks of standard message passing (47; 46). In parallel, another line of work has extended the WL framework to higher-order combinatorial domains, including simplicial complexes (16), cell complexes (14), and hypergraphs (76), leading to TNN architectures with distinguishing power beyond that of standard WL. Recent work has further extended the WL test to CCs (1; 19), providing the foundation for our work.

However, these works rely on a coarse notion of distinguishability. A finer analysis of expressivity requires characterizations beyond merely separating points (54; 74). To quantify differences between distinguishable structures, optimal transport (OT) (63) has been identified as a useful framework (20; 21; 54). Originally formulated to compare probability measures, it has been extended to graphs (53; 62), trees (23), and hypergraphs (22). (54) characterize the fine-grained expressivity of GNNs using a tree-based OT distance. Closer to our work, (20) relates WL refinement to Markov chains and recursively applies OT to define a WL distance with the same distinguishing power as the WL test.

Building on recent developments in combinatorial complex WL (CCWL) (1; 19), we define a CCWL distance with the same distinguishing power as the CCWL test, while providing a fine-grained geometry for comparing combinatorial complexes. We further connect this geometry to an optimal transport formulation of similarity between complexes. We summarize our contributions as follows:

- *We define a CCWL distance based on CCWL refinement and prove that it is nonzero if and only if the CCWL test distinguishes the two attributed combinatorial complexes.*
- *We introduce a co-optimal transport (Co-OT) pseudometric and derive a convex relaxation. We prove that convex Co-OT has the same distinguishing power as CCWL and therefore the same zero set as the CCWL distance.*
- *Empirically, we demonstrate a strong correlation between the CCWL distance and convex Co-OT distances.*

2 Preliminaries

In this section, we review optimal transport distances between probability distributions and the WL test for attributed graphs and combinatorial complexes.

2.1 Optimal Transport and Wasserstein distances

Within the framework of optimal transport (OT), Wasserstein distances form a family of metrics for comparing probability distributions (63). Let (Ω, d) be a Polish metric space. For $p \geq 1$, let $\mathcal{P}_p(\Omega)$ denote the set of Borel probability measures on Ω with finite p-th moment, that is, $\int_\Omega d(x, x_0)^p \, d\mu(x) < \infty$ for some (equivalently, any) fixed $x_0 \in \Omega$. The p-Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(\Omega)$ is defined as the infimum of the transport problem:

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} d(x, y)^p \, d\pi(x, y) \right)^{1/p},$$

where $\Pi(\mu, \nu)$ denotes the set of couplings/transport plans of μ and ν, namely, Borel probability measures on $\Omega \times \Omega$ with marginals μ and ν. Specifically, each $\pi \in \Pi(\mu, \nu)$ satisfies

$$\pi(I \times \Omega) = \mu(I), \qquad \pi(\Omega \times J) = \nu(J)$$

for all Borel sets $I, J \subseteq \Omega$. The distance W_p defines a metric on $\mathcal{P}_p(\Omega)$.

Co-Optimal Transport. Co-optimal transport (Co-OT) (55) extends OT by jointly matching samples and features between two datasets. Represent the datasets by $A^1 \in \mathbb{R}^{m_1 \times n_1}$ and $A^2 \in \mathbb{R}^{m_2 \times n_2}$, where each entry records a feature (column) value for a sample (row). Given sample probability weights μ_1, μ_2 and feature probability weights ν_1, ν_2, the Co-OT problem is

$$\min_{\substack{\pi \in \Pi(\mu_1, \mu_2) \\ \xi \in \Pi(\nu_1, \nu_2)}} \sum_{i,j,k,\ell} c(A^1_{ij}, A^2_{k\ell})\, \pi_{ik} \xi_{j\ell}.$$

where $c(\cdot, \cdot)$ denotes the ground cost function, π matches samples and ξ matches features. The objective is bilinear in the transport plans and generally nonconvex.

2.2 Weisfeiler-Lehman isomorphism test and the distance

Let $G = (V, E)$ be a finite graph with the set of nodes V and the set of edges E. The graph may have initial node attributes $l^{(0)} : V \to Z_0$, where $l^{(0)}$ assigns each node a label in the space of Z_0. For graphs without node attributes, all nodes receive the same initial label. The WL test is a classical procedure for distinguishing non-isomorphic graphs. It iteratively refines node labels according to

$$\ell^{(t+1)}(v) = \left(\ell^{(t)}(v), \left\{\!\left\{ \ell^{(t)}(u) : u \in N(v) \right\}\!\right\}\right),$$

where $N(v)$ denotes the neighborhood of $v \in V$, double braces denote a multiset. $\ell^{(t+1)}(v)$ is often mapped to a label space, e.g. $\mathbb{N}$, by an injective map shared across the graphs being compared.

At each iteration, the test compares the multisets of node attributes in the two graphs. If they differ, the graphs are non-isomorphic, where isomorphisms between attributed graphs must preserve the initial attributes. For finite graphs, refinement on their disjoint union is guaranteed to stabilize after finitely many iterations, meaning that no attribute class splits further. If the multisets still agree at stabilization, the test is inconclusive: the graphs may or may not be isomorphic. Thus, the WL test provides a binary notion of distinguishability, but does not quantify how far apart two graphs are.

(20) obtain a quantitative analogue by viewing neighborhood aggregation as a Markov transition. For any graph $G = (V, E)$ and parameter $q \in [0, 1)$, the q-Markov chain $m^{G,q}_\bullet$ induced by G is defined as follows: for any $v \in V$,

$$m^{G,q}_v := \begin{cases} q\,\delta_v + \dfrac{1-q}{\deg_G(v)} \displaystyle\sum_{v' \in N_G(v)} \delta_{v'}, & N_G(v) \neq \emptyset, \\ \delta_v, & N_G(v) = \emptyset. \end{cases} \tag{1}$$

Then the WL distance associates the label refinement with a hierarchy of probability measures.

Definition 2.1 (WL distance)**.** *Given a finite set V, let $\mathcal{X} = (V, m^V, \mu^V)$ be a Markov chain with kernel $m^V : V \to \mathcal{P}(V)$, and a fully-supported stationary distribution μ^V with respect to m^V. Suppose it has initial attributes $\ell_V : V \to Z$, where (Z, d_Z) is a metric space. Starting from $\ell^{(0)}_V = \ell_V$, define $\ell^{(1)}_V(v) := (\ell^{(0)}_V)_\# m^V_v : V \to \mathcal{P}(Z)$, and*

$$\ell^{(t+1)}_V(v) := (\ell^{(t)}_V)_\# m^V_v : V \to \mathcal{P}^{\circ(t+1)}(Z), \qquad \mathfrak{L}_t(\mathcal{X}, \ell_V) := (\ell^{(t)}_V)_\# \mu^V \in \mathcal{P}^{\circ(t+1)}(Z),$$

where $\#$ denotes the pushforward of a measure, and $\mathcal{P}^{\circ t}(Z)$ is a hierarchy of measures with $\mathcal{P}^{\circ 1}(Z) = \mathcal{P}(Z)$ and $\mathcal{P}^{\circ(t+1)}(Z) = \mathcal{P}(\mathcal{P}^{\circ t}(Z))$. For two such Markov chains $\mathcal{X}, \mathcal{Y}$ with initial attributes $\ell_\mathcal{X}, \ell_\mathcal{Y}$, the depth-$t$ WL distance and the WL distance are defined by

$$d^{(t)}_{\mathrm{WL}}\big((\mathcal{X}, \ell_\mathcal{X}), (\mathcal{Y}, \ell_\mathcal{Y})\big) := W_1(\mathfrak{L}_t(\mathcal{X}, \ell_\mathcal{X}), \mathfrak{L}_t(\mathcal{Y}, \ell_\mathcal{Y})), \qquad d_{\mathrm{WL}} := \sup_{t \geq 0} d^{(t)}_{\mathrm{WL}},$$

where the 1-Wasserstein distance is applied recursively, starting from the ground metric d_Z.

By applying this definition to a graph G-induced Markov chain with kernel $m^V = m^{G,q}_\bullet$ and and node attributes ℓ_G, it yields the WL distance between attributed graphs. With $q \in (\frac{1}{2}, 1)$ to preserve self-identification and a suitable relabeling map to retain multiplicity information lost when normalizing multisets into probability distributions, (20) show that the WL test distinguishes two graphs G_1 and G_2 if and only if $d_{\mathrm{WL}}\big((G_1, \ell_{G_1}), (G_2, \ell_{G_2})\big) > 0$, where ℓ_{G_1} and ℓ_{G_2} denote the resulting relabeled node attributes.

2.3 Combinatorial Complex Weisfeiler–Lehman test

In this work we are interested in combinatorial complexes, which extends a graph by allowing cells containing more than two vertices and assigning each cell a rank to encode its combinatorial structure.

Definition 2.2 (Combinatorial complex (CC), (34))**.** *A combinatorial complex (CC) is a triple* (S, X, rk) *where* $S \subset \mathbb{R}^d$ *is a finite set of vertices,* $X \subseteq \mathcal{P}(S) \setminus \{\emptyset\}$ *is a set of cells containing all singletons* $\{v\}$ *in* S*, and* $\mathrm{rk} : X \to \mathbb{Z}_{\geq 0}$ *is order-preserving (*$x \subseteq y \Rightarrow \mathrm{rk}(x) \leq \mathrm{rk}(y)$*) with* $\mathrm{rk}(\{v\}) = 0$*. Write* $X_r := \mathrm{rk}^{-1}(r)$ *and let* R *be the maximal rank (i.e. the "height"). We identify* $X_0 \cong S$*.*

An attributed combinatorial complex (ACC) additionally carries a map $c : X \to Z$ assigning each cell an attribute in a space Z; we denote it by $\mathcal{X} = (X, \mathrm{rk}, c)$, leaving the underlying vertex set S implicit. Following (14; 1), we define the four types of neighborhoods for CCs:

- Boundary: $\mathcal{N}_B(x) = \{y \subset x : \mathrm{rk}(y) = \mathrm{rk}(x) - 1\}$.
- Co-boundary: $\mathcal{N}_C(x) = \{y \supset x : \mathrm{rk}(y) = \mathrm{rk}(x) + 1\}$.
- Upper adjacency: $\mathcal{N}_\uparrow(x) = \{y : \exists \delta \text{ such that } x \subset \delta,\ y \subset \delta \text{ and } \mathrm{rk}(\delta) - 1 = \mathrm{rk}(y) = \mathrm{rk}(x)\}$.
- Lower adjacency: $\mathcal{N}_\downarrow(x) = \{y : \exists \delta \text{ such that } \delta \subset x,\ \delta \subset y \text{ and } \mathrm{rk}(\delta) + 1 = \mathrm{rk}(y) = \mathrm{rk}(x)\}$.

(1) defines the WL test on ACCs (CCWL) that iteratively refines cell attributes by aggregating information separately over the four neighborhood types defined above:

Definition 2.3 (CCWL)**.** *Let* $\mathcal{X} = (X, \mathrm{rk}, c)$ *be an attributed combinatorial complex. Initialize each cell* $x \in X$ *with* $\chi_0(x) := (c(x), \mathrm{rk}(x))$*. At each refinement step* $t \geq 0$*, form the multisets*

$$M_{B,t}(x) := \{\{\chi_t(y) : y \in \mathcal{N}_B(x)\}\}, \quad M_{C,t}(x) := \{\{\chi_t(y) : y \in \mathcal{N}_C(x)\}\},$$
$$M_{\downarrow,t}(x) := \{\{(\chi_t(y), \chi_t(z)) : y \in \mathcal{N}_\downarrow(x), z \in \mathcal{N}_B(x) \cap \mathcal{N}_B(y)\}\},$$
$$M_{\uparrow,t}(x) := \{\{(\chi_t(y), \chi_t(z)) : y \in \mathcal{N}_\uparrow(x), z \in \mathcal{N}_C(x) \cap \mathcal{N}_C(y)\}\}.$$

The cell attributes are then updated by

$$\chi_{t+1}(x) := (\chi_t(x), M_{B,t}(x), M_{C,t}(x), M_{\downarrow,t}(x), M_{\uparrow,t}(x)),$$

Two ACCs $\mathcal{X}$ *and* $\mathcal{Y}$ *are* CCWL-equivalent*, written* $\mathcal{X} \equiv_{\mathrm{CCWL}} \mathcal{Y}$*, if*

$$\{\{\chi_t^{\mathcal{X}}(x) : x \in X\}\} = \{\{\chi_t^{\mathcal{Y}}(y) : y \in Y\}\} \qquad \text{for every } t \geq 0.$$

For upper and lower adjacency, the multisets $M_{\uparrow,t}(x)$ and $M_{\downarrow,t}(x)$ contain pairs: one for the neighboring cell and one for the corresponding mediating cell.

Similar to the WL test, CCWL stabilizes after finitely many steps (1). The distinguishing power of CCWL upper-bounds that of a broad class of message-passing TNNs, including message-passing simplicial networks (16), CW Networks/CIN (14), CIN++ (29) and CCIN (19). We seek to extend this distinguishability to a quantitative distance that is zero exactly when two complexes are CCWL-equivalent.

3 CCWL distance

Inspired by the WL distance (20), we construct a hierarchy of probability measures and compare them using recursively defined Wasserstein distances. Unlike the graph-based construction, which aggregates through a single Markov kernel, our distance follows the structure of CCWL: it treats the four neighborhood channels separately and represents upper and lower adjacency using joint distributions of neighbor-mediator pairs. Throughout this section, we denote a finite ACC by $\mathcal{X} = (X, \mathrm{rk}, c)$, and let $X_r := \{x \in X : \mathrm{rk}(x) = r\}$.

3.1 Distributional Representations of CCWL

For each cell $x \in X$, we represent its four neighborhood channels by probability measures. For nonempty boundary and coboundary neighborhoods, define

$$\kappa_x^B = \frac{1}{|\mathcal{N}_B(x)|} \sum_{y \in \mathcal{N}_B(x)} \delta_y, \qquad \kappa_x^C = \frac{1}{|\mathcal{N}_C(x)|} \sum_{y \in \mathcal{N}_C(x)} \delta_y.$$

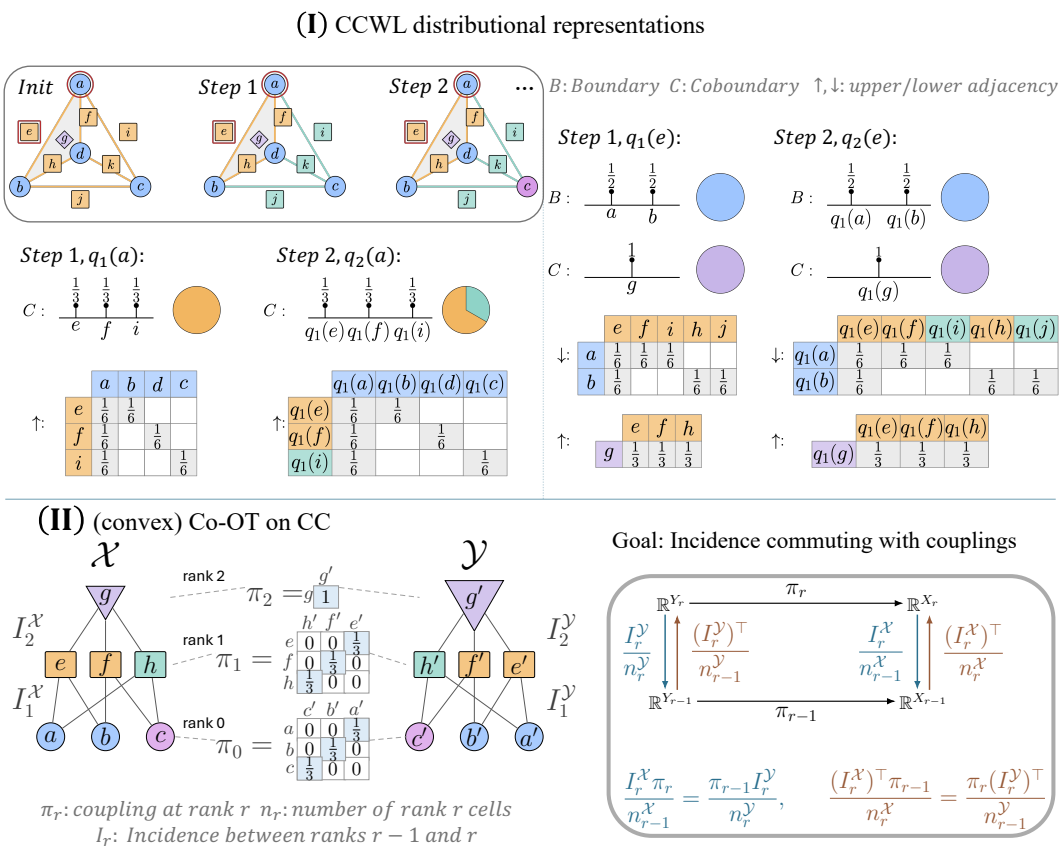


Figure 1: ***(I)*** *CCWL distance recursively represents cell attributes through neighborhood distributions, illustrated for vertex a and edge e. Boundary and coboundary channels aggregate individual attributes, while upper and lower adjacency channels retain joint distributions of neighbor-mediator attributes.* ***(II)*** *Co-OT jointly matches cells per rank via couplings π_r. Convex Co-OT penalizes the failure of these couplings to commute with normalized incidence maps in both directions. Note that the upper and lower adjacent neighbors of a cell include itself whenever the corresponding mediator exists.*

For the lower and upper adjacent neighborhoods, let $\mathcal{A}_\downarrow(x) = \{(y,z) \mid y \in \mathcal{N}_\downarrow(x), z \in \mathcal{N}_B(x) \cap \mathcal{N}_B(y)\}$ and $\mathcal{A}_\uparrow(x) = \{(y,z) \mid y \in \mathcal{N}_\uparrow(x), z \in \mathcal{N}_C(x) \cap \mathcal{N}_C(y)\}$. Then define the distributions

$$\kappa_x^\downarrow = \frac{1}{|\mathcal{A}_\downarrow(x)|} \sum_{(y,z)\in\mathcal{A}_\downarrow(x)} \delta_{(y,z)}, \qquad \kappa_x^\uparrow = \frac{1}{|\mathcal{A}_\uparrow(x)|} \sum_{(y,z)\in\mathcal{A}_\uparrow(x)} \delta_{(y,z)}.$$

When a neighborhood channel $a \in \{B, C, \downarrow, \uparrow\}$ is empty, we set $\kappa_x^a = \delta_{\perp_a}$, where $\perp_a$ is a distinguished null symbol for that channel. Thus, every channel carries a probability distribution. For simplicity, the null symbols and their treatment under pushforward maps are omitted. The detailed treatment of empty channels is given in C.1.

Refining distributions via pushforward. Let (Z_0, d_0) be a common initial attribute space equipped with a bounded metric, and let $q_0^{\mathcal{X}} : X \to Z_0$ be the initial attribute map. In addition to the four neighborhood channels, we introduce an identity channel id to preserve the cell's current label. Let $I := \{id, B, C, \downarrow, \uparrow\}$ and fix channel weights $\alpha_a > 0$, $a \in I$, with $\sum_{a\in I} \alpha_a = 1$, shared across all cells, complexes, and refinement depths. Starting from the initial attribute space Z_0, define recursively, for $t \geq 1$, the *channel-tagged label space*

$$\mathsf{T}_{t-1} := \bigsqcup_{a\in\{id,B,C\}} (\{a\} \times Z_{t-1}) \sqcup \bigsqcup_{a\in\{\downarrow,\uparrow\}} (\{a\} \times Z_{t-1}^2).$$

where $\bigsqcup$ denotes the disjoint union. This space combines all neighborhood information, each tagged by its channel $\{a\}$. Write $\iota_{a,t}(u) := (a, u)$ for the corresponding injection into T_{t-1}, and define

$$Z_t := \left\{ \sum_{a\in I} \alpha_a (\iota_{a,t})_\# \eta_a \;\middle|\; \begin{array}{ll} \eta_a \in \mathcal{P}(Z_{t-1}), & a \in \{id, B, C\}, \\ \eta_a \in \mathcal{P}(Z_{t-1}^2), & a \in \{\downarrow, \uparrow\} \end{array} \right\} \subseteq \mathcal{P}(\mathsf{T}_{t-1}).$$

Each element of Z_t is a probability measure on T_{t-1} combining five tagged channel distributions, with prescribed mass α_a in each channel. The null-symbol conventions for the channel spaces and pushforward maps are left implicit; their detailed treatment is given in Appendix C.1. Now define the attribute maps $q_t^{\mathcal{X}} : X \to Z_t$ recursively for $t \geq 1$ by

$$q_t^{\mathcal{X}}(x) := \alpha_{id}(\iota_{id})_{\#}\delta_{q_{t-1}^{\mathcal{X}}(x)} + \sum_{a\in\{B,C\}} \alpha_a(\iota_a)_{\#}\big((q_{t-1}^{\mathcal{X}})_{\#}\kappa_x^a\big) + \sum_{a\in\{\downarrow,\uparrow\}} \alpha_a(\iota_a)_{\#}\big((q_{t-1}^{\mathcal{X}}\times q_{t-1}^{\mathcal{X}})_{\#}\kappa_x^a\big).$$

Thus, $q_t^{\mathcal{X}}(x)$ combines the cell's previous attribute with the four neighborhood attribute distributions. We now construct the neighborhood attribute distributions and label maps recursively:

Step 1. For each cell $x \in X$ and each channel, push the neighborhood attribute distribution forward under the initial attribute map:

$$\nu_{B,x,0}^{\mathcal{X}} := (q_0^{\mathcal{X}})_{\#}\kappa_x^B \in \mathcal{P}(Z_0), \qquad \nu_{C,x,0}^{\mathcal{X}} := (q_0^{\mathcal{X}})_{\#}\kappa_x^C \in \mathcal{P}(Z_0),$$
$$\nu_{\downarrow,x,0}^{\mathcal{X}} := (q_0^{\mathcal{X}} \times q_0^{\mathcal{X}})_{\#}\kappa_x^{\downarrow} \in \mathcal{P}(Z_0^2), \qquad \nu_{\uparrow,x,0}^{\mathcal{X}} := (q_0^{\mathcal{X}} \times q_0^{\mathcal{X}})_{\#}\kappa_x^{\uparrow} \in \mathcal{P}(Z_0^2).$$

Step t ($t \geq 2$). Given $Z_{t-1}, q_{t-1}^{\mathcal{X}}$ from the previous step, define the updated attribute representations for the cell $x \in X$

$$\nu_{B,x,t-1}^{\mathcal{X}} := (q_{t-1}^{\mathcal{X}})_{\#}\kappa_x^B, \qquad \nu_{C,x,t-1}^{\mathcal{X}} := (q_{t-1}^{\mathcal{X}})_{\#}\kappa_x^C,$$
$$\nu_{\downarrow,x,t-1}^{\mathcal{X}} := (q_{t-1}^{\mathcal{X}} \times q_{t-1}^{\mathcal{X}})_{\#}\kappa_x^{\downarrow}, \qquad \nu_{\uparrow,x,t-1}^{\mathcal{X}} := (q_{t-1}^{\mathcal{X}} \times q_{t-1}^{\mathcal{X}})_{\#}\kappa_x^{\uparrow},$$

where $(q_{k-1}^{\mathcal{X}} \times q_{k-1}^{\mathcal{X}})(y,z) = (q_{k-1}^{\mathcal{X}}(y), q_{k-1}^{\mathcal{X}}(z))$.

Global distributional representations. To represent the entire complex by a distribution of its cell attributes, we equip the cell set with a base probability measure. For each finite nonempty ACC $\mathcal{X}$, let $\mathcal{R}_{\mathcal{X}} := \{r \geq 0 : X_r \neq \emptyset\}$ denote its rank support. Fix positive weights $w_r > 0$ for all $r \geq 0$ and define

$$s_{\mathcal{X}} := \sum_{r\in\mathcal{R}_{\mathcal{X}}} w_r, \qquad \mu_{\mathcal{X}} := \sum_{r\in\mathcal{R}_{\mathcal{X}}} \frac{w_r}{s_{\mathcal{X}}|X_r|} \sum_{x\in X_r} \delta_x. \tag{2}$$

Thus, each rank receives mass proportional to w_r, distributed uniformly among its cells. The measure $\mu_{\mathcal{X}}$ serves as the base measure on the cell set. At each depth t, this measure induces the global attribute distribution

$$L_t(\mathcal{X}) := (q_t^{\mathcal{X}})_{\#}\mu_{\mathcal{X}} = \sum_{r\in\mathcal{R}_{\mathcal{X}}} \frac{w_r}{s_{\mathcal{X}}|X_r|} \sum_{x\in X_r} \delta_{q_t^{\mathcal{X}}(x)} \in \mathcal{P}(Z_t). \tag{3}$$

3.2 The CCWL distance

We next equip these distributional representations with recursively defined Wasserstein metrics to obtain a distance between complexes.

Recursive ground metrics. Starting from the bounded metric d_0 on Z_0, we recursively define a metric d_t on Z_t for each $t \geq 1$. Suppose that Z_{t-1}, d_{t-1} has been constructed. For the identity, boundary, and coboundary channels, define for $z, z' \in Z_{t-1}$,

$$\rho_{a,t-1}(z,z') := d_{t-1}(z,z'), \qquad a \in \{id, B, C\}.$$

For the lower and upper channels, define

$$\rho_{a,t-1}\big((z_1,z_2),(z_1',z_2')\big) := \frac{1}{2}\left[d_{t-1}(z_1,z_1') + d_{t-1}(z_2,z_2')\right], \qquad a \in \{\downarrow,\uparrow\}.$$

On the channel-tagged attribute space T_{t-1}, we use the extended ground cost to avoid mixing comparisons across neighbor channels

$$\rho_{t-1}\big((a,u),(b,v)\big) := \begin{cases} \rho_{a,t-1}(u,v), & a = b, \\ +\infty, & a \neq b. \end{cases} \tag{4}$$

We then define

$$d_t(\eta, \eta') := W(\eta, \eta') = \inf_{\pi \in \Pi(\eta, \eta')} \int_{\mathsf{T}_{t-1} \times \mathsf{T}_{t-1}} \rho_{t-1}(u, v)\, d\pi(u, v), \qquad \eta, \eta' \in Z_t. \tag{5}$$

The prescribed channel masses ensure that a channel-preserving coupling of finite cost exists, while the infinite cross-channel cost forces every finite-cost coupling to preserve channel types. Consequently, d_t decomposes into a weighted sum of channel-wise Wasserstein distances, as shown in Proposition C.1 (Appendix C.2).

Definition 3.1 (CCWL distance)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite nonempty ACCs. For each step $t \geq 0$, let $L_t(\mathcal{X}), L_t(\mathcal{Y}) \in \mathcal{P}(Z_t)$ be their global attribute distributions defined by 3, and let d_t be the recursively defined ground metric on Z_t introduced above. For $t \geq 0$, define the depth-t CCWL distance by*

$$D^{(t)}_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) := W_{d_t}\big(L_t(\mathcal{X}), L_t(\mathcal{Y})\big).$$

where W_{d_t} denotes the 1-Wasserstein distance on $\mathcal{P}(Z_t)$ with ground metric d_t. The CCWL distance is then

$$D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) := \sup_{t \geq 0} D^{(t)}_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}).$$

The following proposition then establishes that D_{CCWL} is a pseudometric on finite nonempty ACCs: it is nonnegative, symmetric, vanishes for identical complexes, and satisfies the triangle inequality. The detailed proof is given in Appendix C.3.

Proposition 3.2 (Pseudometricity)**.** *Let $\mathfrak{C}$ be the class of finite nonempty ACCs, each equipped with an initial attribute map $q_0^{\mathcal{X}} : X \to Z_0$ into a common Polish space (Z_0, d_0) with a bounded complete metric. Fix positive channel weights α_a, $a \in I = \{id, B, C, \downarrow, \uparrow\}$, satisfying $\sum_{a \in I} \alpha_a = 1$, and positive rank weights w_r for every $r \geq 0$, shared across $\mathfrak{C}$. For each $\mathcal{X} \in \mathfrak{C}$, define its base measure $\mu_{\mathcal{X}}$ as in 2 and global representations $L_t(\mathcal{X})$ at step t as in 3. Choose $M > 0$ with $d_0 < M$ and assign null-to-non-null cost M within each neighborhood channel. Then*

$$D_{\mathrm{CCWL}} : \mathfrak{C} \times \mathfrak{C} \to [0, M],$$

defined in Definition 3.1, is a pseudometric on $\mathfrak{C}$.

3.3 Equivalence with CCWL distinguishability

To show that the CCWL distance quantifies differences between complexes while preserving exactly the distinguishing power of the CCWL test, it requires an augmentation of initial attributes to retain the multiplicity information lost when multisets are normalized into probability measures. Specifically, define for an ACC $\mathcal{X}$ that

$$\boldsymbol{m}_{\mathcal{X}}(x) := \big(|\mathcal{N}_B(x)|, |\mathcal{N}_C(x)|, |\mathcal{A}_{\downarrow}(x)|, |\mathcal{A}_{\uparrow}(x)|\big), \forall x \in X; \qquad \boldsymbol{n}(\mathcal{X}) := (|X_r|)_{r \geq 0},$$

where $|\cdot|$ denotes cardinality, and absent ranks have cardinality zero, so $\boldsymbol{n}(\mathcal{X})$ has finite support. For the following result, use the initial attribute map

$$q_0^{\mathcal{X}}(x) := \big(c_{\mathcal{X}}(x), \mathrm{rk}_{\mathcal{X}}(x), \boldsymbol{m}_{\mathcal{X}}(x), \boldsymbol{n}(\mathcal{X})\big). \tag{6}$$

The additional counts allow recovery of neighborhood and global attribute multiplicities: $\boldsymbol{m}_{\mathcal{X}}(x)$ retains the cardinalities of the four neighborhood contexts, while $\boldsymbol{n}(\mathcal{X})$ retains the rank-wise cell counts of the entire complex. Such augmentation of initial attributes preserves the CCWL test equivalence, although they may change the first distinguishing depth; see Lemma C.5 in Appendix C.4. We obtain the following result by first proving correspondence with this augmented CCWL test and then using its equivalence to the original test.

Proposition 3.3 (Equivalence with CCWL distinguishability)**.** *Under the assumptions of Proposition 3.2, for any two finite nonempty ACCs $\mathcal{X}, \mathcal{Y} \in \mathfrak{C}$, use the initial attribute maps $q_0^{\mathcal{X}}, q_0^{\mathcal{Y}}$ as defined in 6, then*

$$D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) = 0 \quad \Longleftrightarrow \quad \mathcal{X} \equiv_{\mathrm{CCWL}} \mathcal{Y},$$

i.e. the CCWL test distinguishes $\mathcal{X}$ and $\mathcal{Y}$ if and only if $D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) > 0$.

Corollary 3.4 (Consequence for expressivity of TNNs)**.** *Let $\mathcal{F}$ be any message-passing TNN family whose expressive power is upper bounded by CCWL as in (1), i.e. $\mathcal{X} \equiv_{\mathrm{CCWL}} \mathcal{Y} \Longrightarrow f(\mathcal{X}) = f(\mathcal{Y}), \forall f \in \mathcal{F}$. Then*

$$D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) = 0 \quad \Longrightarrow \quad f(\mathcal{X}) = f(\mathcal{Y}) \qquad \forall f \in \mathcal{F}.$$

4 Relation with Co-Optimal Transport

We now switch our view from the WL-type geometry to the optimal transport geometry and provide a similarity measure between ACCs directly in terms of cell attributes and adjacent-rank incidence via co-optimal transport (Co-OT). Here, we optimize one transport plan per rank, and these plans couple cells jointly at adjacent ranks. The transport ground cost measures discrepancies between their incidence relations. Then we replace its structural incidence mismatch objective by a convex penalty on fractional matchings. Our main result is that the resulting convex Co-OT distance has the same distinguishability as the CCWL distance. All Co-OT constructions below include attributes, and we omit the prefix "attributed" throughout.

4.1 Combinatorial Complex Co-Optimal Transport

Given ACCs $\mathcal{X}$ and $\mathcal{Y}$, let $a_{\mathcal{X}}(x) := q_0^{\mathcal{X}}(x)$ denote the augmented initial attributes in 6, equipped with the common bounded metric d_0. For non-empty ranks, write $n_r^{\mathcal{X}} := |X_r|$, $r \geq 0$, and define the attribute cost $C_r(x, y) := d_0(a_{\mathcal{X}}(x), a_{\mathcal{Y}}(y))$. For $r \geq 1$, define the adjacent-rank incidence matrix by

$$I_r^{\mathcal{X}}(x, x') := \mathbf{1}[x \subseteq x'], \qquad (x, x') \in X_{r-1} \times X_r,$$

and define $I_r^{\mathcal{Y}}$ analogously. Missing ranks are represented by a single null cell with zero incidence, as detailed in D.1.

Let $\boldsymbol{\pi} = (\pi_r)_{r \geq 0}$ be a feasible coupling family, where π_r couples the uniform probability measures on X_r and Y_r. The Co-OT incidence cost is

$$\mathcal{I}_r(\boldsymbol{\pi}) := \sum_{\substack{x \in X_{r-1},\, x' \in X_r \\ y \in Y_{r-1},\, y' \in Y_r}} |I_r^{\mathcal{X}}(x, x') - I_r^{\mathcal{Y}}(y, y')| \pi_{r-1}(x, y) \pi_r(x', y').$$

This cost encourages consistency between couplings at adjacent ranks: matched cell pairs are penalized when their incidence relations disagree. However, the objective is bilinear in the adjacent couplings and generally nonconvex. A standard implementation approach is block-coordinate descent, which updates one transport plan while holding others fixed (55).

We instead construct a convex relaxation by replacing the expected absolute incidence mismatch with two averaged mismatch residuals $R_r^{\downarrow}$ and $R_r^{\uparrow}$, one for each direction of incidence aggregation. Each residual compares *couplings followed by incidence aggregation* with *incidence aggregation followed by couplings*:

$$\begin{array}{ccc} \mathbb{R}^{Y_r} & \xrightarrow{\pi_r} & \mathbb{R}^{X_r} \\ {\scriptstyle I_r^{\mathcal{Y}}/n_r^{\mathcal{Y}}}\downarrow & & \downarrow{\scriptstyle I_r^{\mathcal{X}}/n_{r-1}^{\mathcal{X}}} \\ \mathbb{R}^{Y_{r-1}} & \xrightarrow[\pi_{r-1}]{} & \mathbb{R}^{X_{r-1}} \end{array} \qquad \begin{array}{ccc} \mathbb{R}^{Y_r} & \xrightarrow{\pi_r} & \mathbb{R}^{X_r} \\ {\scriptstyle (I_r^{\mathcal{Y}})^\top/n_{r-1}^{\mathcal{Y}}}\uparrow & & \uparrow{\scriptstyle (I_r^{\mathcal{X}})^\top/n_r^{\mathcal{X}}} \\ \mathbb{R}^{Y_{r-1}} & \xrightarrow[\pi_{r-1}]{} & \mathbb{R}^{X_{r-1}} \end{array} \tag{7}$$

$$R_r^{\downarrow}(\boldsymbol{\pi}) := \frac{I_r^{\mathcal{X}} \pi_r}{n_{r-1}^{\mathcal{X}}} - \frac{\pi_{r-1} I_r^{\mathcal{Y}}}{n_r^{\mathcal{Y}}}, \qquad R_r^{\uparrow}(\boldsymbol{\pi}) := \frac{(I_r^{\mathcal{X}})^\top \pi_{r-1}}{n_r^{\mathcal{X}}} - \frac{\pi_r (I_r^{\mathcal{Y}})^\top}{n_{r-1}^{\mathcal{Y}}}.$$

Here, $\mathbb{R}^{X_r}$ denotes the space of real-valued functions on X_r, and each arrow acts by matrix multiplication. Each diagram commutes exactly when its corresponding residual vanishes. We penalize these failures to commute through

$$\mathcal{R}_r(\boldsymbol{\pi}) := \frac{1}{2} \left(\|R_r^{\downarrow}(\boldsymbol{\pi})\|_1 + \|R_r^{\uparrow}(\boldsymbol{\pi})\|_1 \right), \|H\|_1 := \sum_{i,j} |H_{ij}|.$$

Both residuals are linear in the coupling family, so $\mathcal{R}_r$ is jointly convex.

Definition 4.1 (Co-OT and convex Co-OT)**.** *Fix weights $\beta_r > 0$ for $r \geq 0$ and $\tau_r > 0$ for $r \geq 1$. Define for ACCs $\mathcal{X}, \mathcal{Y}$ the Co-OT and convex Co-OT distances*

$$\begin{aligned} D_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) &:= \min_{\boldsymbol{\pi} = (\pi_r)_{r \geq 0}} \left\{ \sum_{r \geq 0} \beta_r \langle C_r, \pi_r \rangle + \sum_{r \geq 1} \tau_r \mathcal{I}_r(\boldsymbol{\pi}) \right\}, \\ D_{\text{Co-OT}}^{\text{cvx}}(\mathcal{X}, \mathcal{Y}) &:= \min_{\boldsymbol{\pi} = (\pi_r)_{r \geq 0}} \left\{ \sum_{r \geq 0} \beta_r \langle C_r, \pi_r \rangle + \sum_{r \geq 1} \tau_r \mathcal{R}_r(\boldsymbol{\pi}) \right\}. \end{aligned} \tag{8}$$

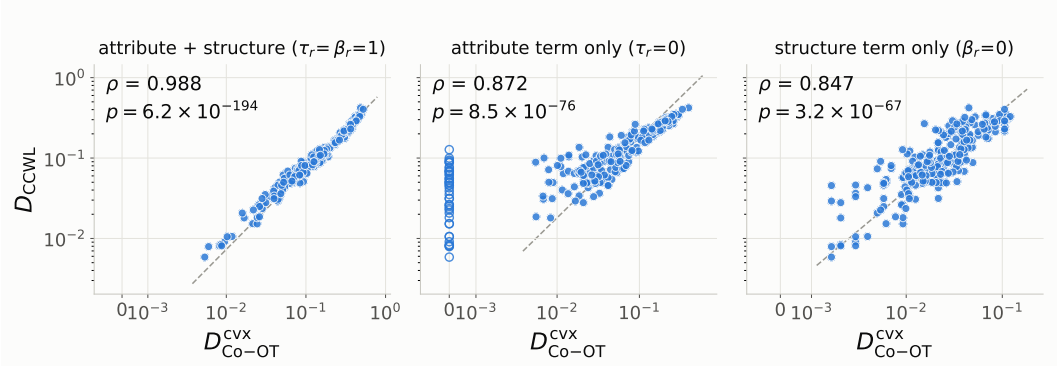


Figure 2: D_{CCWL} *against* $D^{\mathrm{cvx}}_{Co\text{-}OT}$ *on* 240 *pairs of attributed rank-*2 *complexes, for the full cost (left), the attribute channel alone (centre) and the incidence channel alone (right). Hollow markers are pairs with zero attribute cost.*

where $\boldsymbol{\pi} = (\pi_r)_{r\geq 0}$ *is constrained so that each per-rank coupling* π_r *has uniform marginals over the sets of rank-*r *cells.*

Both distances are pseudometrics, and convex Co-OT admits a linear-program formulation. It relaxes the objective on the same transport polytopes: $D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}(\mathcal{X},\mathcal{Y}) \leq D_{\mathrm{Co\text{-}OT}}(\mathcal{X},\mathcal{Y})$ for all $\mathcal{X},\mathcal{Y} \in \mathfrak{C}$.. Precise statements and proofs are given in D.2 and D.3. The main result of this work is the following:

Proposition 4.2 (CCWL equivalence of convex Co-OT)**.** *For any finite nonempty attributed combinatorial complexes* $\mathcal{X},\mathcal{Y}$ *with the augmented initial attributes in 6,*

$$D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}(\mathcal{X},\mathcal{Y}) = 0 \iff \mathcal{X} \equiv_{\mathrm{CCWL}} \mathcal{Y} \iff D_{\mathrm{CCWL}}(\mathcal{X},\mathcal{Y}) = 0.$$

The proof is in D.4. In contrast, unrelaxed Co-OT has a strictly smaller zero set than the CCWL distance and can therefore distinguish complexes that CCWL cannot; see D.5 for a counterexample.

4.2 Strong correlation between D_{CCWL} and $D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}$

We also empirically investigate how D_{CCWL} and $D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}$ agree on measuring the similarities of the distinguishable pairs. We randomly sample 240 pairs of attributed rank-2 combinatorial complexes of varying sizes by progressively sampling vertices, edges, and triangles (rank-2 cells). We calculate the D_{CCWL} and $D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}$ between the pairs and plot the correlations in Figure 2. D_{CCWL} is strongly correlated with $D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}$. To check that the association is not carried by the attribute term alone, we ablate the transport cost to both the attribute and structural terms separately. In some cases, the attribute term alone is not discriminative: it vanishes on 44 of the 240 pairs, all of them equal-cardinality pairs whose attribute multisets agree while their incidence differs, whereas neither the structure term nor the full cost vanishes on any pair. So the high attribute-only correlation reflects agreement in ordering, not an ability to separate.

We provide evidence for the effectiveness of the geometry induced by D_{CCWL} and $D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}$ through 1-Nearest Neighbor classification and TNN regularization in B.

5 Conclusion

We introduced a CCWL distance that quantifies differences between attributed combinatorial complexes through recursive Wasserstein comparisons of neighborhood distributions. This construction defines a pseudometric that vanishes exactly when CCWL cannot distinguish the complexes. We further formulated a co-optimal transport pseudometric and a convex relaxation based on consistency between rank-wise couplings and incidence maps. The convex relaxation has exactly the same zero set as the CCWL distance, whereas unrelaxed Co-OT can distinguish some CCWL-equivalent complexes. These results connect CCWL indistinguishability, and hence the message-passing TNN expressivity, to an optimal transport formulation. Future work includes establishing quantitative relationships between these distances and TNN representations, developing generalizability theories, and designing scalable transport algorithms for larger complexes.

AI use statement

In this work, we used generative AI tools for formulating mathematical claims, especially for assist in the writing of proofs, and providing feedback on the experimental section. We have not used generative AI tools for other tasks with required disclosure. Additionally, we used generative AI tools for modifying scientific figures or images. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

References

[1] Amirreza Akbari, Amauri H. Souza, and Vikas Garg. The logical expressiveness of topological neural networks. In *International Conference on Learning Representations*, 2026.

[2] Mathieu Alain, So Takao, Brooks Paige, and Marc Peter Deisenroth. Gaussian processes on cellular complexes. In *Forty-first International Conference on Machine Learning*, 2024.

[3] Mayar Ali, Sabrina Richter, Ali Ertürk, David S Fischer, and Fabian J Theis. Graph neural networks learn emergent tissue properties from spatial molecular profiles. *Nature Communications*, 16(1):8419, 2025.

[4] Sergio Barbarossa and Stefania Sardellitti. Topological signal processing over simplicial complexes. *IEEE Transactions on Signal Processing*, 68:2992–3007, 2020.

[5] Federico Barbero, Cristian Bodnar, Haitz Sáez de Ocáriz Borde, Michael Bronstein, Petar Veličković, and Pietro Liò. Sheaf neural networks with connection laplacians. In Alexander Cloninger, Timothy Doster, Tegan Emerson, Manohar Kaul, Ira Ktena, Henry Kvinge, Nina Miolane, Bastian Rieck, Sarah Tymochko, and Guy Wolf, editors, *Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022*, volume 196 of *Proceedings of Machine Learning Research*, pages 28–36. PMLR, 25 Feb–22 Jul 2022.

[6] Claudio Battiloro, Paolo Di Lorenzo, and Sergio Barbarossa. Topological slepians: Maximally localized representations of signals over simplicial complexes. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023.

[7] Claudio Battiloro, Ege Karaismailoğlu, Mauricio Tec, George Dasoulas, Michelle Audirac, and Francesca Dominici. E(n) equivariant topological neural networks. In *International Conference on Learning Representations*, 2025.

[8] Claudio Battiloro, Indro Spinelli, Lev Telyatnikov, Michael Bronstein, Simone Scardapane, and Paolo Di Lorenzo. From latent graph to latent topology inference: Differentiable cell complex module. *arXiv preprint arXiv:2305.16174*, 2023.

[9] Claudio Battiloro, Lucia Testa, Lorenzo Giusti, Stefania Sardellitti, Paolo Di Lorenzo, and Sergio Barbarossa. Generalized simplicial attention neural networks. *IEEE Transactions on Signal and Information Processing over Networks*, 10:833–850, 2024.

[10] Claudio Battiloro, Zhiyang Wang, Hans Riess, Paolo Di Lorenzo, and Alejandro Ribeiro. Tangent bundle convolutional learning: from manifolds to cellular sheaves and back. *IEEE Transactions on Signal Processing*, 2024.

[11] Austin R Benson, Rediet Abebe, Michael T Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, 115(48):E11221–E11230, 2018.

[12] Guillermo Bernárdez, Lev Telyatnikov, Marco Montagna, Federica Baccini, Mathilde Papillon, Miquel Ferriol-Galmés, Mustafa Hajij, Theodore Papamarkou, Maria Sofia Bucarelli, Olga Zaghen, Johan Mathe, Audun Myers, Scott Mahan, Hansen Lillemark,

Sharvaree Vadgama, Erik Bekkers, Tim Doster, Tegan Emerson, Henry Kvinge, Katrina Agate, Nesreen K Ahmed, Pengfei Bai, Michael Banf, Claudio Battiloro, Maxim Beketov, Paul Bogdan, Martin Carrasco, Andrea Cavallo, Yun Young Choi, George Dasoulas, Matouš Elphick, Giordan Escalona, Dominik Filipiak, Halley Fritze, Thomas Gebhart, Manel Gil-Sorribes, Salvish Goomanee, Victor Guallar, Liliya Imasheva, Andrei Irimia, Hongwei Jin, Graham Johnson, Nikos Kanakaris, Boshko Koloski, Veljko Kovač, Manuel Lecha, Minho Lee, Pierrick Leroy, Theodore Long, German Magai, Alvaro Martinez, Marissa Masden, Sebastian Mežnar, Bertran Miquel-Oliver, Alexis Molina, Alexander Nikitin, Marco Nurisso, Matt Piekenbrock, Yu Qin, Patryk Rygiel, Alessandro Salatiello, Max Schattauer, Pavel Snopov, Julian Suk, Valentina Sánchez, Mauricio Tec, Francesco Vaccarino, Jonas Verhellen, Frederic Wantiez, Alexander Weers, Patrik Zajec, Blaž Škrlj, and Nina Miolane. Icml topological deep learning challenge 2024: Beyond the graph domain. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM)*, volume 251 of *Proceedings of Machine Learning Research*, pages 420–428, 2024.

[13] Maciej Besta, Florian Scheidl, Lukas Gianinazzi, Grzegorz Kwasniewski, Shachar Klaiman, Jürgen Müller, and Torsten Hoefler. Demystifying higher-order graph neural networks. *arXiv preprint arXiv:2406.12841*, 2024.

[14] Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Lio, Guido F Montufar, and Michael Bronstein. Weisfeiler and lehman go cellular: Cw networks. *Advances in neural information processing systems*, 34:2625–2640, 2021.

[15] Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Lio, Guido F Montufar, and Michael Bronstein. Weisfeiler and lehman go cellular: Cw networks. *Advances in neural information processing systems*, 34:2625–2640, 2021.

[16] Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montufar, Pietro Lio, and Michael Bronstein. Weisfeiler and lehman go topological: Message passing simplicial networks. In *International conference on machine learning*, pages 1026–1037. PMLR, 2021.

[17] Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montufar, Pietro Lio, and Michael Bronstein. Weisfeiler and lehman go topological: Message passing simplicial networks. In *International conference on machine learning*, pages 1026–1037. PMLR, 2021.

[18] Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Lio, and Michael M. Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs. In *Advances in Neural Information Processing Systems*, 2022.

[19] Jiawen Chen, Qi Shao, Zhiqiang Ge, Duxin Chen, and Wenwu Yu. Weisfeiler lehman test on combinatorial complexes: Generalized expressive power of topological neural networks. *arXiv preprint arXiv:2605.00725*, 2026.

[20] Samantha Chen, Sunhyuk Lim, Facundo Mémoli, Zhengchao Wan, and Yusu Wang. Weisfeiler-lehman meets gromov-wasserstein. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 2022.

[21] Samantha Chen, Sunhyuk Lim, Facundo Mémoli, Zhengchao Wan, and Yusu Wang. The weisfeiler-lehman distance: Reinterpretation and connection with gnns. In *Topological, algebraic and geometric learning workshops 2023*, pages 404–425. PMLR, 2023.

[22] Samir Chowdhury, Tom Needham, Ethan Semrad, Bei Wang, and Youjia Zhou. Hypergraph co-optimal transport: metric and categorical properties: S. chowdhury et al. *Journal of Applied and Computational Topology*, 8(5):1171–1230, 2024.

[23] Ching-Yao Chuang and Stefanie Jegelka. Tree mover's distance: Bridging graph metrics and stability of graph neural networks. *Advances in Neural Information Processing Systems*, 35:2944–2957, 2022.

[24] Gabriele D'Acunto and Claudio Battiloro. The relativity of causal knowledge. In Silvia Chiappa and Sara Magliacane, editors, *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pages 863–881. PMLR, 21–25 Jul 2025.

[25] Stefania Ebli, Michaël Defferrard, and Gard Spreemann. Simplicial neural networks. *arXiv preprint arXiv:2010.03633*, 2020.

[26] Yam Eitan, Yoav Gelberg, Guy Bar-Shalom, Fabrizio Frasca, Michael Bronstein, and Haggai Maron. Topological blind spots: Understanding and extending topological deep learning through the lens of expressivity. *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

[27] L. Giusti, C. Battiloro, P. Di Lorenzo, S. Sardellitti, and S. Barbarossa. Simplicial attention neural networks, 2022.

[28] Lorenzo Giusti, Claudio Battiloro, Lucia Testa, Paolo Di Lorenzo, Stefania Sardellitti, and Sergio Barbarossa. Cell attention networks. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2023.

[29] Lorenzo Giusti, Teodora Reu, Francesco Ceccarelli, Cristian Bodnar, and Pietro Liò. Cin++: Enhancing topological message passing. *arXiv preprint arXiv:2306.03561*, 2023.

[30] M. Hajij, G. Zamzmi, T. Papamarkou, N. Miolane, A. Guzmán-Sáenz, and K. N. Ramamurthy. Higher-order attention networks, 2022.

[31] Mustafa Hajij et al. Topological deep learning: Going beyond graph data. *arXiv preprint arXiv:2206.00606*, 2023.

[32] Mustafa Hajij, Kyle Istvan, and Ghada Zamzmi. Cell complex neural networks. *arXiv preprint arXiv:2010.00743*, 2020.

[33] Mustafa Hajij, Mathilde Papillon, Florian Frantzen, Jens Agerberg, Ibrahem AlJabea, Rubén Ballester, Claudio Battiloro, Guillermo Bernárdez, Tolga Birdal, Aiden Brent, Peter Chin, Sergio Escalera, Simone Fiorellino, Odin Hoff Gardaa, Gurusankar Gopalakrishnan, Devendra Govil, Josef Hoppe, Maneel Reddy Karri, Jude Khouja, Manuel Lecha, Neal Livesay, Jan Meißner, Soham Mukherjee, Alexander Nikitin, Theodore Papamarkou, Jaro Prílepok, Karthikeyan Natesan Ramamurthy, Paul Rosen, Aldo Guzmán-Sáenz, Alessandro Salatiello, Shreyas N. Samaga, Simone Scardapane, Michael T. Schaub, Luca Scofano, Indro Spinelli, Lev Telyatnikov, Quang Truong, Robin Walters, Maosheng Yang, Olga Zaghen, Ghada Zamzmi, Ali Zia, and Nina Miolane. Topox: A suite of python packages for machine learning on topological domains. *Journal of Machine Learning Research*, 25(374):1–8, 2024.

[34] Mustafa Hajij, Ghada Zamzmi, Theodore Papamarkou, Nina Miolane, Aldo Guzmán-Sáenz, Karthikeyan Natesan Ramamurthy, Tolga Birdal, Tamal K Dey, Soham Mukherjee, Shreyas N Samaga, et al. Topological deep learning: Going beyond graph data. *arXiv preprint arXiv:2206.00606*, 2022.

[35] J. Hansen and R. Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 2019.

[36] Jakob Hansen and Thomas Gebhart. Sheaf neural networks, 2020.

[37] Josef Hoppe, Vincent P Grande, and Michael T Schaub. Don't be afraid of cell complexes! an introduction from an applied perspective. *arXiv preprint arXiv:2506.09726*, 2025.

[38] Elvin Isufi, Geert Leus, Baltasar Beferull-Lozano, Sergio Barbarossa, and Paolo Di Lorenzo. Topological signal processing and learning: Recent advances and future challenges. *Signal Processing*, page 109930, 2025.

[39] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[40] Manuel Lecha, Andrea Cavallo, Francesca Dominici, Elvin Isufi, and Claudio Battiloro. Higher-order topological directionality and directed simplicial neural networks. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025.

[41] Manuel Lecha, Andrea Cavallo, Francesca Dominici, Ran Levi, Alessio Del Bue, Elvin Isufi, Pietro Morerio, and Claudio Battiloro. Directed semi-simplicial learning with applications to brain activity decoding. In *International Conference on Learning Representations*, 2026.

[42] Lek-Heng Lim. Hodge laplacians on graphs. *Siam Review*, 62(3):685–715, 2020.

[43] Abdul Majeed and Ibtisam Rauf. Graph theory: A comprehensive survey about graph theory applications in computer science and social networks. *Inventions*, 5(1):10, 2020.

[44] Lorenzo Marinucci, Claudio Battiloro, and Paolo Di Lorenzo. Topological adaptive least mean squares algorithms over simplicial complexes. *IEEE Transactions on Signal and Information Processing over Networks*, 12:283–298, 2026.

[45] Christopher Morris, Nils M. Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. In *ICML 2020 Workshop on Graph Representation Learning and Beyond (GRL+ 2020)*, 2020.

[46] Christopher Morris, Yaron Lipman, Haggai Maron, Bastian Rieck, Nils M Kriege, Martin Grohe, Matthias Fey, and Karsten Borgwardt. Weisfeiler and leman go machine learning: The story so far. *Journal of Machine Learning Research*, 24(333):1–59, 2023.

[47] Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4602–4609, 2019.

[48] Theodore Papamarkou, Tolga Birdal, Michael Bronstein, Gunnar Carlsson, Justin Curry, Yue Gao, Mustafa Hajij, Roland Kwitt, Pietro Lio, Paolo Di Lorenzo, et al. Position: Topological deep learning is the new frontier for relational learning. *Proceedings of machine learning research*, 235:39529, 2024.

[49] Mathilde Papillon, Guillermo Bernardez, Claudio Battiloro, and Nina Miolane. Topotune: A framework for generalized combinatorial complex neural networks. In *Forty-second International Conference on Machine Learning*, 2025.

[50] Mathilde Papillon, Mustafa Hajij, Audun Myers, Helen Jenne, Johan Mathe, Theodore Papamarkou, Aldo Guzmán-Sáenz, Neal Livesay, Tamal Dey, Abraham Rabinowitz, Aiden Brent, Alessandro Salatiello, Alexander Nikitin, Ali Zia, Claudio Battiloro, Dmitrii Gavrilev, German Magai, Gleb Bazhenov, Guillermo Bernardez, Indro Spinelli, Jens Agerberg, Kalyan Nadimpalli, Lev Telyatnikov, Luca Scofano, Lucia Testa, Manuel Lecha, Maosheng Yang, Mohammed Hassanin, Odin Hoff Gardaa, Olga Zaghen, Paul Hausner, Paul Snopoff, Rubén Ballester, Sadrodin Barikbin, Sergio Escalera, Simone Fiorellino, Henry Kvinge, Karthikeyan Natesan Ramamurthy, Paul Rosen, Robin Walters, Shreyas N. Samaga, Soham Mukherjee, Sophia Sanborn, Tegan Emerson, Timothy Doster, Tolga Birdal, Abdelwahed Khamis, Simone Scardapane, Suraj Singh, Tatiana Malygina, Yixiao Yue, and Nina Miolane. Icml 2023 topological deep learning challenge: Design and results. In *Proceedings of 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*, volume 221 of *Proceedings of Machine Learning Research*, pages 3–8, 2023.

[51] Mathilde Papillon, Sophia Sanborn, Mustafa Hajij, and Nina Miolane. Architectures of topological deep learning: A survey of message-passing topological neural networks. *arXiv preprint arXiv:2304.10031*, 2023.

[52] Mathilde Papillon, Sophia Sanborn, Johan Mathe, Louisa Cornelis, Abby Bertics, Domas Buracas, Hansen J Lillemark, Christian Shewmake, Fatih Dinc, Xavier Pennec, and Nina Miolane. Beyond euclid: an illustrated guide to modern machine learning with geometric, topological, and algebraic structures. *Machine Learning: Science and Technology*, 6(3):031002, aug 2025.

[53] Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In *International conference on machine learning*, pages 2664–2672. PMLR, 2016.

[54] Levi Rauchwerger, Stefanie Jegelka, and Ron Levie. Generalization, expressivity, and universality of graph neural networks on attributed graphs. In *International Conference on Learning Representations*, volume 2025, pages 91497–91557, 2025.

[55] Ievgen Redko, Titouan Vayer, Rémi Flamary, and Nicolas Courty. Co-optimal transport. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[56] T. M. Roddenberry, N. Glaze, and S. Segarra. Principled simplicial neural networks for trajectory prediction. In *International Conference on Machine Learning*, 2021.

[57] T. Mitchell Roddenberry, Michael T. Schaub, and Mustafa Hajij. Signal processing on cell complexes. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8852–8856, 2022.

[58] S. Sardellitti, S. Barbarossa, and L. Testa. Topological signal processing over cell complexes. In *Asilomar Conference on Signals, Systems, and Computers*, 2021.

[59] Michael T Schaub, Austin R Benson, Paul Horn, and Ali Jadbabaie. Signal processing on higher-order networks: Living on the edge... and beyond. *Signal Processing*, 180:107946, 2021.

[60] Lev Telyatnikov, Guillermo Bernardez, Marco Montagna, Mustafa Hajij, Martin Carrasco, Pavlo Vasylenko, Mathilde Papillon, Ghada Zamzmi, Michael T. Schaub, Jonas Verhellen, Pavel Snopov, Bertran Miquel-Oliver, Manel Gil-Sorribes, Alexis Molina, Victor Guallar, Theodore Long, Julian Suk, Patryk Rygiel, Alexander Nikitin, Giordan Escalona, Michael Banf, Dominik Filipiak, Max Schattauer, Liliya Imasheva, Alvaro Martinez, Halley Fritze, Marissa Masden, Valentina Sánchez, Manuel Lecha, Andrea Cavallo, Claudio Battiloro, Matt Piekenbrock, Mauricio Tec, George Dasoulas, Nina Miolane, Simone Scardapane, and Theodore Papamarkou. Topobench: A framework for benchmarking topological deep learning, 2025.

[61] Lev Telyatnikov, Guillermo Bernardez, Marco Montagna, Mustafa Hajij, Martin Carrasco, Pavlo Vasylenko, Mathilde Papillon, Ghada Zamzmi, Michael T Schaub, Jonas Verhellen, Pavel Snopov, Bertran Miquel-Oliver, Manel Gil-Sorribes, Alexis Molina, VICTOR GUALLAR, Theodore Long, Julian Suk, Patryk Rygiel, Alexander V Nikitin, Giordan Escalona, Michael Banf, Dominik Filipiak, Liliya Imasheva, Max Schattauer, Alvaro L. Martinez, Halley Fritze, Marissa Masden, Valentina Sánchez, Manuel Lecha, Andrea Cavallo, Claudio Battiloro, Matthew Piekenbrock, Mauricio Tec, George Dasoulas, Nina Miolane, Simone Scardapane, and Theodore Papamarkou. Topobench: A framework for benchmarking topological deep learning. *Journal of Data-centric Machine Learning Research*, 2025.

[62] Titouan Vayer, Laetita Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused gromov-wasserstein distance for structured objects: theoretical foundations and mathematical properties. *arXiv preprint arXiv:1811.02834*, 2018.

[63] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[64] Yuyang Wang, Zijie Li, and Amir Barati Farimani. Graph neural networks for molecules. In *Machine learning in molecular sciences*, pages 21–66. Springer, 2023.

[65] Yuyang Wang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3):279–287, 2022.

[66] Boris Weisfeiler and Andrei Leman. The reduction of a graph to canonical form and the algebra which appears therein. *nti, Series*, 2(9):12–16, 1968.

[67] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

[68] Maosheng Yang, Viacheslav Borovitskiy, and Elvin Isufi. Hodge-compositional edge gaussian processes. *International Conference on Artificial Intelligence and Statistics*, 238:754–3762, 2024.

[69] Maosheng Yang and Elvin Isufi. Convolutional learning on simplicial complexes. *arXiv preprint arXiv:2301.11163*, 2023.

[70] Maosheng Yang, Elvin Isufi, and Geert Leus. Simplicial convolutional neural networks. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8847–8851, 2022.

[71] Maosheng Yang, Elvin Isufi, and Geert Leus. Simplicial convolutional neural networks. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8847–8851. IEEE, 2022.

[72] Ruochen Yang, Frederic Sala, and Paul Bogdan. Efficient representation learning for higher-order data with simplicial complexes. In Bastian Rieck and Razvan Pascanu, editors, *Proceedings of the First Learning on Graphs Conference*, volume 198 of *Proceedings of Machine Learning Research*, pages 13:1–13:21. PMLR, 09–12 Dec 2022.

[73] Bohang Zhang, Guhao Feng, Yiheng Du, Di He, and Liwei Wang. A complete expressiveness hierarchy for subgraph gnns via subgraph weisfeiler-lehman tests. In *International Conference on Machine Learning*, pages 41019–41077. PMLR, 2023.

[74] Bohang Zhang, Jingchu Gai, Yiheng Du, Qiwei Ye, Di He, and Liwei Wang. Beyond weisfeiler-lehman: A quantitative framework for gnn expressiveness. In *International Conference on Learning Representations*, volume 2024, pages 53674–53746, 2024.

[75] Bohang Zhang, Lingxiao Zhao, and Haggai Maron. On the expressive power of spectral invariant graph neural networks. *arXiv preprint arXiv:2406.04336*, 2024.

[76] Detian Zhang, Chengqiang Zhang, Yanghui Rao, Li Qing, and Chunjiang Zhu. Improved expressivity of hypergraph neural networks through high-dimensional generalized weisfeiler-leman algorithms. In *Forty-second International Conference on Machine Learning*, 2025.

[77] Xiao-Meng Zhang, Li Liang, Lin Liu, and Ming-Jing Tang. Graph neural networks and their current applications in bioinformatics. *Frontiers in genetics*, 12:690049, 2021.

Appendix

Contents

A Background: Topological deep learning and Higher-order Message Passing 17

B Additional experiments 17

B.1 1-Nearest Neighbor Classification 18

B.2 D_{CCWL} as a regularizer 18

B.3 Correlations with Co-OT and convex Co-OT 18

C Construction details of the CCWL distance 19

C.1 Incorporating Empty Channels 19

C.2 Neighborhood Channel Decomposition of d_t 20

C.3 Proof of Pseudometricity of D_{CCWL} (Proposition 3.2) 21

C.4 Proof of Equivalence with CCWL distinguishability (Proposition 3.3) 23

D Co-OT: formulation, relaxation, and CCWL equivalence 26

D.1 Rank measures and missing ranks 26

D.2 Convexity, attainment, and the relaxation inequality 27

D.3 Pseudometricity 28

D.4 The zero set and CCWL equivalence 29

D.5 Strict separation between Co-OT and CCWL 32

A Background: Topological deep learning and Higher-order Message Passing

Topological deep learning Simplicial and cell complexes extend graphs by representing nodes, edges, faces, and higher-dimensional cells within a unified algebraic framework (59; 37; 13; 11), building on work in topological signal processing (4; 57; 58; 38; 44). This motivates Topological Deep Learning (TDL) (48; 31; 51; 52). Simplicial Neural Networks generalize graph convolutions through Hodge-based filtering (25; 42). Message-passing frameworks extend relational aggregation within and across dimensions (17; 32; 15; 8). Convolutional architectures such as simplicial convolutional filters (SCF), simplicial convolutional neural networks (SCNN) and simplicial complex convolutional neural networks (SCCNN) implement localized multi-hop filtering and explicitly couple simplex orders (71; 72; 69; 56; 30). Attention-based models have incorporated Dirac operators and cell complex attention to generalize graph attention to higher-order domains(9; 27; 28). Expressivity analyses of TDL models have revealed structural blind spots relative to classical WL tests (17; 15; 26). Further extensions include sheaf networks(35; 36; 18; 10; 5; 24), semi-simplicial networks (41; 40), Gaussian processes on simplicial domains (68; 2), continuous architectures citeeinizade2025cosmos, and software benchmarks for TDL (33; 60; 49). These works establish topology as a strong inductive bias for deep learning for relational data.

Higher-order message passing. As a central architectural paradigm in TDL, message-passing TNN architectures on combinatorial complexes are organized by *neighborhood functions*: a cell exchanges information with the cells it is related to, and the choice of relations fixes the model class (31). For a cell $x \in X_r$ of an attributed combinatorial complex $\mathcal{X}$, the adjacent-rank incidences I_r generate four such relations,

$$\mathcal{N}_{\mathcal{X}} = \{\, \mathcal{N}_B,\ \mathcal{N}_C,\ \mathcal{N}_{\downarrow},\ \mathcal{N}_{\uparrow} \,\}, \tag{9}$$

where, for $x \in X_r$,

$$\mathcal{N}_B(x) = \{\, y \in X_{r-1} : I_r[y, x] = 1 \,\}, \qquad \mathcal{N}_{\downarrow}(x) = \{y : \mathcal{N}_B(x) \cap N_B(y) \neq \emptyset\,\}, \tag{10}$$

$$\mathcal{N}_C(x) = \{\, y \in X_{r+1} : I_{r+1}[x, y] = 1 \,\}, \qquad \mathcal{N}_{\uparrow}(x) = \{y : \mathcal{N}_C(x) \cap N_C(y) \neq \emptyset\,\}, \tag{11}$$

are the boundary, coboundary, and lower- and upper-adjacent cells of x.

The t-th layer of message passing updates the embedding $h^{(t)}(x) \in \mathbb{R}^{F_t}$ of a cell x by aggregating within each neighborhood and then across neighborhoods,

$$h^{(t+1)}(x) = \phi\Big(h^{(t)}(x),\ \bigotimes_{\mathcal{N} \in \mathcal{N}_{\mathcal{X}}}\ \bigoplus_{\omega \in \mathcal{N}(x)} \psi_{\mathcal{N},\mathrm{rk}(x)}\big(h^{(t)}(x),\, h^{(t)}(\omega)\big)\Big) \in \mathbb{R}^{F_{t+1}}, \tag{12}$$

where $h^{(0)}(x)$ are the initial cell features, $\bigoplus$ is an intra-neighborhood aggregator, $\bigotimes$ an inter-neighborhood aggregator, and the message maps $\psi_{\mathcal{N},\mathrm{rk}(\cdot)}$ and update map ϕ are learnable, typically shared across neighborhoods and ranks. A complex-level prediction is read out by pooling across ranks such as sum or mean aggregation.

Equation equation 12 is a refinement procedure in disguise, and this is what our analysis attaches to. Taking $\bigoplus$ to be multiset union, $\bigotimes$ to be tupling, and ϕ injective on its five arguments recovers the CCWL refinement procedure.

B Additional experiments

Throughout the experiments, we adopt the following implementation:

Approximation of the D_{CCWL} by truncation. The supremum in Definition 3.1 is replaced by the running maximum

$$\widehat{D}_T(\mathcal{X}, \mathcal{Y}) := \max_{0 \le t \le T} D^{(t)}_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}). \tag{13}$$

For every T this is a pseudometric, being a maximum of finitely many pseudometrics; it is nondecreasing in T, satisfies $\widehat{D}_T \le D_{\mathrm{CCWL}}$, and increases to D_{CCWL} as $T \to \infty$. As a default for the numerical experiments throughout this paper, we set $T = 15$ and use $\widehat{D}_{15}$ as an approximation of D_{CCWL}.

Dataset	WL dist.	Co-OT		$D^{(t)}_{\mathrm{CCWL}}$				
		$D_{\text{Co-OT}}$	$D^{\text{cvx}}_{\text{Co-OT}}$	$t=1$	$t=2$	$t=3$	$t=4$	$\hat{D}_4$
MUTAG	$80.38^{\pm 4.70}$	$84.06^{\pm 4.36}$	$\mathbf{84.15}^{\pm 4.15}$	$82.83^{\pm 2.81}$	$82.45^{\pm 4.54}$	$82.55^{\pm 5.28}$	$82.55^{\pm 5.28}$	$83.57^{\pm 3.61}$
COX2	$74.71^{\pm 3.59}$	$76.80^{\pm 3.30}$	$\mathbf{77.55}^{\pm 3.69}$	$73.69^{\pm 2.97}$	$74.91^{\pm 3.62}$	$74.38^{\pm 3.68}$	$74.72^{\pm 4.08}$	$73.05^{\pm 2.79}$
PROTEINS	$64.12^{\pm 1.37}$	$67.54^{\pm 1.67}$	$\mathbf{67.95}^{\pm 3.31}$	$65.95^{\pm 2.23}$	$67.03^{\pm 2.88}$	$66.88^{\pm 2.85}$	$66.66^{\pm 2.45}$	$65.07^{\pm 2.21}$
NCI1	$73.42^{\pm 1.15}$	$74.03^{\pm 0.68}$	$\mathbf{77.45}^{\pm 1.38}$	$67.53^{\pm 0.24}$	$68.44^{\pm 0.49}$	$68.81^{\pm 0.48}$	$69.23^{\pm 0.50}$	$67.52^{\pm 0.60}$
NCI109	$71.81^{\pm 0.78}$	$74.19^{\pm 0.53}$	$\mathbf{77.26}^{\pm 1.42}$	$67.26^{\pm 1.37}$	$67.69^{\pm 1.19}$	$68.14^{\pm 0.85}$	$68.77^{\pm 0.80}$	$67.20^{\pm 1.03}$

Table 1: *1-NN classification accuracy (%). Results are mean ± standard deviation over splits generated using 5 data seeds.* $\hat{D}_4 = \max_{t\le 4} D^{(t)}_{\mathrm{CCWL}}$.

B.1 1-Nearest Neighbor Classification

To show that the pseudometrics encode meaningful geometry, we test them on the 1-Nearest Neighbor Classification tasks. For a given query test sample, we calculate the distance between the query and all training samples, then assign the label of the nearest one to the query. We compare our distances with the WL distance in (20) on five datasets from TUDataset (45). As shown in Table 1, Convex Co-OT $D^{\text{cvx}}_{\text{Co-OT}}$ achieves the highest mean 1-NN classification accuracy on all five datasets. Overall, CCWL distances at early depths improve over the WL distance (20) on MUTAG and PROTEINS and achieve comparable performance on COX2, but underperform on NCI1 and NCI109. Although convex Co-OT and the full CCWL distance share the same zero set, convex Co-OT yields nearest-neighbor relationships that better align with the class labels than the early-depth CCWL distances evaluated here.

B.2 D_{CCWL} as a regularizer

In this experiment we ask whether the geometry induced by D_{CCWL} can help message-passing TNNs on classification tasks. We use three cellular message-passing backbones: CWN/CIN (14), CIN++ (29) and CCIN (19) and add a label-free regularization term to the supervised objective,

$$\mathcal{L} \;=\; \mathcal{L}_{\mathrm{CE}} \;+\; \lambda\, \mathcal{L}_{\mathrm{CCWL}}, \tag{14}$$

where λ is set to be 0.5 across experiments. Let X denote the input, and $z(X)$ denote the pooled graph representation the backbone feeds to its final linear layer, and let $u(X) = g_\theta\big(z(X)\big)$, where g_θ is a two-layer projection head trained jointly with the backbone whose output is L_2-normalised. Writing $u_i := u(X_i)$ and $D_{ij} := \widehat{D}_{15}(X_i, X_j)$ with $\widehat{D}_{15} = \max_{t\le 15} D^{(t)}$, the term is

$$\mathcal{L}_{\mathrm{CCWL}} = \frac{1}{|\mathcal{T}|} \sum_{(i_1,i_2,i_3)\in\mathcal{T}} \Big[\, \|u_{i_1} - u_{i_2}\| - \|u_{i_1} - u_{i_3}\| + \big(D_{i_1,i_3} - D_{i_1,i_2}\big) \Big]_+, \tag{15}$$

At each optimiser step a fresh set of $|\mathcal{T}| = 32$ triplets is drawn: i_1 uniformly over the training split, i_2 uniformly among the 10 nearest training graphs to i_1 under D, and i_3 uniformly among training graphs strictly farther from i_1 than i_2. The term vanishes exactly when the embedding separates i_3 from i_1 by at least as much as the distance does, $\|u_{i_1} - u_{i_3}\| - \|u_{i_1} - u_{i_2}\| \ge D_{i_1,i_3} - D_{i_1,i_2}$, and is positive otherwise; it therefore penalizes not only triplets whose embedded order contradicts $D_{i_1,i_2} < D_{i_1,i_3}$, but also correctly ordered triplets separated by less than the margin the distance itself supplies. Table 2 shows that the regularizer is beneficial in the majority of the configurations on the graph classification tasks.

B.3 Correlations with Co-OT and convex Co-OT

Here we provide the strong correlations between the D_{CCWL} and $D_{\text{Co-OT}}$.

Dataset	CCIN		CWN		CIN++	
	Base	Regularized (ours)	Base	Regularized (ours)	Base	Regularized (ours)
MUTAG	$94.44^{\pm 1.35}$	$\mathbf{95.80}^{\pm 1.06}$	$94.81^{\pm 1.39}$	$\mathbf{95.19}^{\pm 1.70}$	$95.19^{\pm 1.23}$	$\mathbf{95.19}^{\pm 0.91}$
PTC	$61.96^{\pm 1.63}$	$\mathbf{62.25}^{\pm 0.90}$	$61.96^{\pm 1.69}$	$\mathbf{63.04}^{\pm 1.81}$	$62.75^{\pm 1.96}$	$\mathbf{62.94}^{\pm 1.44}$
PROTEIN.	$\mathbf{75.05}^{\pm 0.53}$	$74.92^{\pm 0.59}$	$75.20^{\pm 0.59}$	$\mathbf{75.44}^{\pm 0.70}$	$74.74^{\pm 0.61}$	$\mathbf{75.17}^{\pm 0.74}$
NCI1	$81.97^{\pm 0.27}$	$\mathbf{83.08}^{\pm 0.67}$	$\mathbf{84.75}^{\pm 0.50}$	$84.54^{\pm 0.36}$	$\mathbf{85.02}^{\pm 0.48}$	$84.67^{\pm 0.37}$
NCI109	$80.42^{\pm 0.30}$	$\mathbf{80.61}^{\pm 0.33}$	$\mathbf{85.11}^{\pm 0.35}$	$84.75^{\pm 0.31}$	$84.89^{\pm 0.28}$	$\mathbf{85.28}^{\pm 0.34}$

Table 2: *Classification accuracy (%) over 3 seeds.* ***Base*** *denotes the original model training protocol;* ***Regularized*** *adds the* D_{CCWL} *ordering regularizer.*

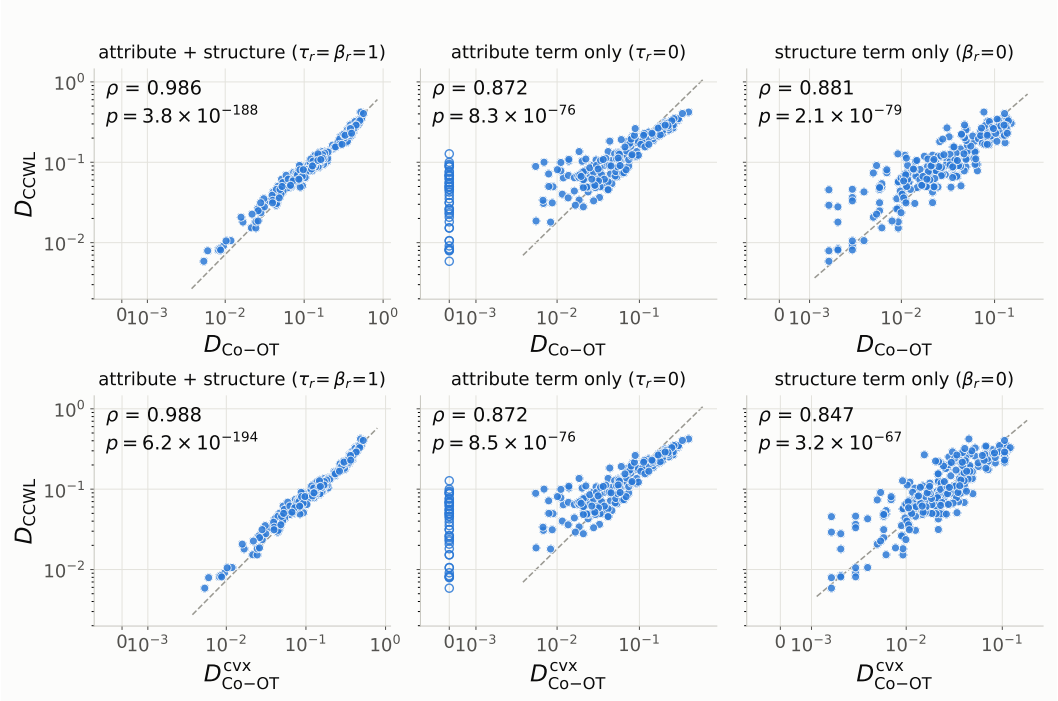


Figure 3: D_{CCWL} *against both* $D_{\mathrm{Co\text{-}OT}}$ *and* $D^{\mathrm{cvx}}_{\mathrm{Co\text{-}OT}}$ *on* 240 *pairs of attributed rank-2 complexes, for the full cost (left), the attribute channel alone (centre) and the incidence channel alone (right). Hollow markers are pairs with zero attribute cost.*

C Construction details of the CCWL distance

C.1 Incorporating Empty Channels

Starting from Z_0, define

$$E_{id,0} := Z_0, \qquad E_{a,0} := \begin{cases} Z_0 \sqcup \{\bot_a\}, & a \in \{B, C\}, \\ Z_0^2 \sqcup \{\bot_a\}, & a \in \{\downarrow, \uparrow\}, \end{cases} \qquad \mathsf{T}_0 := \bigsqcup_{a \in I} \big(\{a\} \times E_{a,0}\big).$$

where $\bigsqcup$ denotes the disjoint union and $\{a\}$ is the channel tag. Here, $E_{a,0}$ is the space of potential values carried by channel a and the identity channel id carries the cell's own attribute. The space T_0 collects the values from all channels, with each value tagged by its channel.

Then, recursively for $t \geq 1$, define

$$Z_t := \{\eta \in \mathcal{P}(\mathsf{T}_{t-1}) : \eta(\{a\} \times E_{a,t-1}) = \alpha_a \text{ for every } a \in I\},$$

$$E_{id,t} := Z_t, \qquad E_{a,t} := \begin{cases} Z_t \sqcup \{\perp_a\}, & a \in \{B, C\}, \\ Z_t^2 \sqcup \{\perp_a\}, & a \in \{\downarrow, \uparrow\}, \end{cases}$$

and

$$\mathsf{T}_t := \bigsqcup_{a \in I} \big(\{a\} \times E_{a,t}\big). \tag{16}$$

For each $t \geq 0$, write $\iota_a : E_{a,t} \to \mathsf{T}_t$, $\iota_a(u) := (a, u)$, for the canonical injection, leaving its dependence on t implicit. The null symbol $\perp_a$ is understood separately at each depth; we suppress its depth index for readability.

For every $a \neq id$, extend the channel ground metric by

$$\rho_{a,t}(\perp_a, \perp_a) := 0, \qquad \rho_{a,t}(\perp_a, u) = \rho_{a,t}(u, \perp_a) := M \quad \text{for } u \neq \perp_a.$$

for some $M > 0$ such that the initial attribute metric d_0 is bounded by M.

C.2 Neighborhood Channel Decomposition of d_t

Proposition C.1 (Channel decomposition)**.** *For $\eta, \eta' \in Z_{t+1}$, with a slight abuse of notation, rewrite η_a, η'_a as their normalized channel restrictions, defined by*

$$\eta_a(A) := \frac{\eta(\iota_a(A))}{\alpha_a}, \qquad \eta'_a(A) := \frac{\eta'(\iota_a(A))}{\alpha_a}$$

for Borel sets $A \subseteq E_{a,t}$. Then

$$W_{\rho_t}(\eta, \eta') = \sum_{a \in I} \alpha_a W_{\rho_{a,t}}(\eta_a, \eta'_a). \tag{17}$$

In particular, for $x \in X$ and $y \in Y$,

$$\begin{aligned} &d_{t+1}\big(q^{\mathcal{X}}_{t+1}(x), q^{\mathcal{Y}}_{t+1}(y)\big) \\ &\quad = \alpha_0 d_t\big(q^{\mathcal{X}}_t(x), q^{\mathcal{Y}}_t(y)\big) + \sum_{a \in I \setminus \{0\}} \alpha_a W_{\rho_{a,t}}\big(\nu^{\mathcal{X}}_{a,x,t}, \nu^{\mathcal{Y}}_{a,y,t}\big). \end{aligned} \tag{18}$$

Proof. For each $a \in I$, choose a coupling $\pi_a \in \Pi(\eta_a, \eta'_a)$. Then

$$\pi := \sum_{a \in I} \alpha_a (\iota_a \times \iota_a)_{\#} \pi_a$$

has marginals η and η'. It is supported on matching channel types and satisfies

$$\int_{\mathsf{T}_t \times \mathsf{T}_t} \rho_t \, d\pi = \sum_{a \in I} \alpha_a \int_{S_{a,t} \times S_{a,t}} \rho_{a,t} \, d\pi_a.$$

The measures involved have finite support, so each channel transport problem is a finite-dimensional minimization over a compact coupling polytope with a finite linear objective. Each therefore admits a minimizer. Choosing π_a to be a minimizer gives

$$W_{\rho_t}(\eta, \eta') \leq \sum_{a \in I} \alpha_a W_{\rho_{a,t}}(\eta_a, \eta'_a).$$

Conversely, let $\pi \in \Pi(\eta, \eta')$ have finite cost. The infinite cross-type cost implies that π gives zero mass to pairs with different channel tags. For each a, its restriction to $\iota_a(S_{a,t}) \times \iota_a(S_{a,t})$ has mass α_a. Removing the tags and dividing by α_a therefore gives a probability measure π_a on $S_{a,t} \times S_{a,t}$. For every Borel $A \subseteq S_{a,t}$,

$$\begin{aligned} \pi_a(A \times S_{a,t}) &= \frac{1}{\alpha_a} \pi\big(\iota_a(A) \times \iota_a(S_{a,t})\big) \\ &= \frac{1}{\alpha_a} \pi\big(\iota_a(A) \times \mathsf{T}_t\big) = \eta_a(A). \end{aligned}$$

The second marginal is η'_a by the same argument, so $\pi_a \in \Pi(\eta_a, \eta'_a)$. Consequently,

$$\int \rho_t \, d\pi = \sum_{a \in I} \alpha_a \int \rho_{a,t} \, d\pi_a \geq \sum_{a \in I} \alpha_a W_{\rho_{a,t}}(\eta_a, \eta'_a).$$

Infinite-cost couplings cannot improve this bound. Taking the infimum proves equation 17.

For the attributes, the normalized channel restrictions are precisely $\nu^{\mathcal{X}}_{a,x,t}$ and $\nu^{\mathcal{Y}}_{a,y,t}$. In the identity channel, the only coupling of the two Dirac measures has cost $d_t(q^{\mathcal{X}}_t(x), q^{\mathcal{Y}}_t(y))$. Substitution gives equation 18. □

C.3 Proof of Pseudometricity of D_{CCWL} (Proposition 3.2)

Proof. Let M and the base measures $\mu_{\mathcal{X}}$ be as specified in the proposition. We first prove by induction that d_t is a complete metric on Z_t, bounded by M, and that (Z_t, d_t) is separable.

The assertion holds for $t = 0$ by assumption. Suppose it holds at depth $t-1$. For clarity, write the channel spaces, including the null symbols, as

$$E_{id,t-1} := Z_{t-1}, \qquad E_{a,t-1} := \begin{cases} Z_{t-1} \sqcup \{\perp_a\}, & a \in \{B, C\}, \\ Z^2_{t-1} \sqcup \{\perp_a\}, & a \in \{\downarrow, \uparrow\}. \end{cases}$$

The non-null channel metrics are either d_{t-1} or the arithmetic mean of two copies of d_{t-1}. The latter is a metric because positivity, symmetry, and the triangle inequality hold coordinatewise. A sequence is Cauchy in this product metric if and only if each coordinate sequence is Cauchy. Thus, completeness follows from that of d_{t-1}. Separability follows by taking the Cartesian product of countable dense subsets. All non-null channel metrics are therefore complete, separable, and bounded by M.

Adjoining a null point at distance M preserves these properties. To verify the additional triangle inequalities, for non-null u, v observe that

$$\rho_{a,t-1}(u, v) \leq M \leq \rho_{a,t-1}(u, \perp_a) + \rho_{a,t-1}(\perp_a, v)$$

and

$$\rho_{a,t-1}(u, \perp_a) = M \leq \rho_{a,t-1}(u, v) + \rho_{a,t-1}(v, \perp_a).$$

The remaining cases follow by symmetry or involve repeated points. The extended space is separable, since a countable dense subset is obtained by adjoining the null point to one for the non-null space. Since $M > 0$, any Cauchy sequence is eventually either constantly null or entirely non-null; in either case, it converges. Hence every $(E_{a,t-1}, \rho_{a,t-1})$ is a Polish space with a complete metric bounded by M.

We now use the standard Wasserstein-space theorem: if (E, ρ) is complete and separable, then $(\mathcal{P}_1(E), W_\rho)$ is complete and separable, and W_ρ is a metric. Here, each ground metric is bounded, so every Borel probability measure has finite first moment. Consequently, $\mathcal{P}_1(E_{a,t-1}) = \mathcal{P}(E_{a,t-1})$, and each channel Wasserstein space is complete and separable. Moreover, its metric is bounded by M.

For $\eta \in Z_t$, let η_a denote its normalized channel restriction, characterized by

$$\eta_a(A) = \frac{1}{\alpha_a} \eta\big(\iota_{a,t}(A)\big)$$

for Borel $A \subseteq E_{a,t-1}$. By Proposition C.1,

$$d_t(\eta, \eta') = \sum_{a \in I} \alpha_a W_{\rho_{a,t-1}}(\eta_a, \eta'_a).$$

This expression is nonnegative and symmetric. Since $\alpha_a > 0$ for every a, it vanishes if and only if $\eta_a = \eta'_a$ for every channel. In that case,

$$\eta = \sum_{a \in I} \alpha_a (\iota_{a,t})_{\#} \eta_a = \sum_{a \in I} \alpha_a (\iota_{a,t})_{\#} \eta'_a = \eta'.$$

For any $\eta, \eta', \eta'' \in Z_t$,

$$\begin{aligned} d_t(\eta, \eta'') &= \sum_{a \in I} \alpha_a W_{\rho_{a,t-1}}(\eta_a, \eta''_a) \\ &\leq \sum_{a \in I} \alpha_a \left[W_{\rho_{a,t-1}}(\eta_a, \eta'_a) + W_{\rho_{a,t-1}}(\eta'_a, \eta''_a) \right] \\ &= d_t(\eta, \eta') + d_t(\eta', \eta''). \end{aligned}$$

Thus, d_t is a metric. Its bound follows from

$$d_t(\eta, \eta') \leq \sum_{a \in I} \alpha_a M = M.$$

The map

$$\eta \longmapsto (\eta_a)_{a \in I}$$

is a bijection from Z_t onto $\prod_{a \in I} \mathcal{P}(E_{a,t-1})$, with inverse

$$(\eta_a)_{a \in I} \longmapsto \sum_{a \in I} \alpha_a (\iota_{a,t})_{\#} \eta_a.$$

By the decomposition formula, it is an isometry when the product is equipped with the weighted sum of its channel Wasserstein metrics. This finite product is complete: a Cauchy sequence is Cauchy in every coordinate because each $\alpha_a > 0$, and its coordinatewise limits give convergence in the weighted sum metric. It is also separable, as the finite product of countable dense subsets is countable and dense. Therefore, (Z_t, d_t) is complete and separable. This completes the induction.

Next, for each $\mathcal{X} \in \mathfrak{C}$, finiteness and nonemptiness imply that $\mathcal{R}_{\mathcal{X}}$ is finite and nonempty. Since $w_r > 0$,

$$0 < s_{\mathcal{X}} = \sum_{r \in \mathcal{R}_{\mathcal{X}}} w_r < \infty.$$

Every denominator $|X_r|$ appearing in $\mu_{\mathcal{X}}$ is positive, and

$$\mu_{\mathcal{X}}(X) = \sum_{r \in \mathcal{R}_{\mathcal{X}}} \frac{w_r}{s_{\mathcal{X}}} = 1.$$

Thus $\mu_{\mathcal{X}}$ is a probability measure. The recursive attribute map $q_t^{\mathcal{X}}$ takes values in Z_t: each channel distribution is a probability measure on its corresponding channel space, and the tagged mixture assigns mass α_a to channel a. Since X is finite, this map is measurable, and hence

$$L_t(\mathcal{X}) = (q_t^{\mathcal{X}})_{\#} \mu_{\mathcal{X}} \in \mathcal{P}(Z_t).$$

No common rank support is needed.

Applying the Wasserstein-space theorem to (Z_t, d_t) shows that W_{d_t} is a metric on $\mathcal{P}(Z_t)$ bounded by M. Therefore, $D_{\mathrm{CCWL}}^{(t)}$ is nonnegative, symmetric, and zero on the diagonal. For any $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathfrak{C}$,

$$\begin{aligned} D_{\mathrm{CCWL}}^{(t)}(\mathcal{X}, \mathcal{Z}) &= W_{d_t}\big(L_t(\mathcal{X}), L_t(\mathcal{Z})\big) \\ &\leq W_{d_t}\big(L_t(\mathcal{X}), L_t(\mathcal{Y})\big) + W_{d_t}\big(L_t(\mathcal{Y}), L_t(\mathcal{Z})\big) \\ &= D_{\mathrm{CCWL}}^{(t)}(\mathcal{X}, \mathcal{Y}) + D_{\mathrm{CCWL}}^{(t)}(\mathcal{Y}, \mathcal{Z}). \end{aligned}$$

Consequently, each $D_{\mathrm{CCWL}}^{(t)}$ is a pseudometric bounded by M.

Finally,

$$0 \leq D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) = \sup_{t \geq 0} D_{\mathrm{CCWL}}^{(t)}(\mathcal{X}, \mathcal{Y}) \leq M.$$

Symmetry and vanishing on the diagonal follow from the corresponding properties at every depth. For every $t \geq 0$,

$$\begin{aligned} D_{\mathrm{CCWL}}^{(t)}(\mathcal{X}, \mathcal{Z}) &\leq D_{\mathrm{CCWL}}^{(t)}(\mathcal{X}, \mathcal{Y}) + D_{\mathrm{CCWL}}^{(t)}(\mathcal{Y}, \mathcal{Z}) \\ &\leq D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) + D_{\mathrm{CCWL}}(\mathcal{Y}, \mathcal{Z}). \end{aligned}$$

Taking the supremum over t on the left proves the triangle inequality. Hence $D_{\mathrm{CCWL}} : \mathfrak{C} \times \mathfrak{C} \to [0, M]$ is a pseudometric. □

C.4 Proof of Equivalence with CCWL distinguishability (Proposition 3.3)

This subsection proves Proposition 3.3. Throughout, we use the initial attributes in equation 6, the base measures in equation 2, and the global attribute distributions in equation 3. Channel weights $\alpha_a > 0$ and rank weights $w_r > 0$ are shared across all complexes. Empty channels are treated according to Appendix C.1.

Original and augmented CCWL refinement. Let $\chi_t^{\mathcal{X}}(x)$ denote the original discrete CCWL color, initialized by

$$\chi_0^{\mathcal{X}}(x) := \big(c_{\mathcal{X}}(x), \mathrm{rk}_{\mathcal{X}}(x)\big).$$

Let $\widetilde{\chi}_t^{\mathcal{X}}(x)$ denote the color obtained by the same refinement rule, but initialized by

$$\widetilde{\chi}_0^{\mathcal{X}}(x) := q_0^{\mathcal{X}}(x).$$

Both procedures use injective encodings shared across all complexes being compared. We refer to the latter procedure as augmented CCWL.

For any cell map $h : X \to U$, define its four context multisets by

$$\mathcal{M}_a^{\mathcal{X}}(h, x) := \begin{cases} \{\{h(y) : y \in \mathcal{N}_a(x)\}\}, & a \in \{B, C\}, \\ \{\{(h(y), h(z)) : (y, z) \in \mathcal{A}_a(x)\}\}, & a \in \{\downarrow, \uparrow\}. \end{cases}$$

Thus, both discrete procedures update a cell's color by injectively encoding its current color and these four context multisets.

Write

$$m_a^{\mathcal{X}}(x) := \begin{cases} |\mathcal{N}_a(x)|, & a \in \{B, C\}, \\ |\mathcal{A}_a(x)|, & a \in \{\downarrow, \uparrow\}, \end{cases}$$

for the corresponding entries of $\boldsymbol{m}_{\mathcal{X}}(x)$.

Retention of initial attributes. For a realized attribute $q_{t+1}^{\mathcal{X}}(x) \in Z_{t+1}$, its normalized identity-channel restriction is $\delta_{q_t^{\mathcal{X}}(x)}$. Because $\alpha_{id} > 0$, this restriction is uniquely determined by $q_{t+1}^{\mathcal{X}}(x)$, and its unique support point recovers $q_t^{\mathcal{X}}(x)$. Iterating this observation recovers $q_0^{\mathcal{X}}(x)$ from every $q_t^{\mathcal{X}}(x)$.

In particular, every realized recursive attribute determines the cell's original attribute, rank, neighborhood-count vector, and global rank-count sequence. These recovery statements concern attributes realized by cells; they need not hold for arbitrary elements of the ambient space Z_t.

Lemma C.2 (Multiset recovery)**.** *For every $t \geq 0$ and $x \in X$, $q_{t+1}^{\mathcal{X}}(x)$ determines $q_t^{\mathcal{X}}(x)$ and all four multisets*

$$\mathcal{M}_a^{\mathcal{X}}(q_t^{\mathcal{X}}, x), \qquad a \in \{B, C, \downarrow, \uparrow\}.$$

Proof. The disjoint channel tags and positive channel weights allow recovery of each normalized channel restriction $\nu_{a,x,t}^{\mathcal{X}}$ from $q_{t+1}^{\mathcal{X}}(x)$. The identity restriction recovers $q_t^{\mathcal{X}}(x)$, as explained above, and hence also the counts $m_a^{\mathcal{X}}(x)$.

For a nonempty unary channel $a \in \{B, C\}$ and $z \in Z_t$,

$$\nu_{a,x,t}^{\mathcal{X}}(\{z\}) = \frac{\big|\{y \in \mathcal{N}_a(x) : q_t^{\mathcal{X}}(y) = z\}\big|}{m_a^{\mathcal{X}}(x)}.$$

Multiplying by the recovered positive integer $m_a^{\mathcal{X}}(x)$ gives the exact multiplicity of z in $\mathcal{M}_a^{\mathcal{X}}(q_t^{\mathcal{X}}, x)$.

For a nonempty pair channel $a \in \{\downarrow, \uparrow\}$ and $(z_1, z_2) \in Z_t^2$,

$$\nu_{a,x,t}^{\mathcal{X}}(\{(z_1, z_2)\}) = \frac{\big|\{(y, z) \in \mathcal{A}_a(x) : q_t^{\mathcal{X}}(y) = z_1,\ q_t^{\mathcal{X}}(z) = z_2\}\big|}{m_a^{\mathcal{X}}(x)}.$$

Again, multiplying by the recovered channel cardinality gives the exact multiplicity.

If a channel is empty, its count is zero and its distribution is the designated null Dirac measure. Its context multiset is therefore the empty multiset. These cases cover all four channels. □

Lemma C.3 (Cellwise correspondence)**.** *For every $t \geq 0$, any two finite nonempty complexes $\mathcal{X}, \mathcal{Y}$, and any $x \in X$, $y \in Y$,*

$$q_t^{\mathcal{X}}(x) = q_t^{\mathcal{Y}}(y) \quad \Longleftrightarrow \quad \widetilde{\chi}_t^{\mathcal{X}}(x) = \widetilde{\chi}_t^{\mathcal{Y}}(y).$$

Proof. We proceed by induction on t. At depth zero, both constructions use the same augmented tuples, so the assertion is immediate.

Suppose the assertion holds at depth t. It follows that, on the realized values at this depth, there is a well-defined bijection between recursive attributes and augmented colors. Indeed, associate $q_t^{\mathcal{X}}(x)$ with $\widetilde{\chi}_t^{\mathcal{X}}(x)$. The forward implication of the induction hypothesis makes this association well-defined, and the reverse implication makes it injective. It is surjective onto the realized augmented colors by construction.

First assume $q_{t+1}^{\mathcal{X}}(x) = q_{t+1}^{\mathcal{Y}}(y)$. By Lemma C.2, the current attributes agree, and the four context multisets of depth-t attributes agree. Applying the bijection from the preceding paragraph to each unary entry and to both coordinates of each pair gives equality of the current augmented colors and all four augmented-color context multisets. The discrete update therefore gives

$$\widetilde{\chi}_{t+1}^{\mathcal{X}}(x) = \widetilde{\chi}_{t+1}^{\mathcal{Y}}(y).$$

Conversely, assume these depth-$(t+1)$ augmented colors agree. Injectivity of the discrete update implies equality of their current augmented colors and of all four context multisets. Applying the inverse bijection gives equality of the current recursive attributes and the corresponding attribute multisets. Equal nonempty multisets have equal cardinalities and equal normalized probability measures. Equal empty multisets give the same null measure. Hence all four neighborhood distributions agree, as do the identity-channel Dirac measures. Their tagged mixtures, formed using the same weights, are equal:

$$q_{t+1}^{\mathcal{X}}(x) = q_{t+1}^{\mathcal{Y}}(y).$$

This completes the induction. □

Lemma C.4 (Global multiset recovery)**.** *For every $t \geq 0$,*

$$L_t(\mathcal{X}) = L_t(\mathcal{Y}) \quad \Longleftrightarrow \quad \left\{\!\left\{ \widetilde{\chi}_t^{\mathcal{X}}(x) : x \in X \right\}\!\right\} = \left\{\!\left\{ \widetilde{\chi}_t^{\mathcal{Y}}(y) : y \in Y \right\}\!\right\}.$$

Proof. Every realized recursive attribute retains the global rank-count sequence of its complex. Consequently, if $\boldsymbol{n}(\mathcal{X}) \neq \boldsymbol{n}(\mathcal{Y})$, no realized recursive attribute can occur in both complexes. Their global attribute distributions, which are finitely supported probability measures, then have disjoint supports and cannot be equal. By Lemma C.3, their augmented-color supports are also disjoint. Since both complexes are nonempty, their global augmented-color multisets cannot agree either. Thus both sides of the claimed equivalence are false in this case.

It remains to consider $\boldsymbol{n}(\mathcal{X}) = \boldsymbol{n}(\mathcal{Y})$. Then

$$\mathcal{R}_{\mathcal{X}} = \mathcal{R}_{\mathcal{Y}}, \qquad |X_r| = |Y_r|, \qquad s_{\mathcal{X}} = s_{\mathcal{Y}}.$$

Every realized recursive attribute determines its rank. Hence, for a realized value z of rank r,

$$L_t(\mathcal{X})(\{z\}) = \frac{w_r}{s_{\mathcal{X}}|X_r|}\left|\{x \in X_r : q_t^{\mathcal{X}}(x) = z\}\right|.$$

The analogous formula for $\mathcal{Y}$ has the same strictly positive coefficient. Therefore, equality of the two global distributions is equivalent to equality of the multiplicity of every recursive attribute. This follows by comparing singleton masses on the finite union of their supports.

Finally, Lemma C.3 identifies realized recursive attributes bijectively with realized augmented colors. Equality of their multiplicities is therefore equivalent to equality of the global augmented-color multisets. □

Lemma C.5 (Augmentation preserves stable equivalence)**.** *The original and augmented CCWL procedures have the same stable global indistinguishability relation. More precisely,*

$$\begin{aligned}&\{\{\chi_t^{\mathcal{X}}(x) : x \in X\}\} = \{\{\chi_t^{\mathcal{Y}}(y) : y \in Y\}\} \quad \text{for every } t \geq 0\\ &\qquad\Longleftrightarrow\\ &\{\{\widetilde{\chi}_t^{\mathcal{X}}(x) : x \in X\}\} = \{\{\widetilde{\chi}_t^{\mathcal{Y}}(y) : y \in Y\}\} \quad \text{for every } t \geq 0.\end{aligned}$$

Proof. We first show that an augmented color at depth t determines the original color at the same depth. At $t = 0$, discard the neighborhood counts and global rank-count sequence from the augmented tuple, retaining only the original attribute and rank. Suppose such a deterministic map exists at depth t. An augmented color at depth $t + 1$ determines its previous color and four context multisets, because its update is injective. Apply the depth-t map to the previous color, to each unary multiset entry, and to both coordinates of every pair entry. This recovers the complete input of the original depth-$(t + 1)$ update, and therefore the original color. By induction, deterministic maps F_t exist such that

$$\chi_t^{\mathcal{X}}(x) = F_t\big(\widetilde{\chi}_t^{\mathcal{X}}(x)\big)$$

for every complex and cell. Applying F_t to equal global augmented-color multisets proves that augmented indistinguishability implies original indistinguishability.

For the converse, fix a rank-count sequence $\boldsymbol{n}$ and restrict attention to complexes with this sequence. We show inductively that the original color at depth $t + 1$ determines the augmented color at depth t. Equivalently, there is a deterministic map $G_{t,\boldsymbol{n}}$ on realized original colors such that

$$\widetilde{\chi}_t^{\mathcal{X}}(x) = G_{t,\boldsymbol{n}}\big(\chi_{t+1}^{\mathcal{X}}(x)\big).$$

For $t = 0$, the original color $\chi_1^{\mathcal{X}}(x)$ determines $\chi_0^{\mathcal{X}}(x)$ and all four depth-zero context multisets. The first gives the original attribute and rank, and the cardinalities of the latter give $\boldsymbol{m}_{\mathcal{X}}(x)$. Together with the fixed sequence $\boldsymbol{n}$, these determine $\widetilde{\chi}_0^{\mathcal{X}}(x)$.

Suppose $G_{t,\boldsymbol{n}}$ has been constructed. The color $\chi_{t+2}^{\mathcal{X}}(x)$ determines $\chi_{t+1}^{\mathcal{X}}(x)$ and its four context multisets of depth-$(t + 1)$ original colors. Apply $G_{t,\boldsymbol{n}}$ to the current color and entrywise to these multisets, acting on both coordinates in each pair channel. The result is precisely the current augmented color at depth t and its four augmented context multisets. These determine $\widetilde{\chi}_{t+1}^{\mathcal{X}}(x)$ through the augmented update. This constructs $G_{t+1,\boldsymbol{n}}$ and completes the induction.

Now suppose that $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable by the original CCWL procedure at every depth. Equality of their depth-zero global color multisets implies equality of their rank histograms, so $\boldsymbol{n}(\mathcal{X}) = \boldsymbol{n}(\mathcal{Y}) = \boldsymbol{n}$. For every $t \geq 0$, apply the same map $G_{t,\boldsymbol{n}}$ to their equal global original-color multisets at depth $t + 1$. The resulting multisets are their global augmented-color multisets at depth t, which are therefore equal. This proves the converse. □

Proof of Proposition 3.3. By the metric construction established in the proof of Proposition 3.2, d_t is a metric on Z_t and W_{d_t} is a metric on $\mathcal{P}(Z_t)$. Hence, for every $t \geq 0$,

$$\begin{aligned}D_{\mathrm{CCWL}}^{(t)}(\mathcal{X}, \mathcal{Y}) = 0 &\iff L_t(\mathcal{X}) = L_t(\mathcal{Y})\\ &\iff \{\{\widetilde{\chi}_t^{\mathcal{X}}(x) : x \in X\}\} = \{\{\widetilde{\chi}_t^{\mathcal{Y}}(y) : y \in Y\}\},\end{aligned}$$

where the second equivalence is Lemma C.4.

All depth distances are nonnegative, so

$$D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) = 0 \iff D_{\mathrm{CCWL}}^{(t)}(\mathcal{X}, \mathcal{Y}) = 0 \quad \text{for every } t \geq 0.$$

Combining the preceding equivalences shows that the absolute distance vanishes exactly when the augmented global color multisets agree at every depth. By Lemma C.5, this is equivalent to $\mathcal{X} \equiv_{\mathrm{CCWL}} \mathcal{Y}$.

Since D_{CCWL} is nonnegative and finite, its failure to vanish is equivalent to strict positivity. This gives the stated characterization of CCWL distinguishability. □

Remark C.6 (Refinement depth). The correspondence with augmented CCWL holds at each depth. The correspondence with the original CCWL procedure is a statement about indistinguishability across all depths, not necessarily about matching the first distinguishing depth.

In particular, the proof of Lemma C.5 shows that augmented depth-t colors determine original depth-t colors. Conversely, within a fixed rank-count class, original depth-$(t+1)$ colors determine augmented depth-t colors. The global rank counts are already determined by the original global color multiset at depth zero.

Remark C.7 (Choice of weights). Under the initial-attribute convention equation 6, any fixed strictly positive channel weights summing to one and any fixed positive rank weights give the same zero set for D_{CCWL}. Indeed, Proposition 3.3 identifies that zero set with CCWL equivalence, independently of these choices.

The same conclusion holds for the zero set at each fixed depth relative to augmented CCWL, by Lemmas C.3 and C.4. Numerical distances may nevertheless change.

Strict positivity ensures that every channel can be recovered from the tagged representation and that every cell has positive reference mass. Allowing zero weights would require a separate analysis of the information discarded.

Remark C.8 (A common initial attribute space). The finite-support rank-count sequences form the countable set

$$\mathcal{N} := \Big\{ (n_r)_{r\geq 0} \in \mathbb{N}_0^{\mathbb{N}_0} : n_r = 0 \text{ for all but finitely many } r \Big\}.$$

Thus, if the original attributes take values in a Polish space C, the augmented tuples can be placed in the common space

$$Z_0 := C \times \mathbb{N}_0 \times \mathbb{N}_0^4 \times \mathcal{N}.$$

Give each countable factor the discrete topology. If d_C is a compatible complete metric on C, then, for any $M > 0$,

$$d_0\big((c,r,\boldsymbol{m},\boldsymbol{n}),(c',r',\boldsymbol{m}',\boldsymbol{n}')\big) := \frac{M}{4}\Big[\min\{1, d_C(c,c')\} + \mathbf{1}_{\{r\neq r'\}} + \mathbf{1}_{\{\boldsymbol{m}\neq\boldsymbol{m}'\}} + \mathbf{1}_{\{\boldsymbol{n}\neq\boldsymbol{n}'\}}\Big]$$

is a bounded complete metric making Z_0 Polish. This is one admissible choice; the distinguishability result requires only that d_0 be a metric separating distinct augmented tuples under the stated assumptions.

D Co-OT: formulation, relaxation, and CCWL equivalence

This appendix provides additional properties of Co-Optimal Transport formulations and proves Proposition 4.2. The attribute maps are the augmented maps in equation 6; the channel weights used to construct D_{CCWL} need not coincide with the weights β_r, τ_r used here. All these weights are fixed independently of the compared complexes and strictly positive. We use raw binary adjacent-rank incidence matrices, not degree-normalized incidence matrices.

D.1 Rank measures and missing ranks

To define the distances on all finite nonempty complexes without pair-dependent changes to their representations, complete each rank separately. For every $r \geq 0$, set

$$\widehat{X}_r := \begin{cases} X_r, & X_r \neq \varnothing, \\ \{\star_r\}, & X_r = \varnothing, \end{cases} \qquad n_r^{\mathcal{X}} := |\widehat{X}_r|, \qquad u_r^{\mathcal{X}} := \frac{1}{n_r^{\mathcal{X}}} \sum_{x\in\widehat{X}_r} \delta_x.$$

The symbol $\star_r$ is a null cell used only for this transport representation. It is not inserted into the original complex or into CCWL refinement. In particular, the augmented sequence $\boldsymbol{n}(\mathcal{X}) = (|X_r|)_{r\geq 0}$ always counts genuine cells; its entries remain zero at absent ranks.

Choose $M > 0$ with $d_0 \le M$. Adjoin a symbol † to the augmented attribute space Z_0 and define the metric

$$\overline{d}_0(a,b) = d_0(a,b) \quad (a,b \in Z_0), \qquad \overline{d}_0(\dagger,a) = \overline{d}_0(a,\dagger) = M, \qquad \overline{d}_0(\dagger,\dagger) = 0.$$

This is a metric: the only new triangle inequalities are $d_0(a,b) \le 2M$ and $M \le M + d_0(a,b)$. Use $\widehat{a}_{\mathcal{X}}(x) = q_0^{\mathcal{X}}(x)$ for real cells and $\widehat{a}_{\mathcal{X}}(\star_r) = \dagger$. Define $B_r^{\mathcal{X}}$ on $\widehat{X}_{r-1} \times \widehat{X}_r$ to be the original binary incidence on real cells and zero whenever either argument is null.

The rank transport polytope is

$$\Pi_r(\mathcal{X},\mathcal{Y}) := \left\{ \pi \ge 0 : \pi\mathbf{1} = \frac{1}{n_r^{\mathcal{X}}}\mathbf{1}, \pi^\top\mathbf{1} = \frac{1}{n_r^{\mathcal{Y}}}\mathbf{1} \right\}.$$

Its matrices may be rectangular. Set

$$C_r(x,y) := \overline{d}_0(\widehat{a}_{\mathcal{X}}(x), \widehat{a}_{\mathcal{Y}}(y)).$$

These definitions specify the null conventions suppressed in the main text. Each complex is represented independently of its comparison partner. In particular, the rank measures used here are uniform within ranks; the global rank weights defining $\mu_{\mathcal{X}}$ for the CCWL distance are not transport marginals in the Co-OT objectives.

If K is the larger of the maximal occupied ranks of two complexes, every rank above K consists of a null singleton on both sides. Its coupling is uniquely [1], its attribute cost is zero, and all incidence costs with index greater than K are zero. Thus equation 8 is a finite optimization over ranks $0, \dots, K$. No summability assumption on the positive weights is needed for pairwise finiteness. The same fixed sequences of weights must nevertheless be used for all pairs; no pair-dependent renormalization is made.

D.2 Convexity, attainment, and the relaxation inequality

Lemma D.1 (Attainment and convexity)**.** *Both minima in equation 8 are attained. The convex Co-OT objective is jointly convex in its coupling family and admits a linear-programming formulation.*

Proof. After truncation as above, the feasible set is a finite product of nonempty compact transport polytopes. Both objectives are continuous: Co-OT has linear attribute terms and bilinear incidence terms, whereas convex Co-OT has linear attribute terms and norms of linear residual maps. Thus both attain their minima. The latter objective is jointly convex.

For its linear-programming formulation, introduce matrices U_r, V_r with the same dimensions as E_r, F_r, impose

$$U_r \ge E_r, \quad U_r \ge -E_r, \qquad V_r \ge F_r, \quad V_r \ge -F_r,$$

entrywise, and replace $\mathcal{R}_r$ by $\frac{1}{2}(\sum_{i,j}(U_r)_{ij} + \sum_{i,j}(V_r)_{ij})$. These inequalities imply nonnegativity of the auxiliary entries. For fixed couplings their minimizing values are $|E_r|, |F_r|$, since the objective coefficients are positive. Hence the linear program has exactly the convex Co-OT optimum. □

Lemma D.2 (Pointwise relaxation)**.** *For every feasible coupling family and every adjacent-rank index r, $\mathcal{R}_r(\boldsymbol{\pi}) \le \mathcal{I}_r(\boldsymbol{\pi})$. If corresponding rank sizes agree and both couplings are induced by bijections, equality holds.*

Proof. Fix r and write

$$m = n_{r-1}^{\mathcal{X}}, \quad n = n_r^{\mathcal{X}}, \quad p = n_{r-1}^{\mathcal{Y}}, \quad q = n_r^{\mathcal{Y}}, \qquad A = B_r^{\mathcal{X}}, \quad B = B_r^{\mathcal{Y}}, \quad P = \pi_{r-1}, \quad Q = \pi_r.$$

Thus $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$, $P \in \mathbb{R}^{m\times p}$, and $Q \in \mathbb{R}^{n\times q}$, and

$$\mathcal{I}_r = \sum_{i,j,k,\ell} |A_{ij} - B_{k\ell}| P_{ik} Q_{j\ell}.$$

The uniform marginal constraints imply

$$(E_r)_{i\ell} = \left(\frac{AQ}{m} - \frac{PB}{q}\right)_{i\ell} = \sum_{j,k} P_{ik} Q_{j\ell} (A_{ij} - B_{k\ell}),$$

$$(F_r)_{jk} = \left(\frac{A^\top P}{n} - \frac{QB^\top}{p}\right)_{jk} = \sum_{i,\ell} P_{ik} Q_{j\ell} (A_{ij} - B_{k\ell}).$$

For example, the first identity uses $\sum_k P_{ik} = 1/m$ and $\sum_j Q_{j\ell} = 1/q$. Taking absolute values, applying the triangle inequality, and summing gives

$$\|E_r\|_{\mathrm{ent},1} \le \mathcal{I}_r, \qquad \|F_r\|_{\mathrm{ent},1} \le \mathcal{I}_r.$$

Averaging proves the claimed inequality.

If $m = p$, $n = q$, and f, g are rank-wise bijections, their couplings are $P_{ik} = \mathbf{1}[k = f(i)]/m$ and $Q_{j\ell} = \mathbf{1}[\ell = g(j)]/n$. Every nonzero average above then contains a single term. Both residual norms equal

$$\frac{1}{mn} \sum_{i,j} |A_{ij} - B_{f(i),g(j)}| = \mathcal{I}_r.$$

This proves equality in the bijection case. □

Summing Lemma D.2 with the fixed positive weights, leaving the attribute terms unchanged, and minimizing proves the relaxation inequality in Proposition D.3.

Relation to the equal-size formulation. When $n_r^{\mathcal{X}} = n_r^{\mathcal{Y}} = n_r$, the variables $S_r = n_r \pi_r$ are doubly stochastic. Substitution into equation 7 gives

$$E_r = \frac{B_r^{\mathcal{X}} S_r - S_{r-1} B_r^{\mathcal{Y}}}{n_{r-1} n_r}, \qquad F_r = \frac{(B_r^{\mathcal{X}})^\top S_{r-1} - S_r (B_r^{\mathcal{Y}})^\top}{n_{r-1} n_r}.$$

Thus the rectangular formulation specializes exactly to the two-sided penalty divided by $2n_{r-1}n_r$. Equal-size doubly stochastic matrices are not used when sizes differ.

D.3 Pseudometricity

Proposition D.3 (Pseudometricity and convex relaxation)**.** *Let $\mathfrak{C}$ be the class of finite nonempty attributed combinatorial complexes equipped with initial attribute maps into a common bounded metric space (Z_0, d_0). Use uniform per-rank probability measures, fixed positive weights β_r and τ_r, and the missing-rank conventions of Appendix D.1. Then the distances in equation 8 satisfy the following properties:*

1. *Both minima are attained, and $D_{\text{Co-OT}}$ and $D^{\text{cvx}}_{\text{Co-OT}}$ are finite pseudometrics on $\mathfrak{C}$: they are nonnegative, symmetric, vanish on the diagonal, and satisfy the triangle inequality.*
2. *The convex Co-OT objective is jointly convex in the rank-wise coupling family and admits an equivalent linear-programming formulation.*
3. *For every feasible coupling family $\boldsymbol{\pi}$ and every $r \ge 1$,*

   $$\mathcal{R}_r(\boldsymbol{\pi}) \le \mathcal{I}_r(\boldsymbol{\pi}).$$

   Consequently,

   $$D^{\text{cvx}}_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) \le D_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) \qquad \text{for all } \mathcal{X}, \mathcal{Y} \in \mathfrak{C}.$$

4. *When corresponding rank sizes agree, the two objectives coincide on every coupling family induced by rank-wise bijections.*

No equality of rank sizes or rank supports is required for the first three assertions.

Proof. Attainment was proved in Lemma D.1. Both objectives are nonnegative. For self-comparison choose $\pi_r = I/n_r^{\mathcal{X}}$. The attribute costs vanish. The Co-OT incidence mismatches vanish, and $B_r\pi_r/n_{r-1}^{\mathcal{X}} = \pi_{r-1}B_r/n_r^{\mathcal{X}}$, with the analogous transposed identity, so both convex residuals vanish. Thus both distances have zero self-distance.

Transposing each coupling proves symmetry for Co-OT. For convex Co-OT, swapping the complexes and transposing the couplings sends E_r to $-F_r^\top$ and F_r to $-E_r^\top$. The sum of the two entrywise norms and the attribute costs are unchanged. This proves symmetry.

For the triangle inequalities, let π_r couple $\mathcal{X}$ to $\mathcal{Y}$ and σ_r couple $\mathcal{Y}$ to $\mathcal{Z}$. Work up to the largest occupied rank of the three complexes, padding as in Appendix D.1. Write $b_r = n_r^{\mathcal{Y}}$ and define

$$\gamma_r(x, y, z) := b_r\pi_r(x, y)\sigma_r(y, z), \qquad T_r := b_r\pi_r\sigma_r.$$

Summing γ_r over z gives π_r, since the row sums of σ_r are $1/b_r$. Summing over x gives σ_r, since the column sums of π_r are $1/b_r$. Its (x, z) marginal is T_r, which is therefore a feasible coupling between $\mathcal{X}$ and $\mathcal{Z}$.

The triangle inequality for $\overline{d}_0$, integrated against γ_r, gives

$$\langle C_r^{\mathcal{XZ}}, T_r\rangle \le \langle C_r^{\mathcal{XY}}, \pi_r\rangle + \langle C_r^{\mathcal{YZ}}, \sigma_r\rangle.$$

For Co-OT, apply the scalar inequality

$$|B_r^{\mathcal{X}}(x_-, x) - B_r^{\mathcal{Z}}(z_-, z)| \le |B_r^{\mathcal{X}}(x_-, x) - B_r^{\mathcal{Y}}(y_-, y)| + |B_r^{\mathcal{Y}}(y_-, y) - B_r^{\mathcal{Z}}(z_-, z)|$$

and integrate against $\gamma_{r-1} \otimes \gamma_r$. Its relevant marginals give

$$\mathcal{I}_r^{\mathcal{XZ}}(\boldsymbol{T}) \le \mathcal{I}_r^{\mathcal{XY}}(\boldsymbol{\pi}) + \mathcal{I}_r^{\mathcal{YZ}}(\boldsymbol{\sigma}).$$

For convex Co-OT, direct expansion of the residuals gives

$$E_r^{\mathcal{XZ}}(\boldsymbol{T}) = b_r E_r^{\mathcal{XY}}(\boldsymbol{\pi})\sigma_r + b_{r-1}\pi_{r-1}E_r^{\mathcal{YZ}}(\boldsymbol{\sigma}),$$
$$F_r^{\mathcal{XZ}}(\boldsymbol{T}) = b_{r-1}F_r^{\mathcal{XY}}(\boldsymbol{\pi})\sigma_{r-1} + b_r\pi_r F_r^{\mathcal{YZ}}(\boldsymbol{\sigma}).$$

For example, the two middle terms in the first identity are $-\pi_{r-1}B_r^{\mathcal{Y}}\sigma_r$ and $+\pi_{r-1}B_r^{\mathcal{Y}}\sigma_r$, which cancel. For any compatible real matrix H, nonnegativity and the marginals imply

$$\|b_r H\sigma_r\|_{\mathrm{ent},1} \le b_r \sum_{i,j,k} |H_{ij}|\sigma_r(j, k) = \|H\|_{\mathrm{ent},1},$$

and

$$\|b_r\pi_r H\|_{\mathrm{ent},1} \le b_r \sum_{i,j,k} \pi_r(i, j)|H_{jk}| = \|H\|_{\mathrm{ent},1}.$$

The first uses the row sums of σ_r and the second the column sums of π_r. Applying these contractions to the residual identities proves

$$\mathcal{R}_r^{\mathcal{XZ}}(\boldsymbol{T}) \le \mathcal{R}_r^{\mathcal{XY}}(\boldsymbol{\pi}) + \mathcal{R}_r^{\mathcal{YZ}}(\boldsymbol{\sigma}).$$

For either objective, sum the attribute and incidence inequalities with their fixed weights. Choose minimizing families for the two input pairs and minimize over couplings of the outer pair. The resulting inequality is the triangle inequality for the corresponding distance. All sums are finite after the common truncation, so no exchange of infinite sums or limits is needed. □

D.4 The zero set and CCWL equivalence

Lemma D.4 (Zero cost forces equal rank counts)**.** *If a feasible family has zero total attribute cost, then $\boldsymbol{n}(\mathcal{X}) = \boldsymbol{n}(\mathcal{Y})$. In particular, this holds when either Co-OT distance vanishes.*

Proof. Suppose the two sequences differ. Choose an occupied rank r of $\mathcal{X}$, which exists because the complex is nonempty. Every cell of $\widehat{X}_r = X_r$ is real and its augmented attribute contains $\boldsymbol{n}(\mathcal{X})$. If Y_r is nonempty, all of its attributes contain the different sequence $\boldsymbol{n}(\mathcal{Y})$, so every entry of C_r is strictly positive because d_0 is a metric. If Y_r is empty, every entry is the null cost $M > 0$. In either case the finite matrix C_r has a strictly positive minimum entry. Every feasible π_r has total mass one, so $\beta_r\langle C_r, \pi_r\rangle > 0$. This contradicts zero attribute cost. For the final assertion, use attainment, nonnegativity of all terms, and positivity of the weights. □

Lemma D.5 (Support rigidity)**.** *Let S be doubly stochastic and let u, v satisfy $u = Sv$ and $v = S^\top u$. Then $S(x, y) > 0$ implies $u(x) = v(y)$.*

Proof. The identities give $u^\top Sv = \|u\|_2^2 = \|v\|_2^2$. Using both marginal constraints of S,

$$\sum_{x,y} S(x,y)(u(x) - v(y))^2 = \|u\|_2^2 + \|v\|_2^2 - 2u^\top Sv = 0.$$

Each summand is nonnegative, proving the assertion. □

Proof of Proposition 4.2. We first compare convex Co-OT with CCWL initialized by the augmented attributes $a_{\mathcal{X}} = q_0^{\mathcal{X}}$. Write $\widetilde{\chi}_t$ for these discrete colors. The original and augmented procedures have the same stable global equivalence relation by Lemma C.5.

Zero convex Co-OT implies CCWL equivalence. Suppose the convex distance is zero. A minimizing coupling family exists. Lemma D.4 implies equality of the genuine rank sizes and supports. Discard ranks absent in both complexes; their incidences and costs are zero. At each occupied rank, write $n_r = |X_r| = |Y_r|$ and $S_r = n_r \pi_r$. Then S_r is doubly stochastic. Vanishing of each residual gives

$$B_r^{\mathcal{X}} S_r = S_{r-1} B_r^{\mathcal{Y}}, \qquad (B_r^{\mathcal{X}})^\top S_{r-1} = S_r (B_r^{\mathcal{Y}})^\top \tag{19}$$

whenever both adjacent ranks are occupied. If an adjacent rank is absent, the corresponding genuine incidence block is empty and imposes no constraint. Zero attribute cost also gives

$$S_r(x, y) > 0 \implies a_{\mathcal{X}}(x) = a_{\mathcal{Y}}(y). \tag{20}$$

Let $A_{\mathcal{X}}$ be the symmetric adjacency matrix of the adjacent-rank incidence graph on the genuine cell set X. Its $(r-1, r)$ block is $B_r^{\mathcal{X}}$, its reverse block is $(B_r^{\mathcal{X}})^\top$, and all other blocks are zero. Define $A_{\mathcal{Y}}$ analogously, and let S be the block diagonal matrix with blocks S_r on the occupied ranks. It is doubly stochastic, and equation 19 gives

$$A_{\mathcal{X}} S = S A_{\mathcal{Y}}, \qquad A_{\mathcal{Y}} S^\top = S^\top A_{\mathcal{X}}.$$

The second identity is the transpose of the first.

We prove by induction that

$$S(x, y) > 0 \implies \widetilde{\chi}_t^{\mathcal{X}}(x) = \widetilde{\chi}_t^{\mathcal{Y}}(y) \tag{21}$$

at every depth. The base case follows from equation 20. Suppose the assertion holds at depth t. For each color γ, let $e_\gamma^{\mathcal{X}}$ be its indicator vector and $H_\gamma^{\mathcal{X}} = \mathrm{diag}(e_\gamma^{\mathcal{X}})$, with analogous notation on Y. The support condition and the stochastic constraints imply

$$S e_\gamma^{\mathcal{Y}} = e_\gamma^{\mathcal{X}}, \qquad S^\top e_\gamma^{\mathcal{X}} = e_\gamma^{\mathcal{Y}}, \qquad H_\gamma^{\mathcal{X}} S = S H_\gamma^{\mathcal{Y}}.$$

Indeed, a nonzero entry of S joins equal colors, and the corresponding row and column sums are one.

For unary neighborhoods, set $u_\gamma^{\mathcal{X}} = A_{\mathcal{X}} e_\gamma^{\mathcal{X}}$. Then

$$u_\gamma^{\mathcal{X}} = A_{\mathcal{X}} S e_\gamma^{\mathcal{Y}} = S A_{\mathcal{Y}} e_\gamma^{\mathcal{Y}} = S u_\gamma^{\mathcal{Y}}, \qquad u_\gamma^{\mathcal{Y}} = S^\top u_\gamma^{\mathcal{X}}.$$

Support rigidity shows equality of each color's neighbor count at positively matched cells. Since colors retain rank, boundary and coboundary counts are distinguished.

For mediated pairs, define

$$v_{\gamma,\zeta}^{\mathcal{X}} := A_{\mathcal{X}} H_\zeta^{\mathcal{X}} A_{\mathcal{X}} e_\gamma^{\mathcal{X}}.$$

At a rank-r cell x, when γ has rank r and ζ rank $r-1$, its entry counts exactly the pairs $(y, z) \in \mathcal{A}_\downarrow(x)$ with colors (γ, ζ). Taking ζ of rank $r+1$ gives the upper-pair count. Each mediator is counted separately; the case $y = x$ is included, as in the neighborhood definitions of the manuscript. The intertwining and color-mask identities give

$$\begin{aligned} v_{\gamma,\zeta}^{\mathcal{X}} &= A_{\mathcal{X}} H_\zeta^{\mathcal{X}} A_{\mathcal{X}} S e_\gamma^{\mathcal{Y}} \\ &= A_{\mathcal{X}} H_\zeta^{\mathcal{X}} S A_{\mathcal{Y}} e_\gamma^{\mathcal{Y}} \\ &= A_{\mathcal{X}} S H_\zeta^{\mathcal{Y}} A_{\mathcal{Y}} e_\gamma^{\mathcal{Y}} \\ &= S A_{\mathcal{Y}} H_\zeta^{\mathcal{Y}} A_{\mathcal{Y}} e_\gamma^{\mathcal{Y}} = S v_{\gamma,\zeta}^{\mathcal{Y}}. \end{aligned}$$

The transposed identities yield $v^{\mathcal{Y}}_{\gamma,\zeta} = S^\top v^{\mathcal{X}}_{\gamma,\zeta}$. Support rigidity therefore gives equality of all pair-color multiplicities at positively matched cells. Their current colors and all four context multisets agree, so the common injective CCWL update proves equation 21 at depth $t+1$.

At every depth, double stochasticity also gives

$$\mathbf{1}^\top e^{\mathcal{X}}_\gamma = \mathbf{1}^\top S e^{\mathcal{Y}}_\gamma = \mathbf{1}^\top e^{\mathcal{Y}}_\gamma.$$

Thus all global augmented-color multisets agree. Lemma C.5 implies original CCWL equivalence.

CCWL equivalence implies zero convex Co-OT. Conversely, suppose $\mathcal{X} \equiv_{\mathrm{CCWL}} \mathcal{Y}$. Their initial rank histograms agree, so their rank sizes and supports agree. By Lemma C.5, their augmented CCWL color multisets agree at every depth. Run augmented CCWL on the disjoint union with a common injective encoding until its partition stabilizes. Finiteness and retention of the preceding color ensure that this occurs after finitely many steps. Choose such a depth T. Matching color classes $C^{\mathcal{X}}_i, C^{\mathcal{Y}}_i$ have equal sizes $h_i > 0$. Define

$$S(x,y) := \begin{cases} 1/h_i, & x \in C^{\mathcal{X}}_i,\ y \in C^{\mathcal{Y}}_i, \\ 0, & \text{otherwise.} \end{cases}$$

Each row and column sums to one. Colors retain ranks and initial attributes, so S is rank-block diagonal and is supported on equal augmented attributes.

Stability and injectivity imply that every cell of color i, in either complex, has the same number d_{ij} of incidence neighbors of color j. The two unary channels supply these counts, with rank distinguishing the direction. Counting edges between the two color classes in either complex gives

$$h_i d_{ij} = h_j d_{ji}.$$

For $x \in C^{\mathcal{X}}_i$ and $y \in C^{\mathcal{Y}}_j$,

$$(A_{\mathcal{X}} S)(x,y) = d_{ij}/h_j, \qquad (S A_{\mathcal{Y}})(x,y) = d_{ji}/h_i.$$

These are equal by the preceding identity, proving $A_{\mathcal{X}} S = S A_{\mathcal{Y}}$. Reading both adjacent blocks gives the two identities in equation 19. Set $\pi_r = S_r/n_r$ on occupied ranks and use the unique null coupling on absent ranks. All residuals and attribute costs vanish, proving zero convex Co-OT distance.

We have proved the first equivalence in Proposition 4.2. The second is Proposition 3.3, which uses the same augmented initial attributes. This completes the proof. □

Remark D.6 (What the equivalence does and does not assert). The equality of zero sets concerns convex Co-OT and the CCWL distance. It does not assert equality of their positive values or of finite-depth behavior. Nor does it identify the zero set of unrelaxed Co-OT with CCWL equivalence. The relaxation implies only

$$D_{\text{Co-OT}} = 0 \implies D^{\text{cvx}}_{\text{Co-OT}} = 0.$$

The reverse implication can fail: with identical original attributes, a six-cycle and the disjoint union of two triangles, regarded as vertex–edge complexes, have the same augmented CCWL colors and multiplicities at every depth. Their convex Co-OT distance is zero, but their Co-OT distance is positive. Indeed, a zero Co-OT coupling would, after equal-size rescaling, admit rank-wise permutations supported on its doubly stochastic blocks (by the Birkhoff–von Neumann theorem). Nonnegativity of the original incidence objective forces all incidence comparisons along those supported permutations to agree. This would give an isomorphism of the two incidence graphs, impossible because one is connected and the other has two connected components.

Remark D.7 (Role of the augmented attributes). The full rank-count sequence removes the need for equal rank sizes as an input assumption by forcing them at zero cost. This argument gives no uniform positive separation between different size classes: its positive lower bound is pair-dependent. The neighborhood counts ensure consistency with the initial attributes of the CCWL distance; stable-equivalence preservation is exactly the content of Lemma C.5. The zero-set proof uses exact equality of augmented attributes, not a Lipschitz comparison between different attribute metrics. The relaxation and pseudometric arguments require the same attribute metric in the two Co-OT objectives.

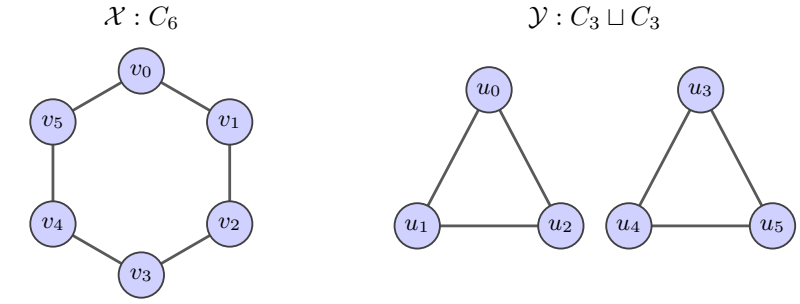
D.5 Strict separation between Co-OT and CCWL

The zero-set equivalence with CCWL applies to convex Co-OT, but not to unrelaxed Co-OT. We give a counterexample using the same augmented initial attributes in all three constructions. The example has equal rank sizes, so the separation does not rely on missing ranks or different cardinalities.

The two complexes. Let $\mathcal{X}$ be the vertex–edge complex of the six-cycle C_6, and let $\mathcal{Y}$ be that of two disjoint triangles $C_3 \sqcup C_3$. Explicitly, write their vertices as $v_0, \dots, v_5$ and $u_0, \dots, u_5$, respectively, and take the edges

$$X_1 = \big\{\{v_i, v_{i+1 \bmod 6}\} : 0 \le i \le 5\big\},$$
$$Y_1 = \big\{\{u_0, u_1\}, \{u_1, u_2\}, \{u_2, u_0\}, \{u_3, u_4\}, \{u_4, u_5\}, \{u_5, u_3\}\big\}.$$

The rank-zero cells are the singleton vertices, identified with their vertices. Neither complex has cells of rank two or higher; in particular, the triangles in $\mathcal{Y}$ are not filled faces. Assign the same original attribute c_* to every cell in both complexes. Figure 4 illustrates the pair.



$$D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) = D^{\mathrm{cvx}}_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) = 0$$
$$D_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) > 0$$

Figure 4: *A strict separation between unrelaxed Co-OT and CCWL. Both vertex–edge complexes have six vertices of degree two and six edges, with identical original attributes. Vertex names identify cells in the illustration and are not input attributes. The triangular regions on the right do not represent rank-two cells.*

Augmented initial attributes. Use the prescribed augmentation

$$q_0^{\mathcal{Z}}(x) = \big(c_*, \mathrm{rk}(x), \boldsymbol{m}_{\mathcal{Z}}(x), \boldsymbol{n}(\mathcal{Z})\big), \qquad \mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}\},$$

where the neighborhood counts are ordered as $(B, C, \downarrow, \uparrow)$. Both complexes have $\boldsymbol{n} = (6, 6, 0, \dots)$. Every vertex has two incident edges, and every edge has two endpoints. With self-neighbors included in the mediated adjacency contexts, this gives

$$\boldsymbol{m}_{\mathcal{Z}}(x) = \begin{cases} (0, 2, 0, 4), & \mathrm{rk}(x) = 0, \\ (2, 0, 4, 0), & \mathrm{rk}(x) = 1. \end{cases}$$

For a vertex, its four upper contexts consist of the two endpoints of each of its two incident edges, paired with that edge. For an edge, its four lower contexts consist of the two incident edges at each of its two endpoints, paired with that endpoint. Thus the augmentation is constant within each rank and agrees between the two complexes.

Proposition D.8 (Strict separation)**.** *For the complexes above, with the augmented initial attributes and any common attribute metric separating distinct attributes,*

$$D_{\mathrm{CCWL}}(\mathcal{X}, \mathcal{Y}) = D^{\mathrm{cvx}}_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) = 0, \qquad D_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) > 0.$$

This holds for every choice of the fixed strictly positive weights in the respective constructions.

Proof. CCWL does not distinguish the complexes. We prove inductively that all vertices in both complexes share one depth-t CCWL attribute, denoted a_t, and all edges share another, denoted b_t. At depth zero this follows from the augmented attributes computed above. If it holds at depth t, the four refinement multisets at each vertex are

$$M_{B,t} = \varnothing, \qquad M_{C,t} = \{\{b_t, b_t\}\}, \qquad M_{\downarrow,t} = \varnothing, \qquad M_{\uparrow,t} = \{\{(a_t, b_t) \text{ repeated four times}\}\}.$$

At each edge they are

$$M_{B,t} = \{\{a_t, a_t\}\}, \qquad M_{C,t} = \varnothing, \qquad M_{\downarrow,t} = \{\{(b_t, a_t) \text{ repeated four times}\}\}, \qquad M_{\uparrow,t} = \varnothing.$$

The current attributes and all four multisets therefore agree within corresponding ranks, proving the induction step. At every depth, each global multiset consists of six copies of a_t and six copies of b_t.

The distributional refinement obeys the same induction: its depth-t attributes have common values $z_{0,t}$ on vertices and $z_{1,t}$ on edges in both complexes. Empty channels agree, and all nonempty channel distributions are identical Dirac measures on the corresponding attributes or attribute pairs. Since both rank supports are $\{0, 1\}$, the reference measures give

$$L_t(\mathcal{X}) = L_t(\mathcal{Y}) = \frac{w_0}{w_0 + w_1}\delta_{z_{0,t}} + \frac{w_1}{w_0 + w_1}\delta_{z_{1,t}}.$$

Every depth-t Wasserstein distance is consequently zero, and so is D_{CCWL}.

Convex Co-OT is zero. Let J denote the 6×6 all-ones matrix and choose

$$\pi_0 = \pi_1 = \frac{1}{36}J.$$

Each row and column sums to $1/6$, so these are feasible rank couplings. Their attribute costs vanish because the augmented attributes agree within corresponding ranks. Both incidence matrices $I_1^{\mathcal{X}}, I_1^{\mathcal{Y}}$ have every row sum and every column sum equal to two. Consequently,

$$I_1^{\mathcal{X}}\pi_1 = \pi_0 I_1^{\mathcal{Y}} = \frac{1}{18}J, \qquad (I_1^{\mathcal{X}})^\top \pi_0 = \pi_1 (I_1^{\mathcal{Y}})^\top = \frac{1}{18}J.$$

All four rank sizes in the two normalized residuals are six, so both residuals vanish. Higher ranks are absent in both complexes and contribute zero under the common null convention. This feasible family has zero total convex Co-OT cost.

Unrelaxed Co-OT is positive. Suppose, to the contrary, that its minimum were zero. The feasible set on the two occupied ranks is compact and the objective is continuous, so a minimizing family (π_0, π_1) exists. Nonnegativity of every term and positivity of the structural weight imply

$$\sum_{v,e,u,f} \left|I_1^{\mathcal{X}}(v, e) - I_1^{\mathcal{Y}}(u, f)\right| \pi_0(v, u)\pi_1(e, f) = 0.$$

Therefore,

$$\pi_0(v, u) > 0, \quad \pi_1(e, f) > 0 \quad \Longrightarrow \quad I_1^{\mathcal{X}}(v, e) = I_1^{\mathcal{Y}}(u, f). \tag{22}$$

For each $r \in \{0, 1\}$, the matrix $S_r = 6\pi_r$ is doubly stochastic, and its positive support contains a perfect matching. Indeed, for any subset A of its row indices, let $N(A)$ be the columns receiving positive mass from A. The marginal constraints give

$$|A| = \sum_{i \in A} \sum_{j \in N(A)} S_r(i, j) \le \sum_{j \in N(A)} \sum_{i} S_r(i, j) = |N(A)|.$$

Thus Hall's marriage theorem supplies such a matching. Choose the resulting bijections $\sigma_0 : X_0 \to Y_0$ and $\sigma_1 : X_1 \to Y_1$, with $\pi_r(x, \sigma_r(x)) > 0$. Equation equation 22 then yields

$$I_1^{\mathcal{X}}(v, e) = I_1^{\mathcal{Y}}(\sigma_0(v), \sigma_1(e)) \qquad \text{for every } (v, e) \in X_0 \times X_1.$$

These bijections preserve vertex–edge incidence. They would therefore give an isomorphism between the two incidence graphs. This is impossible: the incidence graph of $\mathcal{X}$ is a connected twelve-cycle, whereas that of $\mathcal{Y}$ consists of two disjoint six-cycles. No minimizing family can have zero cost, and attainment implies that the minimum is strictly positive. □

Together with the relaxation inequality and the zero-set equivalence between convex Co-OT and CCWL, this example proves the strict inclusion

$$\{(\mathcal{X}, \mathcal{Y}) : D_{\text{Co-OT}}(\mathcal{X}, \mathcal{Y}) = 0\} \subsetneq \{(\mathcal{X}, \mathcal{Y}) : D_{\text{CCWL}}(\mathcal{X}, \mathcal{Y}) = 0\}.$$

Thus unrelaxed Co-OT can distinguish some complexes that both CCWL and convex Co-OT regard as equivalent, even when all three use the same augmented initial attributes.