sustainability

MDPI

Article

Geometric Optimal Transport for Sustainable Closed-Loop Supply Chain: A Fused Gromov–Wasserstein Framework for Structural and Attribute Inefficiency Diagnosis

Iman Seyedi [1,*], Antonio Candelieri [2,*] and Francesco Archetti [1]

[1] Department of Computer Science Systems and Communication, University of Milano-Bicocca, 20126 Milan, Italy; francesco.archetti@unimib.it
[2] Department of Economics Management and Statistics, University of Milano-Bicocca, 20126 Milan, Italy
* Correspondence: seyediman.seyedi@unimib.it (I.S.); antonio.candelieri@unimib.it (A.C.)

Abstract

Designing sustainable closed-loop supply chain (CLSC) networks requires jointly assessing node-level operational attributes (recovery efficiency, processing capacity, unit cost) and inter-node spatial structure. Existing methods, including mixed-integer programming, multi-objective metaheuristics, and graph-matching, typically optimize a single cost dimension and do not decompose structural connectivity from attribute-level inefficiency. We propose a Fused Gromov–Wasserstein (FGW) diagnostic framework that combines the Wasserstein distance (attribute similarity) and the Gromov–Wasserstein distance (structural alignment) via a convex trade-off parameter α, solved using the conditional gradient algorithm. Supply–capacity imbalances are resolved by marginal rescaling, with residual unabsorbed mass reported as a diagnostic indicator of infrastructure shortfall. The framework is applied to an eight-echelon PET bottle recovery and filament manufacturing network across 24 synthetic benchmark instances at three scale classes. The FGW cost decomposes exactly into feature and structural components, allowing bottleneck arcs to be diagnosed as attribute-driven or structure-driven. Under this benchmark, bottleneck cost decreases with network size, the most frequent bottleneck arc shifts from the collection interface in small networks to the mid-chain processing handoff in large networks, and attribute heterogeneity accounts for the majority of FGW cost (57.9%, conditional on the normalization and weighting scheme used) across all 144 arc–instance combinations. These results position FGW as a tractable, interpretable diagnostic layer for circular supply chain analysis, complementing rather than replacing classical CLSC design models.

Keywords: closed-loop supply chain; Fused Gromov–Wasserstein; optimal transport; reverse logistics; PET bottle recovery; sustainable network design; circular economy; multi-echelon optimization

Check for updates

Academic Editor: Jiang Wu

Received: 27 May 2026
Revised: 29 June 2026
Accepted: 29 June 2026
Published: 7 July 2026

1. Introduction

The transition toward a circular economy has placed closed-loop supply chain (CLSC) design at the center of sustainable operations research. Unlike traditional forward supply chains, CLSCs must simultaneously coordinate product collection, sorting, and end-of-life processing across multiple heterogeneous echelons, while balancing competing economic and environmental objectives [1]. In the context of multi-echelon circular material recovery, such as post-consumer PET bottle collection and filament remanufacturing, this challenge is particularly acute: collection centers and recycling facilities are spatially distributed, operate at different capacity scales, and exhibit substantially different material recovery rates and carbon footprints, making naive nearest-facility assignment strategies insufficient for system-level sustainability goals [2]. Critically, each echelon in a CLSC can be represented as a probability distribution over facilities, weighted by capacity or throughput, so that the assignment problem between echelons reduces to finding an optimal coupling between two such distributions. This distributional representation is not merely a mathematical convenience; it naturally encodes facility heterogeneity and allows assignment decisions to be sensitive to the full geometry of supply and demand, rather than to individual point-to-point costs alone.

Optimal transport (OT) theory offers a principled mathematical framework for this class of problems. Rooted in the classical works of Monge and Kantorovich, OT seeks a minimum-cost coupling between two probability measures, and its Wasserstein distance provides a geometrically meaningful metric over distributions of supply and demand [3]. Recent advances in computational OT, particularly the introduction of entropic regularization and the Sinkhorn algorithm, have made these methods tractable for large-scale logistics applications [4]. However, the standard Wasserstein formulation focuses on the feature space and ignores the relational structure among nodes. A critical limitation when the inter-facility spatial topology carries independent information about routing feasibility and network resilience.

The Gromov–Wasserstein (GW) distance addresses this gap by comparing metric measure spaces (mm-space) rather than distributions in a fixed ambient space, enabling transport plans to respect the internal geometry of each echelon independently [5]. However, GW optimizes purely for structural alignment, discarding all node-level attribute information. In the CLSC context, this means GW can align the spatial geometry of two echelons without any regard for whether the matched facilities are compatible in terms of capacity, cost, or recovery rate, exactly the attributes that determine sustainability performance. For CLSC design, where both sustainability attributes and spatial structure must jointly determine routing decisions, neither Wasserstein nor Gromov–Wasserstein alone is sufficient. The Fused Gromov–Wasserstein (FGW) distance, introduced by [6], resolves this tension by minimizing a convex combination of the two costs, parameterized by a trade-off coefficient α that controls the relative emphasis on feature similarity versus structural alignment. While FGW has found applications in graph classification, molecular comparison, and domain adaptation, its potential for supply chain network diagnostics remains largely unexplored.

This paper proposes one of the first applications of the FGW optimal transport to multi-echelon CLSC diagnostic analysis. We formulate the arc-level assignment problem across six arc interfaces of an eight-echelon network as an FGW problem in which the feature cost encodes processing capacity, unit cost, and material recovery efficiency, while the structural cost captures inter-node spatial alignment within each echelon. The exact decomposition of FGW cost into feature and structural components provides a principled diagnostic: it identifies not only which arc is the binding bottleneck but whether the inefficiency stems from attribute heterogeneity or spatial misalignment, determining the appropriate policy lever. The transport plan is solved via the conditional gradient algorithm for different α values, yielding an interpretable trade-off between attribute alignment and spatial efficiency. Experiments on 24 benchmark instances drawn from the CLSC network introduced by [2] reveal that the binding bottleneck arc shifts with network scale, moving from the collection interface in small networks to the mid-chain processing handoff in large networks, and that 57.9% of total FGW cost derives from attribute heterogeneity rather than spatial misalignment, establishing operational standardization as the dominant efficiency lever. The α trade-off frontier, bottleneck arc identification, and scale-effect analysis provide practical guidance for infrastructure planning and policy design in circular supply chain networks.

The remainder of the paper is organized as follows. Section 2 reviews related work across CLSC network design, computational OT, and GW distances. Section 3 develops the FGW formulation, including the measure network (m-net) representation of CLSC echelons, marginal balancing, feature cost matrix, and cost decomposition. Section 4 instantiates the framework on an eight-echelon PET bottle recovery network across 24 benchmark instances. Section 5 reports the scale effect, bottleneck arc identification, feature-structural cost decomposition, and statistical validation. Section 6 concludes with policy implications and future research directions.

2. Related Work

This section reviews three streams of literature directly relevant to this paper: (i) closed-loop supply chain network design and sustainability, (ii) optimal transport theory and its theoretical framework, and (iii) Gromov–Wasserstein and Fused Gromov–Wasserstein distances and their emerging applications in structured data.

2.1. Closed-Loop Supply Chain Network Design

The design of CLSC networks has received sustained attention since the seminal review of [7], which synthesized studies on reverse logistics and identified multi-echelon network design under uncertainty as a central open challenge. Early contributions formulated CLSC design as mixed-integer linear programming (MILP), seeking to minimize total cost over forward and reverse flow simultaneously [8,9]. Ref. [10] introduced a three-objective CLSC model balancing economic cost, environmental impact, and social equity, while ref. [11] extended this to stochastic demand with carbon emission constraints under multiple regulatory scenarios.

The relationship between CLSC design and the circular economy has been systematically examined by [12], who reviewed CLSC articles and concluded that the majority apply a reductionist interpretation of circularity by focusing almost exclusively on monetary costs, neglecting material recovery efficiency and echelon-level attribute heterogeneity as explicit design criteria. This gap motivates the present work: rather than assigning flows to minimize cost alone, we treat recovery rates and carbon intensities as first-class features in the optimization objective. More recent contributions have addressed multi-echelon 3D printing and PET bottle waste CLSCs specifically [2], confirming that the asymmetric structure of collection-to-recycling flows, where the number of sources and sinks differs, and their attribute distributions are heterogeneous, poses challenges that standard MILP formulations do not address at the level of individual flow coupling.

2.2. Optimal Transport Theory and Computational Advances

Optimal transport theory, originating in the works of Monge (1787) [13] and extended by Kantorovich (1942) [14], provides a mathematically principled framework for comparing probability distributions by minimizing the expected cost of moving mass from one distribution to another. The modern treatment by [3] established the Wasserstein distance as a geometrically meaningful metric with strong theoretical properties, including continuity with respect to weak convergence and sensitivity to distributional support structure. Unlike information-theoretic divergences such as the Kullback–Leibler or total variation distances, the Wasserstein distance satisfies the axioms of a metric and reflects the spatial arrangement of mass within the distributions [15,16]. It offers a practical discrepancy measure between probability distributions by incorporating the underlying geometry of the sample space, which has been recently applied in the distributionally robust optimization setting [17].

The field underwent a computational renaissance following the introduction of entropic regularization and the Sinkhorn-Knopp algorithm by [18], which reduced the complexity of solving discrete OT problems from cubic to near-linear in the number of support points, enabling application at logistics-relevant scales.

Peyré and Cuturi (2019) [4] provide the most comprehensive review of computational OT, covering Sinkhorn variants, multiscale methods, stochastic approaches, and applications in imaging, machine learning, and statistics. Their monograph establishes the theoretical and algorithmic foundations that this work builds upon. The open-source POT library [19] implements these methods, including the FGW solver used in this paper, and has become the standard computational reference for OT research in machine learning and applied mathematics. Despite the maturity of OT as a computational tool, its application to supply chain network diagnostics remains nascent: prior work has used Wasserstein distances for demand distribution comparison [20] and inventory control [21] and distributionally robust facility location under uncertain demand [22], but the joint feature-structure formulation introduced here has not previously been applied to CLSC routing.

2.3. Gromov–Wasserstein and Fused Gromov–Wasserstein Distances

A key limitation of the Wasserstein distance is that it needs both measures to be in the same metric space [23]. The GW distance, introduced by [24] as a metric between mm-space, generalizes OT to settings where the two distributions live in incomparable metric spaces. It extends optimal transport to cases where distributions cannot be compared point by point [25]. Rather than matching points based on a fixed ground cost, GW matches relational structures by minimizing a quadratic distortion term that penalizes dissimilarities between pairwise distances within each space. This makes GW naturally suited to problems where the internal geometry of two networks, not just their node attributes, must be aligned. Ref. [26] established the first scalable computational framework for GW using entropic regularization, demonstrating its utility for shape matching, graph comparison, and heterogeneous domain adaptation.

Ref. [6] introduced the Fused Gromov–Wasserstein distance as a linear interpolation between the Wasserstein feature cost and the GW structural cost, controlled by a trade-off parameter α. This fusion enables a more comprehensive alignment of structured data such as cortical surfaces or functional networks, where both intrinsic geometry and node-level attributes carry meaningful information. FGW has since been applied to graph classification [27], molecular property prediction [28], and most recently to transportation network vulnerability assessment via its unbalanced extension, Fused Unbalanced GW (FUGW) [29]. The last of these is the closest existing application to the present work, ref. [29] apply FUGW to identify critical infrastructure in a benchmark transportation network, demonstrating that jointly capturing topological structure and demand characteristics yields more stable and interpretable rankings than graph-theoretic approaches alone. Our contribution differs in that we apply FGW to the reverse logistics assignment problem in an eight-echelon CLSC, using the α trade-off frontier as an explicit decision-support tool rather than a hyperparameter to be tuned. To the best of our knowledge, no prior work has applied FGW optimal transport to CLSC network diagnostics, leaving the diagnostic potential of its exact cost decomposition unexplored in this domain.

Beyond optimal transport, recent work has explored graph neural networks for representing and analyzing supply chain structure, including demand forecasting, anomaly detection, and route optimization tasks framed as learning problems over the supply chain graph [30]. Reinforcement learning has similarly been combined with graph-based state representations for adaptive inventory and routing decisions under disruption [31]. Stochastic and robust optimization approaches, typically formulated as two-stage or multistage stochastic programs, remain the dominant paradigm for incorporating uncertainty into CLSC network diagnostics, particularly under uncertain return quantity and quality [32]. These approaches generally require either large historical datasets for training (GNN/RL) or an explicit probabilistic model of uncertain parameters (stochastic programming), in contrast to the geometry-based, training-free decomposition into feature and structural cost provided by FGW, which instead requires only point-level attribute and distance information.

2.4. Contributions

This paper presents the first application of Fused Gromov–Wasserstein optimal transport to multi-echelon CLSC network diagnostics, filling a gap identified across all three literature streams reviewed above. A sustainability-weighted feature cost matrix encoding processing capacity, unit cost, and material recovery efficiency is introduced, with independent per-echelon normalization to prevent scale collapse across incommensurable features. Asymmetric non-square source-sink topologies and supply–capacity imbalances are handled through a principled mass-rescaling procedure that quantifies residual unabsorbed material as a policy-relevant infrastructure gap. The framework is evaluated on 24 benchmark instances across three scale classes, revealing a strong scale effect and a shift in the binding bottleneck arc across echelon interfaces as network scale increases. This pattern is undetectable by single-metric evaluations. Finally, exact FGW cost decomposition across all 144 arc-instance combinations establishes that 57.9% of total cost derives from attribute heterogeneity rather than spatial misalignment, positioning operational standardization, not facility relocation, as the dominant lever for CLSC efficiency improvement. The full α trade-off frontier is exposed as an interpretable decision-support tool for balancing attribute alignment against spatial restructuring objectives. We emphasize that the proposed framework is a diagnostic and decision-support layer, not a network design optimization model. It does not solve facility location, flow conservation, capacity-investment, or demand-satisfaction decisions, and is intended to complement, rather than replace, MILP-based, multi-objective, or simulation-based CLSC design models.

3. FGW Formulation for CLSC Network Diagnostics

3.1. Problem Statement and Network Structure

Consider a multi-echelon closed-loop supply chain comprising K echelons connected by directed arc interfaces. The central design problem is to determine, for each arc interface $A \rightarrow B$, the optimal coupling between the source echelon A and the sink echelon B, where the number of source nodes n and sink nodes m differ in general, yielding asymmetric non-square assignment structures. Throughout this paper, coupling refers to the FGW transport plan T between source and sink nodes. Assignment and routing are used informally to describe the resulting flow pattern, and optimization refers specifically to the FGW solver's minimization, not to a full network design optimization problem (see Section 2.4).

For a given arc $A \rightarrow B$, let $\{x_1, \ldots, x_n\}$ denote the set of source nodes in echelon A and $\{y_1, \ldots, y_m\}$ denote the set of sink nodes in echelon B. Each source node x_i is characterized by a feature vector $a_i \in \mathbb{R}^d$ encoding operational attributes relevant to the arc under analysis, and each sink node y_j by a feature vector $b_j \in \mathbb{R}^d$ encoding corresponding sink-side attributes, where d is the number of attributes. The source and sink marginals $\mu \in \mathbb{R}^n$ and $\nu \in \mathbb{R}^m$ are probability vectors encoding the relative throughput capacity of each node within its echelon, defined formally in Section 3.6.

The internal spatial structure of each echelon is captured by a normalized pairwise distance matrix $C_1 \in \mathbb{R}^{n \times n}$ among source nodes and $C_2 \in \mathbb{R}^{m \times m}$ among sink nodes.

These matrices encode the relational geometry within each echelon independently, without requiring a common coordinate space across A and B.

The arc-level FGW formulation is applied independently to each arc interface in the network. This yields, for each arc, an optimal transport plan $T^* \in \mathbb{R}^{\mathrm{n}\times\mathrm{m}}$ that jointly minimizes attribute dissimilarity and structural distortion between the two echelons. The cross-arc comparison of FGW costs identifies the binding bottleneck arc of the circular loop and decomposes total network inefficiency into its attribute-level and structural components. The specific echelon definitions, arc interfaces, and node attributes for the case study instantiation of this framework are given in Section 4.

The present formulation treats network parameters deterministically, establishing a baseline diagnostic layer; the FGW formulation is compatible with uncertainty-aware extensions, since both feature vectors and distance matrices can be replaced by their expected values or sampled scenarios without altering the underlying transport mechanism.

3.2. Closed-Loop Supply Chains as Measure–Network Spaces

3.2.1. Metric Measure (mm) Spaces

Graphs can be represented as mm-space (X, d_X, μ), an mm-space is a triple (X, d_X, μ), where (X, d_X) is a metric space and μ is a probability measure on X. In the discrete graph setting, mm-spaces are represented by weighted graphs whose node distributions are encoded by probability vectors. This framework provides a unified mathematical language for describing spaces endowed with both geometric structure (via the metric d) and distributional information (via the measure μ). Mm-spaces arise naturally in various contexts, including Riemannian geometry, probability theory, and graph analysis, and serve as the foundation for comparing structured data using optimal transport [29,33]. A broader review of mm-spaces is provided in Appendix A.

3.2.2. Measure Networks (m-Net) Space

The concept of an m-net space generalizes that of an mm-space by relaxing the requirement that the pairwise interaction function be a distance metric. This broadens the applicability of the GW framework to complex relational data structures. An m-net is a triple $\mathcal{X} = (X, \omega_X, \mu)$ where:

X is a separable and completely metrizable topological space (a Polish space);

μ_X is a fully supported Borel probability measure on X;

$\omega_X : X \times X \to R$ is a measurable function, called the network function.

The collection of all m-nets is denoted by $\mathcal{N}$. When ω_X satisfies the metric axioms and generates the topology of X, it is denoted d_X, and the corresponding triple (X, d_X, μ) is a mm-space belonging to the class $\mathcal{M}$. A broader review of mm-spaces is provided in Appendix A.

3.2.3. CLSC as m-Net Space

The network structure of a CLSC is fundamentally more complex than that of a forward chain: material flows are bidirected, facilities operate simultaneously in both channels, and the cost, lead time, and reliability of a forward arc from x to x′ are generically distinct from those of the corresponding return arc. Classical directed graph models and network flow formulations cannot adequately represent this bidirectionality, nor the coupling between forward and reverse flows it induces [34].

A CLSC is naturally formalized as a m-net space $X = (X, \omega_X, \mu_X)$, where X is a Polish space (which is a complete, separable metric space, equipped with their Borel σ-algebra [3]) of facilities, μ_X is a fully supported Borel probability measure over X encoding capacity or importance weights, and $\omega_X : X \times X \to \mathbb{R}$ is a measurable network function encoding pairwise relational structure. Critically, ω_X is not required to satisfy any metric axioms, and in the CLSC setting, it generically does not, since transportation costs are asymmetric and reverse-channel routing violates the triangle inequality [35]. The m-net therefore belongs to N rather than the subclass M of mm-spaces, and structural comparison between CLSC networks proceeds via the Gromov–Wasserstein distance, which operates directly on N without requiring pointwise facility alignment across networks [23,24]. Supply-chain networks may also include edge attributes. Each transportation arc can carry information such as lead-time distributions, time-dependent travel costs, reliability scores, or congestion patterns. These attributes can be formalized as elements of a metric space Z [36]—for example, a 1-D probability distribution of shipping delays or a time-varying distance that reflects daily fluctuations in traffic. Such representations create Z-networks, which can be compared using distances defined for the chosen attribute space Z. In this study, the CLSC is modeled as a static H-network in which forward-channel facility attributes, processing capacity, unit cost, CO_2 intensity, and material recovery rate are encoded in ψ^{x+}, and reverse-channel waste generation and logistics cost attributes are encoded in ψ^{x-}. Arc-level relational structure is captured by normalized Haversine distance matrices ω_X, which satisfy the requirements of a measurable network function but are not required to be metrics, consistent with the asymmetric cost structure of reverse logistics flows.

The bidirected nature of CLSC flows requires that the structural kernel and attribute spaces be decomposed by channel. Building on the attributed network framework of [37], a CLSC network is governed by hyperparameters $H = (p, \Omega, d_\Omega, \Psi, d_\Psi)$, where $p \in [1, \infty]$, (Ω, d_Ω) is a separable metric space of arc attributes, and $\left(\Psi^+, d_{\Psi^+}\right)$, $\left(\Psi^-, d_{\Psi^-}\right)$, are separable metric spaces encoding forward- and reverse-channel facility attributes, respectively. The resulting "CLSC H-network" is a 7-tuple $\left(X, \psi_x^+, \psi_x^-, \phi_X^+, \phi_X^-, \omega_X, \mu_X\right)$ in which:

- $\phi_X^+ \in L^p(\mu_X \otimes \mu_X)$ encodes the forward supply-chain structure, with $\phi_X^+(x, x')$ representing transportation cost or flow intensity from x to x′;
- $\phi_X^- \in L^p(\mu_X \otimes \mu_X)$ encodes the reverse logistics structure, with $\phi_X^-(x, x')$ representing the return arc from x to x′.
- $\psi_x^+ \in L^p(X, \mu_X; \Psi^+)$ represents forward-channel facility attributes such as production capacity, outbound service levels, and demand forecasts.
- $\psi_x^- \in L^p(X, \mu_X; \Psi^-)$ represents reverse-channel facility attributes such as return processing capacity, remanufacturing yield rates, and quality-grade distributions of recovered items.
- $\omega_X \in L^p(X \times X, \mu_X \otimes \mu_X; \Omega)$ models arc-level attributes jointly across both channels—for example, lead-time distributions or reliability scores encoded as elements of (Ω, d_Ω).

The separation of ϕ_X^+ and ϕ_X^- rather than their merger into a single asymmetric kernel is deliberate: it permits independent comparison of forward and reverse channel structures across networks, consistent with the operational reality that the two channels are governed by structurally different cost functions and service constraints [34]. Similarly, disaggregating ψ_x^+ and ψ_x^- reflects the existence of hybrid nodes, facilities simultaneously active in both channels with distinct attribute profiles for each role, whose properties would be suppressed by a single node feature vector. Arc attributes in Ω are naturally probabilistic, formalizing each transportation link as carrying a distribution over lead times or time-varying costs, consistent with the Z-network construction of [36]. In this study, the benchmark CLSC network includes forward demand attributes, reverse return-rate distributions at each facility, and time-varying arc conditions in both channels, yielding a partially probabilistic bidirected H-network illustrated in Figure 1.

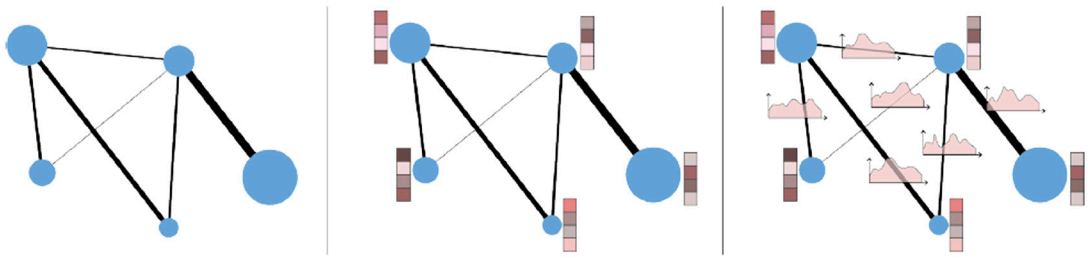

Figure 1. Illustration of network representations relevant to transport modeling. (**a**) m-net: A weighted graph where node size and edge thickness encode distance or flow intensity; (**b**) Node-Attributed Network: Extends the m-net by associating each node with feature vectors (e.g., socioeconomic or demand attributes), represented as small column bars. (**c**) Z-Network: Incorporates edge-level features (e.g., temporal or probabilistic profiles of congestion), represented here as distributions attached to edges. This setting captures dynamic or stochastic edge behavior—analogous to the transportation network in this study [36].

Of this general H-network formalism, the computational experiments in Section 4 instantiate a subset: forward-channel facility attributes (ψ_X^+) are represented by static, deterministic feature vectors (processing capacity, unit cost, CO_2 intensity, recovery rate), and the structural kernel ω_X is computed as a normalized Haversine distance matrix. The probabilistic edge-attribute formalism (Ω, d_Ω), the explicit separation of ϕ_X^+ and ϕ_X^-, and the treatment of arc attributes as distributions over lead times are presented as the general theoretical scaffolding within which the present case study is a static, non-probabilistic special case; their full instantiation, e.g., with time-varying or distributional arc costs, is left for future extensions of the framework.

3.3. Classical Optimal Transport

Optimal transport theory seeks the most efficient way to redistribute mass from a source distribution to a target distribution. Given discrete measures $\mu = \sum_i \mu_i \delta_{x_i}$ and $\nu = \sum_j \nu_j \delta_{y_j}$, the Kantorovich formulation finds a transport plan $\mathrm{T} \in \mathbb{R}_+^{n \times m}$ minimising the total transport cost:

$$\mathcal{W}(\mu, \nu) = \min_{\mathrm{T} \in \Pi(\mu,\nu)} \sum_{i,j} T_{ij} M_{ij} \tag{1}$$

where M_{ij} is the ground cost between the source x_i and sink y_j, and $\Pi(\mu, \nu) = \left\{ \mathrm{T} \geq 0 : \mathrm{T}1 = \mu, \mathrm{T}^\top 1 = \nu \right\}$ is the set of admissible couplings [3]. The resulting distance $\mathcal{W}$ is the Wasserstein distance, which metrises weak convergence of probability measures and is sensitive to the geometric arrangement of support points. In the CLSC context, the Wasserstein formulation assigns, for instance, the collection centers to recycling facilities purely based on node attribute similarity, ignoring the relational structure and the pairwise spatial topology within each echelon.

3.4. Gromov–Wasserstein Distance

The GW distance extends OT to settings where the source and target spaces are not directly comparable. Rather than matching nodes based on a fixed ground cost matrix, GW matches the internal metric structures of the two spaces by minimizing the distortion of pairwise distances under the coupling:

$$\begin{aligned} &\mathcal{GW}(\boldsymbol{\mu},\boldsymbol{\nu}) = \min_{\mathbf{T}\in\Pi(\boldsymbol{\mu},\boldsymbol{\nu})} \sum_{i,j,k,l} (C_X(i,k) - C_Y(j,l))^2 T_{ij} T_{kl} \\ &\text{Subject to} \\ &\mathrm{T}1_m = \mu, \mathrm{T}^\top 1_n = \nu, T_{ij} \geq 0 \forall i,j \end{aligned} \tag{2}$$

where $C_X(i,k)$ is the distance between collection centers x_i and x_k, and $C_Y(j,l)$ is the distance between recycling facilities y_j and y_l. The GW objective penalises couplings that distort the relative arrangement of nodes: if x_i and x_k are spatially close, but their matched sinks y_j and y_l are far apart, a large distortion cost is incurred. The coupling T represents a soft assignment between points in X and Y. The GW objective compares the pairwise distances in X and Y under this soft correspondence, enabling comparison of metric–measure spaces even when no cross-domain distances exist (Figure 2).

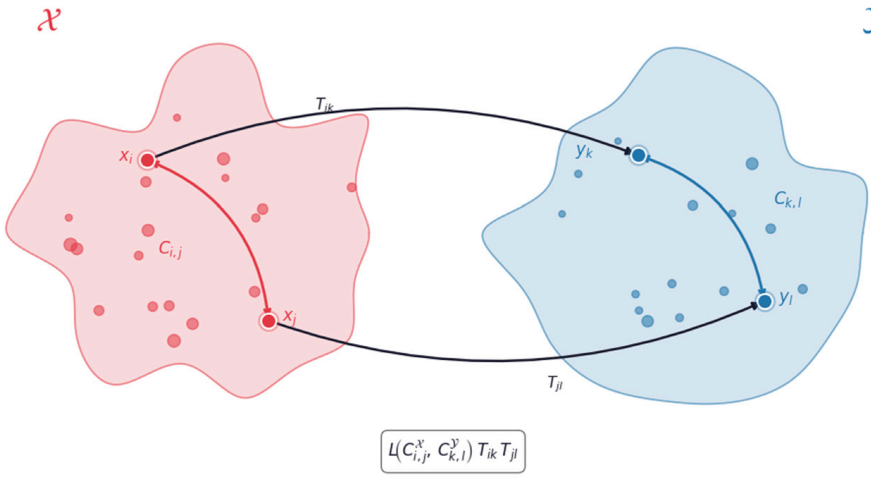

Figure 2. Illustration of GW.

This makes GW sensitive to network topology and routing feasibility in a way that standard Wasserstein is not. However, GW discards all node-level feature information, thereby optimizing purely for structural alignment and ignoring the sustainability attributes central to CLSC design. A broader review of computational GW solvers from the literature is provided in Appendix B.

3.5. Fused Gromov–Wasserstein

The Fused Gromov–Wasserstein distance [6] resolves the complementary limitations of Wasserstein and GW by minimizing a convex combination of both costs over the same transport plan:

$$\begin{gathered} \mathcal{L}^{\alpha}_{\mathrm{FGW}}(\mathrm{T}) = (1-\alpha)\sum_{i,j} T_{ij} M_{ij} + \alpha \sum_{i,j,k,l} (C_1(i,k) - C_2(j,l))^2 T_{ij} T_{kl} \\ \text{subject to:} \\ \mathrm{T}1_m = \mu, \mathrm{T}^\top 1_n = \nu, T_{ij} \geq 0 \forall i,j \end{gathered} \tag{3}$$

The parameter $\alpha \in [0,1]$ controls the trade-off between feature similarity and structural alignment. At $\alpha = 0$ the problem reduces to a standard Wasserstein assignment; at $\alpha = 1$ it reduces to pure GW structural matching. For CLSC design, intermediate values of α are operationally relevant: the transport plan must simultaneously respect compatibility of sustainability attributes between collection centers and recycling facilities and preserve the spatial coherence of routing decisions within each echelon.

Figure 3 illustrates the key idea behind the FGW distance. It visually represents how FGW aligns nodes between two graphs by considering both feature similarity (e.g., node attributes or embeddings) and structural similarity (relationships between nodes captured by the adjacency or structure matrices).

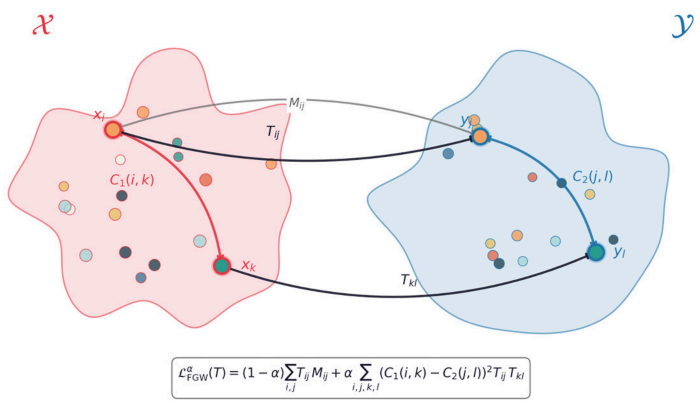

Figure 3. Feature and Structural Alignment in Fused Gromov–Wasserstein Coupling.

3.6. Marginal Distributions and Supply–Capacity Balancing

FGW is applied independently to each arc interface $A \rightarrow B$, where A denotes the source echelon with node set $\{x_1, \ldots, x_n\}$ and B denotes the sink echelon with node set $\{y_1, \ldots, y_m\}$, with n and m not necessarily equal. For a given arc, let $v_i > 0$ denote the raw throughput volume at the source node $x_i \in A$, and $c_j > 0$ denote the processing capacity of the sink node $y_j \in B$.

Before normalization, the absolute supply–capacity balance is assessed. When total supply exceeds total sink capacity, the infrastructure is insufficient to absorb all available material. The residual unabsorbed mass is defined as:

$$\Delta = \max\left(0, \Sigma_i v_i - \Sigma_j c_j\right) \tag{4}$$

and recorded as a policy-relevant infrastructure gap for the arc under analysis. While Δ is consistent with an infrastructure capacity gap, it should be interpreted as a diagnostic indicator rather than a definitive cause, since other operational or organizational factors not modeled here can also produce supply–capacity mismatches. Δ is a diagnostic quantity computed from raw data before any normalization and is not an input to the FGW solver.

The source and sink marginals required by the FGW formulation (Equation (3)) are then defined by normalizing throughput volumes and capacities to probability vectors:

$$\mu_i = v_i / V, \text{where } V = \sum_{i=1}^{n} v_i \tag{5}$$

$$\nu_j = c_j / C, \text{where } C = \sum_{j=1}^{m} c_j \tag{6}$$

By construction, $\Sigma_i \mu_i = 1$ and $\Sigma_j \nu_j = 1$, satisfying the balanced marginal condition required by the FGW solver. The normalization encodes the relative importance of each facility within its echelon; source nodes with higher throughput volume carry greater mass in the coupling, and sink nodes with higher capacity attract proportionally more routed material. The absolute imbalance Δ, having been recorded prior to normalization, is not obscured by this step and is reported separately alongside the FGW transport plan as a quantified infrastructure shortfall.

https://doi.org/10.3390/su18136906

This balanced formulation is adopted because the present diagnostic objective is to evaluate routing and matching quality independently of capacity-sizing decisions, which are treated as a separate planning question outside the scope of the transport problem itself. We note explicitly that this means absolute capacity shortages are reported as a diagnostic quantity (Δ) outside the solver, rather than optimized within the model. The FGW transport plan does not select storage, rerouting, outsourcing, investment, or disposal decisions in response to unabsorbed mass. An unbalanced or partial Gromov–Wasserstein formulation, such as that of [38], would allow capacity constraints to be incorporated directly into the optimization and represents a natural extension of the present framework for settings where capacity-sizing decisions must be made jointly with routing.

3.7. Feature Cost Matrix

The feature cost matrix $M \in \mathbb{R}^{n\times m}$ encodes pairwise attribute dissimilarity between source nodes in echelon A and sink nodes in echelon B. Each source node x_i is described by a feature vector $a_i \in \mathbb{R}^d$, and each sink node y_j by a feature vector $b_j \in \mathbb{R}^d$, where d is the number of matched operational attributes for the arc under consideration. d varies by arc (Section 4.4), reflecting the attributes shared by both echelons of that interface. These attributes were selected to capture the core economic and environmental dimensions of CLSC performance. The framework is attribute-agnostic and can be extended to incorporate additional dimensions such as lead time, reliability, risk, and regulatory constraints where data permits. In this study, $d = 4$, with attributes ordered as: throughput volume, unit handling cost, CO_2 emission intensity, and material recovery rate. To prevent features with larger absolute ranges from dominating the cost matrix, each feature dimension k is normalized independently within its echelon using min-max scaling:

$$\widetilde{a}_{ik} = \frac{a_{ik} - \min_i a_{ik}}{\max_i a_{ik} - \min_i a_{ik} + \varepsilon}, \mathrm{i} = 1, \ldots, \mathrm{n} \tag{7}$$

$$\widetilde{b}_{jk} = \frac{b_{jk} - \min_j b_{jk}}{\max_j b_{jk} - \min_j b_{jk} + \varepsilon}, \mathrm{j} = 1, \ldots, \mathrm{m} \tag{8}$$

where $\varepsilon > 0$ is a small constant added for numerical stability in case of zero range ($\varepsilon = 10^{-8}$ in this study). Independent min-max normalization rescales each attribute to [0, 1] within its own echelon, which prevents attributes with large numerical ranges from dominating the cost matrix but also removes absolute scale differences between source and sink echelons: a source node and a sink node can both receive high normalized values even where their absolute throughput or capacity differ substantially. This is a deliberate trade-off, since the present diagnostic objective is to compare relative attribute alignment within the matched feature set; settings where absolute scale itself drives feasibility (e.g., a sink echelon that is in absolute terms too small regardless of its relative position) would warrant joint normalization, physically meaningful scaling, or engineering-threshold-based normalization instead. We note this as a limitation, since the reported feature-cost share is computed under independent min-max normalization and would shift under an alternative scheme.

The pairwise feature cost between source node x_i and sink node y_j is defined as the weighted squared Euclidean distance between their normalized feature vectors:

$$M_{ij} = \sum_{k=1}^{d} w_k \left(\widetilde{a}_{ik} - \widetilde{b}_{jk}\right)^2 \tag{9}$$

where $w_k \geq 0$ are sustainability weights with $\Sigma_k w_k = 1$. The conditional gradient solver for the FGW objective requires the squared Euclidean form [6]. The linearization step in the Frank-Wolfe algorithm relies on the feature cost being a valid squared distance. By construction, after independent min-max normalization, each term $\left(\tilde{a}_{ik} - \tilde{b}_{jk}\right)^2 \in [0,1]$, so $M_{ij} \in [0,1]$ for all i, j.

Weights are set as $w = [0.10, 0.10, 0.40, 0.40]$ for throughput volume, unit cost, CO_2 emission intensity, and material recovery rate, respectively. The higher weights assigned to CO_2 intensity and recovery rate reflect their primacy as design criteria in circular economy regulation and sustainability assessment frameworks [2,39]. A sensitivity analysis over alternative weight specifications is reported in Appendix D.4.

3.8. Solution Algorithm and Cost Decomposition

The FGW problem is non-convex due to the quadratic GW term and is solved to a local minimum via the conditional gradient (Frank-Wolfe) algorithm, which iteratively linearizes the objective and solves a sequence of linear OT subproblems [6], implemented via the POT library [19]. The total cost at the optimal plan T^* decomposes as:

$$\mathcal{L}^{\alpha}_{\mathrm{FGW}}(\mathrm{T}^*) = \underbrace{(1-\alpha)\langle \mathrm{T}^*, \mathrm{M}\rangle}_{\text{feature cost (W)}} + \underbrace{\alpha \sum_{i,j,k,l} \left(C_1(i,k) - C_2(j,l)\right)^2 T^*_{ij} T^*_{kl}}_{\text{structural cost (GW)}} \quad (10)$$

The relative magnitudes of the two terms reveal the dominant design challenge in the network: a high feature share indicates that attribute heterogeneity is the binding constraint, while a high structural share indicates spatial misalignment dominates.

4. Case Study: Circular CLSC for 3D Printing Filament from Recycled PET Bottles

4.1. Case Study Overview

To demonstrate the proposed FGW framework, we instantiate it on a CLSC that combines two material streams: post-process waste from 3D printing centers and discarded PET bottles collected from end-users. This case is adapted from [2], who established a comprehensive MILP model for a network in which recycled PET bottles serve as the primary raw material for 3D printing filament production. The network is a canonical instance of a multi-echelon circular economy design problem: material flows are bidirectional, echelons are heterogeneous in capacity and sustainability attributes, and both forward and reverse channels must be jointly optimized.

Rather than applying FGW to a single arc, the central methodological contribution of this paper is a Multi-Arc FGW Diagnostic Framework: FGW optimal transport is applied independently to each of the six primary arcs of the CLSC. The resulting per-arc cost decompositions, which separate feature mismatch from structural misalignment, are compared across all arcs to identify the binding sustainability bottleneck of the circular loop and to diagnose the appropriate policy lever for each arc.

4.2. Network Structure and Echelons

The proposed network, illustrated in Figure 4, comprises eight echelons organized across two interdependent sub-networks [2]. The echelons are End-User (E), Collection Center (CC), Processing Center (PC), Treatment Center (TC), Recycling Center (RC), 3D Printing Center (DPC), Filament Customer (FC), and Market. The two sub-networks are described below.

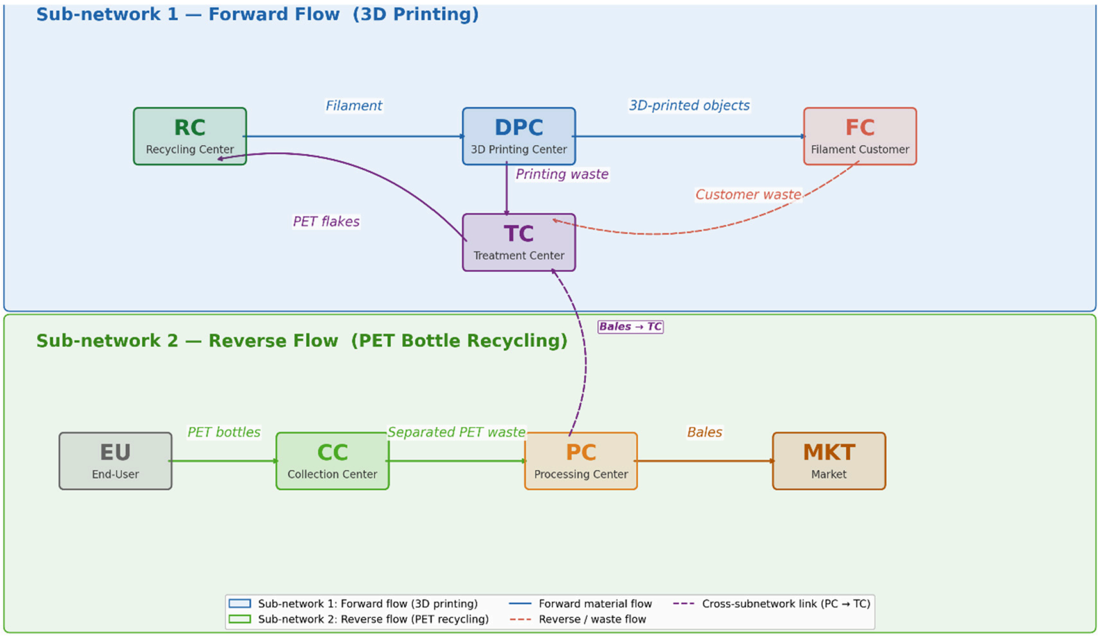

Figure 4. Proposed eight-echelon circular CLSC network. Sub-network 1 (3D printing loop) and Sub-network 2 (PET bottle recycling loop) are integrated via the PC $\rightarrow$ TC cross-subnetwork arc. The six FGW arcs A1–A6 are highlighted.

4.2.1. Sub-Network 1: 3D Printing Loop

Recycled filament produced at Recycling Centers (RC) is shipped to 3D Printing Centers (DPC), where it is consumed to manufacture products delivered to Filament Customers (FC). Two reverse waste flows re-enter the network: printing waste from DPCs and returned products from FCs are both directed to the Treatment Center (TC), where they are processed into PET flakes and cycled back to RC, closing the loop.

4.2.2. Sub-Network 2: PET Bottle Recycling Loop

Post-consumer PET bottles discarded by End-Users (E) are collected at Collection Centers (CC). Separated PET waste is forwarded to Processing Centers (PC), where bottles are sorted and compressed into bales. A portion of the bales is sold to Markets (recycling companies) to recover economic value. The remainder is forwarded via the cross-subnetwork arc to TC, where it is co-processed with 3D printing waste to produce filament-grade flakes that feed RC.

4.3. Multi-Arc FGW Diagnostic Framework

The central contribution of this paper is to apply FGW independently to each arc of the CLSC and compare the resulting cost decompositions across arcs. For each arc A $\rightarrow$ B, the FGW transport plan $T^{arc} \in \mathbb{R}^{n \times m}$ governs the optimal soft routing of material mass from echelon A to echelon B, solving Equation (3) where $\mu \in \mathbb{R}^{n}$ and $\nu \in \mathbb{R}^{m}$ are the capacity-proportional marginals defined in Section 3.6. Each arc receives its own feature cost matrix M^{arc} defined in Section 3.7, its own pairwise distance matrices $C_1^{arc} \in \mathbb{R}^{n \times n}$ and $C_2^{arc} \in \mathbb{R}^{m \times m}$ defined in Section 4.5, and its own α-frontier swept over $\alpha \in \{0.0, 0.1, \ldots, 1.0\}$. The six primary arcs to which FGW is applied are listed in Table 1.

Table 1. Six primary arcs of the circular CLSC to which the FGW framework is applied.

Arc	Source Echelon	Sink Echelon	Material Routed
A1	End-User (E)	Collection Center (CC)	Post-consumer PET bottle waste
A2	Collection Center (CC)	Processing Center (PC)	Separated PET waste
A3	Processing Center (PC)	Treatment Center (TC)	Sorted PET bales
A4	Treatment Center (TC)	Recycling Center (RC)	PET flakes
A5	Recycling Center (RC)	3D Printing Center (DPC)	Recycled filament
A6	DPC/FC	Treatment Center (TC)	Post-process waste (reverse)

Rather than fixing α arbitrarily, the full α-frontier is computed for every arc and instance. The arc that consistently carries the highest total FGW cost across $\alpha \in [0.3, 0.7]$ is identified as the binding sustainability bottleneck of the circular loop.

4.4. Node Attributes and Feature Cost Matrices

Each echelon node is characterized by a feature vector encoding operational and sustainability attributes, as defined generically in Section 3.7. Source and sink echelons carry their own attribute sets, with dimensionality d_a and d_b respectively, which vary by arc. Feature vectors are normalized independently within each echelon before computing M^{arc}. The per-arc attribute specifications and sustainability weight vectors are as follows. Empirical validation of the selected arc-level sustainability weight vectors, including sensitivity analysis across four alternative weighting schemes, is provided in Appendix D.4.

Arc A1 (End-User $\rightarrow$ CC): End-user nodes are characterized by PET bottle generation rate and spatial accessibility score. CC nodes are characterized by collection capacity and unit collection cost.

Arc A2 (CC $\rightarrow$ PC): CC nodes are characterized by outbound PET waste volume (kg/year) and separation efficiency. PC nodes are characterized by sorting capacity (kg/year), unit processing cost (USD/kg), and sorting yield rate (dimensionless). Equal weights are assigned across attributes (w = 0.33, 0.33, 0.34).

Arc A3 (PC $\rightarrow$ TC): PC nodes are characterized by bale output rate, compression cost, and yield. TC nodes are characterized by treatment capacity, unit cost, and flake yield rate.

Arc A4 (TC $\rightarrow$ RC): TC nodes are characterized by flake output volume, unit output cost, and yield. RC nodes are characterized by filament capacity, unit cost, and CO_2 intensity.

Arc A5 (RC $\rightarrow$ DPC): RC nodes are characterized by filament production rate, unit cost, and CO_2 intensity. DPC nodes are characterized by printing capacity, unit cost, and demand rate.

Arc A6 (DPC/FC $\rightarrow$ TC): DPC and FC nodes jointly form the source echelon, characterized by the waste generation rate (kg/year), the unit reverse logistics cost (USD/kg), and the waste CO_2 intensity. TC nodes carry the same attributes as in A3. For this reverse arc, equal weights are assigned (w = 0.33, 0.33, 0.34), as minimizing reverse flow cost and carbon burden are equally important.

The pairwise feature cost between the source node x_i and sink node y_j for each arc, Equation (9) of Section 3.7, with independent min-max normalization applied within each echelon prior to cost computation.

4.5. Spatial Distance Matrices

For each arc A → B, the internal pairwise distance matrix $C_1^{arc} \in \mathbb{R}^{n \times n}$ among source nodes and $C_2^{arc} \in \mathbb{R}^{m \times m}$ among sink nodes are computed from geodesic great-circle distances between facility coordinates using the Haversine formula. By definition, diagonal entries satisfy $C_1(i,i) = C_2(j,j) = 0$ for all i, j. Both matrices are subsequently normalized to the unit interval by dividing by their respective maximum entry:

$$\widetilde{C}_1 = C_1 / \max_{i,k} C_1(i,k), \widetilde{C}_2 = C_2 / \max_{j,l} C_2(j,l) \tag{11}$$

This ensures the structural GW cost is dimensionless and commensurable with the normalized feature cost M^{arc}. Independent per-echelon normalization is applied consistently across all six arcs, so that the structural cost at each arc reflects the relative spatial arrangement within that echelon rather than absolute inter-facility distances.

4.6. Bottleneck Identification and Cost Decomposition

At the optimal transport plan T^{*arc} for each arc, the FGW cost decomposes into feature and structural components as defined in Equation (10) of Section 3.8. The feature cost share φ^{arc}, also defined in Section 3.8, is computed for every arc and instance at α = 0.5 and reported in Appendix D.

The arc with the highest total FGW cost at α = 0.5 is identified as the binding sustainability bottleneck for that instance. We use 'bottleneck' in the model-internal sense of the arc carrying the highest FGW cost. The distinction between per-instance bottleneck frequency and aggregate arc-cost ranking is discussed in Section 5.3. Since the bottleneck ranking may in principle depend on α, stability is assessed via the full α-frontier: the bottleneck is considered robust when it is consistently the highest-cost arc across α ∈ [0.3, 0.7], as verified in Section Sensitivity Analysis and Justification of α = 0.5 for all 24 instances.

The feature cost share φ^{arc} diagnoses the nature of each bottleneck and determines the appropriate policy lever. When $\varphi^{arc} > 50\%$, attribute heterogeneity is the dominant source of inefficiency, and investment should target operational standardization, improving recycling yield, reducing CO_2 intensity, or upgrading process technology. When $\varphi^{arc} < 50\%$, structural misalignment dominates, and spatial reconfiguration of facility locations will yield greater efficiency gains. The 50% threshold is used as a descriptive midpoint distinguishing the two cost components, not as a theoretically derived decision boundary. The appropriate threshold for triggering a specific managerial intervention would depend on the relative cost of standardization versus relocation in a given operational setting. Plan sparsity, which is the fraction of zero entries in T^{*arc} reported as a supplementary stability indicator: a constant sparsity pattern across α confirms that the qualitative routing topology is robust to changes in the feature–structure trade-off. Full per-arc results and cross-arc bottleneck rankings are reported in Sections 5.3 and 5.4.

5. Results and Analysis

5.1. Overview of Computational Experiments

This section presents the results of applying the FGW-based CLSC diagnostic framework to the 24 benchmark instances defined in Table 2 and fully detailed in Appendix C (Table A1), spanning small (Problems 1–8), medium (Problems 9–16), and large (Problems 17–24) scale configurations. For each instance, the FGW transport plan is computed across six arc interfaces, A1 (E → CC), A2 (CC → PC), A3 (PC → TC), A4 (TC → RC), A5 (RC → DPC), and A6 (DPC/FC → TC), at eleven α values from 0.0 to 1.0 in steps of 0.1, yielding 66 transport plans per instance and 1584 in total. The reference trade-off parameter is α = 0.5, which assigns equal weight to structural (spatial) misalignment and feature (attribute) divergence; its empirical justification is established in Section 5.3. Infrastructure gap diagnostics are reported for arcs where total source throughput exceeds total sink capacity, as defined in Section 3.6. Detailed per-instance cost decomposition and feature share results for all arcs and α levels are reported in Appendix D (Tables A2 and A3; Figures A3–A7).

Table 2. Summary of FGW Bottleneck Cost by Problem Size (α = 0.5).

Size	Instances	Mean Cost	Std Dev	Min Cost	Max Cost	Dominant Arc
Small	1–8	0.2262	0.0757	0.1450	0.3414	A2 (3/8)
Medium	9–16	0.0819	0.0102	0.0662	0.1010	A2 (4/8)
Large	17–24	0.0502	0.0068	0.0419	0.0598	A3 (4/8)

5.2. Scale Effect

Table 2 summarizes the bottleneck arc FGW cost, which is the maximum FGW cost across the six arcs, grouped by problem size at α = 0.5. The bottleneck cost decreases systematically across size groups, with a reduction of approximately 88% between the highest-cost small instance (Problem 1, FGW = 0.341) and the lowest-cost large instance (Problem 21, FGW = 0.042). Within each size group, costs are not strictly monotone, as individual instances fluctuate due to differences in node counts and cost distributions. However, the group means decrease sharply and without overlap across the three scale classes (Table 2), confirming a robust scale effect. Because costs are computed under independent min-max normalization within each instance, larger networks—which offer more matching opportunities—may mechanically produce smoother, more distributed couplings independent of any real efficiency gain. The observed scale effect should therefore be interpreted as a property of the FGW coupling structure under this benchmark construction, not as evidence that larger real-world CLSC networks are inherently more efficient. The Spearman rank correlation between the number of FC-layer nodes (N) and the bottleneck cost is $\varrho = -0.967$, confirming a strong, consistent scale effect. The standard deviation also contracts from 0.076 (small) to 0.007 (large), indicating that the framework converges to stable, predictable cost levels in large-scale networks under the present benchmark construction. Figure 5 plots the bottleneck FGW cost trajectory across all 24 instances, and Figure 6 presents the corresponding boxplot by size group.

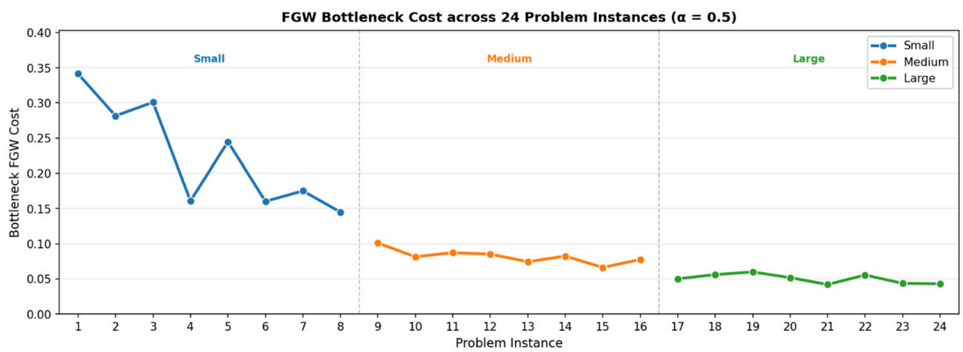

Figure 5. FGW Bottleneck Cost across 24 Problem Instances (α = 0.5).

https://doi.org/10.3390/su18136906

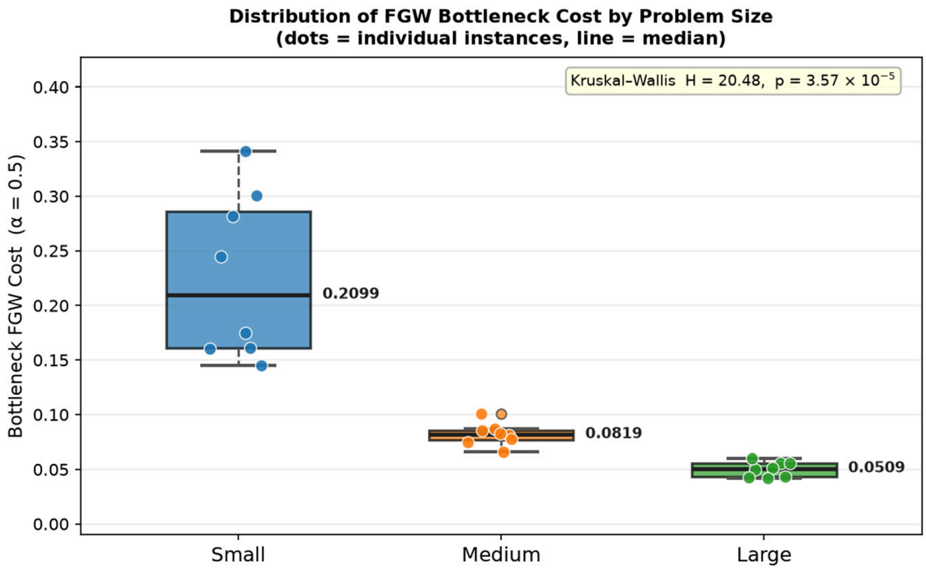

Figure 6. Boxplot of FGW Bottleneck Cost by Problem Size. Dots = individual instances. Kruskal–Wallis test confirms statistically significant group differences ($p < 0.001$).

5.3. Bottleneck Arc Identification

A central contribution of the FGW framework is its ability to identify which arc constitutes the binding constraint, which means the arc where structural and attribute misalignment is most severe. As shown in Figure 7, the dominant bottleneck arc is not fixed but tends to shift with problem size. In small instances, Arc A2 (CC → PC) is the bottleneck in 3 of 8 instances, A3 (PC → TC) and A6 (DPC/FC → TC) in 2 each, and A4 (TC → RC) in 1, indicating that forward-supply and collection interfaces are the primary constraints at a small scale. In medium instances, A2 remains dominant (4/8) while A4 (TC → RC) emerges strongly (3/8) and A3 appears once, reflecting growing reverse-channel pressure as the network grows. In large instances, A3 (PC → TC) becomes the primary bottleneck (4/8), with A4 appearing twice, A2 and A6 once each—indicating a shift in the most frequent bottleneck location from the collection interface toward the mid-chain processing handoff as network scale increases. Figure 8 shows the grouped boxplot of FGW cost per arc by problem size, confirming that A1 is consistently the lowest-cost arc across all size groups, while A2 and A3 have the highest median costs.

To avoid confusion between these two related but distinct quantities, note that these per-instance bottleneck counts, drawn from Table A2 at $\alpha = 0.5$, differ from the aggregate arc cost ranking reported in Table 3, which sums costs across all instances. Both analyses are complementary: the per-instance counts identify where bottlenecks occur most frequently, while the aggregate ranking quantifies which arc accumulates the most total coupling cost across the full benchmark.

This scale-dependent bottleneck migration implies that CLSC managers optimizing small networks should prioritize redesign of the collection interface (A2) and reverse waste feed-in (A6). In contrast, large-scale networks should focus investment on the processing-to-treatment handoff (A3) and the recovery pathway (A4). Arc A1 (E → CC) is never the binding constraint across any of the 24 instances (see Figure A4, Appendix D).

Table 3. Mean FGW Cost per Arc across the α-Frontier (mean across 24 instances). Bold = bottleneck arc at each α.

α	A1	A2	A3	A4	A5	A6	Bottleneck
0.0	0.0696	**0.1306**	0.0932	0.0962	0.0720	0.1036	A2
0.1	0.0712	**0.1272**	0.0935	0.0964	0.0735	0.1037	A2
0.2	0.0715	**0.1235**	0.0934	0.0956	0.0741	0.1033	A2
0.3	0.0710	**0.1190**	0.0929	0.0942	0.0737	0.1019	A2
0.4	0.0696	**0.1138**	0.0911	0.0917	0.0725	0.0998	A2
0.5	0.0674	**0.1078**	0.0877	0.0882	0.0707	0.0973	A2
0.6	0.0626	**0.1009**	0.0826	0.0834	0.0677	0.0936	A2
0.7	0.0579	**0.0912**	0.0769	0.0758	0.0629	0.0886	A2
0.8	0.0520	0.0803	0.0676	0.0686	0.0574	**0.0824**	A6
0.9	0.0476	0.0663	0.0555	0.0564	0.0460	**0.0715**	A6
1.0	0.0306	0.0408	0.0324	0.0373	0.0374	**0.0538**	A6

Notes: Blue-shaded rows mark the stable region $\alpha \in [0.3, 0.7]$. Bold and entries indicate the binding bottleneck arc at each α level. Arc interfaces: A1 (E $\rightarrow$ CC), A2 (CC $\rightarrow$ PC), A3 (PC $\rightarrow$ TC), A4 (TC $\rightarrow$ RC), A5 (RC $\rightarrow$ DPC), A6 (DPC/FC $\rightarrow$ TC).

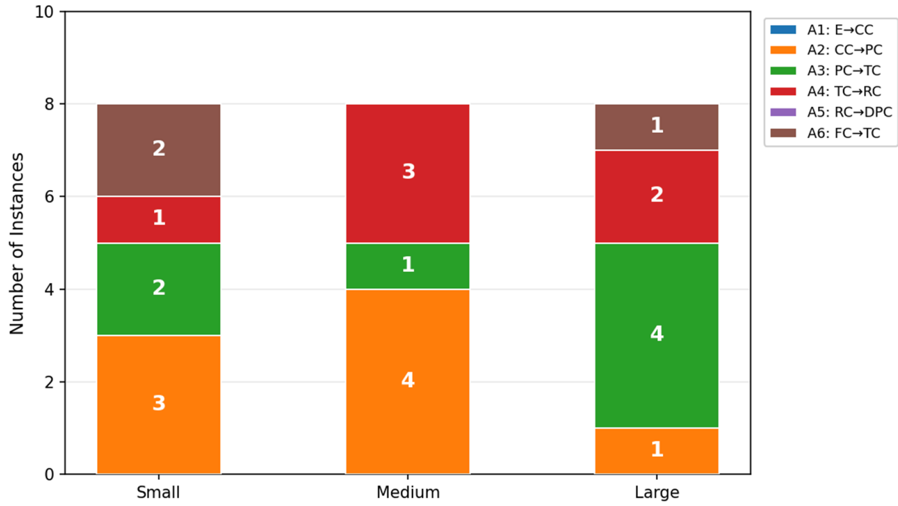

Figure 7. Bottleneck Arc Identity by Problem Size (α = 0.5).

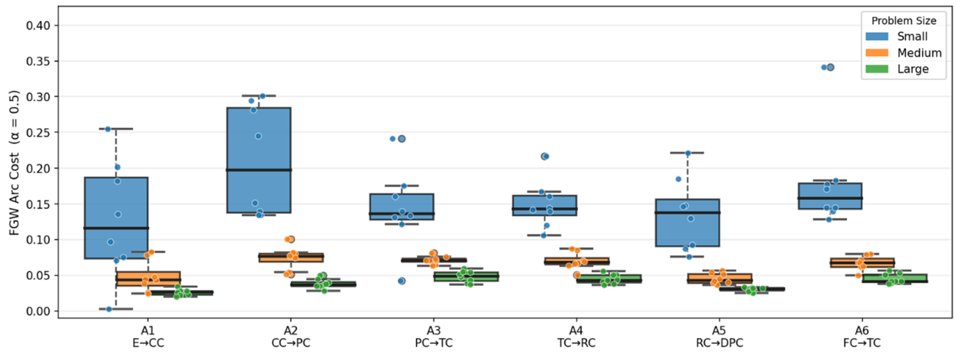

Figure 8. Grouped Boxplot of FGW Cost per Arc and Problem Size. A1 is consistently the lowest-cost arc. A2 and A3 have the highest median costs. (dots = individual instances, line = median).

Sensitivity Analysis and Justification of α = 0.5

To verify that the bottleneck identification above is not an artefact of the specific choice of α, Table 3 reports the mean FGW cost for each arc averaged across all 24 instances at each α level, with the binding bottleneck arc identified at each level. Figure 9 plots the resulting α-frontiers for all six arcs. Note that α = 0 and α = 1 in Table 3 correspond exactly to classical Wasserstein-only (pure attribute) and Gromov–Wasserstein-only (pure structural) bottleneck identification, respectively, providing a direct comparison against these established single-metric methods. The bottleneck ranking is consistent across both, with the discrepancy emerging only as $\alpha \to 1$.

Three findings establish $\alpha = 0.5$ as a robust and well-justified reporting reference.

First, A2 (CC $\to$ PC) is the binding bottleneck arc in aggregate at every $\alpha \in [0.0, 0.7]$, confirming that the collection-to-processing interface is consistently the highest-cost arc across the full feature-informative region of the trade-off space. Only at $\alpha \geq 0.8$ does the bottleneck shift to A6, a regime where the heavy structural weight substantially amplifies the reverse arc's spatial distortion, making it increasingly relevant only when facility relocation rather than attribute standardization is the dominant policy concern.

Second, the aggregate arc cost ranking A2 > A6 > A4 > A3 > A5 > A1 is fully preserved across $\alpha \in [0.3, 0.7]$ for all arcs except a minor positional exchange between A3 and A4 at $\alpha = 0.7$, where A3 (0.0769) edges ahead of A4 (0.0758), a difference of less than 1.5% that is operationally negligible. The top two positions (A2, A6) and the identification of the binding bottleneck are unaffected across the entire stable region.

Third, both A2 and A6 decline monotonically across the full α-frontier. A2 decreases from 0.1306 at $\alpha = 0.0$ to 0.0408 at $\alpha = 1.0$, driven by the declining weight on its dominant feature cost term. A6 decreases from 0.1036 at $\alpha = 0.0$ to 0.0538 at $\alpha = 1.0$, with a notably slower decline; its structural cost remains persistently elevated due to the spatial commingling of DPC and FC waste streams in the joint source echelon. This slower decline explains why A6 eventually overtakes A2 as the bottleneck above $\alpha = 0.8$: A6 is more structurally costly relative to A2, and this gap widens as α increases. All six arcs decline monotonically across the full α-frontier, consistent with the dominance of feature cost terms network-wide. The full α-frontier is shown in Figure 9.

The choice $\alpha = 0.5$ sits within the stable region where the ranking A2 > A6 > A4 > A3 > A5 > A1 is fully preserved, is well below the A6 crossover at $\alpha = 0.8$, and assigns equal weight to feature and structural cost without privileging either objective. It is therefore both a natural equal-weighting reference and an empirically validated robust choice for this benchmark network.

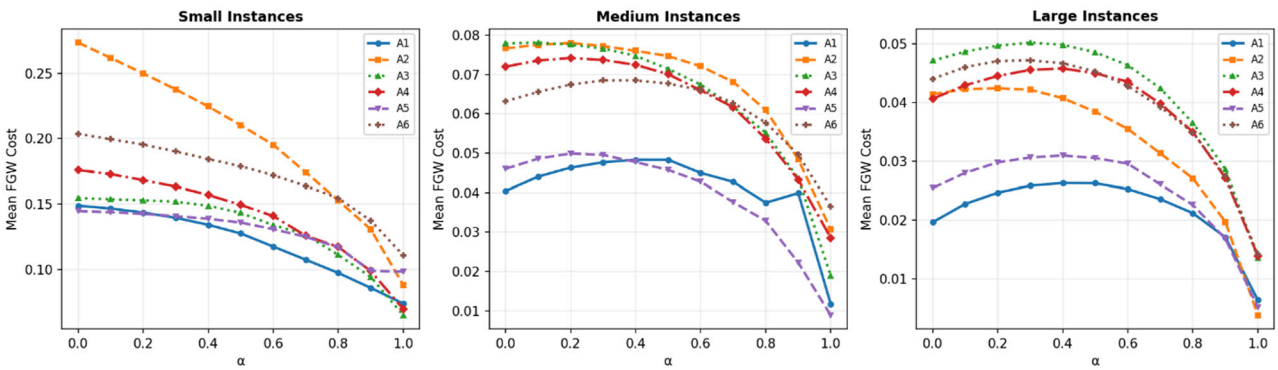

Figure 9. Mean FGW Cost vs. α per Problem Size Group.

5.4. Feature vs. Structural Cost

Figure 10 presents a heatmap of the feature cost share, which is the proportion of total FGW cost attributable to attribute divergence rather than spatial misalignment, for all six arcs across all 24 instances at $\alpha = 0.5$. Full per-instance values are reported in Table A3 (Appendix D).

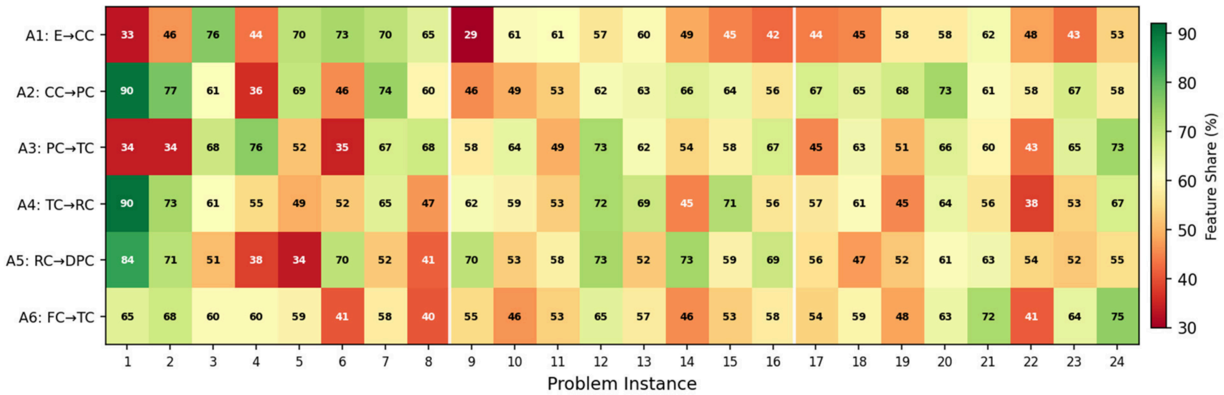

Figure 10. Feature Cost Share (%) at α = 0.5 for All Six Arcs × 24 Instances. Green = feature-dominated (φ > 65%); Red = structure-dominated (φ < 40%). White lines separate size groups.

Three key observations emerge.

First, feature cost consistently dominates. Across all 144 arc–instance combinations at α = 0.5, the mean feature share is 57.9%, and the feature component exceeds the structural component in 108 of 144 arc–instance pairs (75.0%). Per-arc mean shares are: A1 = 53.9%, A2 = 62.0%, A3 = 57.8%, A4 = 59.3%, A5 = 57.8%, A6 = 56.8%. This proportion is conditional on the independent min-max normalization, the four-attribute feature representation, the sustainability weighting scheme, and the α = 0.5 reference point adopted in this study (Section 3.7); a different normalization or weighting scheme would shift the exact feature/structural split, though Section Sensitivity Analysis and Justification of α = 0.5 shows the qualitative dominance of the feature term is stable across the α frontier. The feature share distribution per arc is shown in Figure A5 (Appendix D).

Second, A2 carries the highest mean feature share at 62.0%, consistent with its role as the aggregate bottleneck arc identified in Section Sensitivity Analysis and Justification of α = 0.5. The dominance of feature cost in A2 confirms that the CC → PC interface is primarily challenged by attribute heterogeneity, mismatches in collection capacity profiles, separation efficiency, and processing cost structures, rather than by the spatial arrangement of collection centers relative to processing centers. A4 follows at 59.3%, with A3 and A5 both at 57.8%.

Third, A1 has the lowest mean feature share at 53.9%, the closest to structural parity of any arc. While feature cost still marginally dominates on average, A1 exhibits the highest within-arc variance in feature share across instances, ranging from 33.4% (Problem 1, strongly structure-dominated) to 76.0% (Problem 3, feature-dominated), reflecting the sensitivity of the end-user collection interface to instance-specific geographic configurations.

The mean feature share decreases as α increases for all arcs, as expected from the FGW objective. At the mean level, the crossover point at which structural cost exceeds feature cost (φ < 50%) occurs at α = 0.7 for A1, A3, A4, and A6, and at α = 0.8 for A2 and A5. This confirms that feature cost dominance is robust well into the upper half of the trade-off space and is not an artefact of the α = 0.5 reference choice. The full α-sensitivity of the feature share decomposition is shown in Figure A6 (Appendix D).

The direct policy implication is that CLSC redesign should prioritize attribute standardization—harmonizing collection capacity profiles, processing yields, and cost structures across echelons—rather than spatial relocation of facilities. This finding holds across 75% of all arc–instance combinations and is most pronounced at the aggregate bottleneck arc A2, where investment in standardizing collection-center and processing-center attribute profiles will deliver the largest system-level efficiency gains.

5.5. Infrastructure Capacity Gaps

The mass balance diagnostic defined in Section 3.6 identifies supply–capacity imbalances for each arc prior to normalization. Two arcs exhibit structural capacity gaps across the majority of benchmark instances.

Arc A3 (PC $\rightarrow$ TC) has a positive gap in 23 of 24 instances, indicating that processing center bale output systematically exceeds treatment center intake capacity across nearly the entire benchmark. The gap grows with network scale, ranging from 442 units in Problem 2 (Small) to 6133 units in Problem 23 (Large), reflecting that treatment center capacity does not scale proportionally with processing center output as the network grows. This persistent and worsening gap across scale classes constitutes the most significant infrastructure shortfall in the network and represents a structural investment priority independent of FGW assignment quality.

Arc A5 (RC $\rightarrow$ DPC) has a positive gap in 20 of 24 instances, concentrated in small and medium instances. Recycling center filament output exceeds 3D printing center absorption capacity in these cases, with gaps ranging from 462 units (Problem 1) to 5195 units (Problem 10). The four instances without a gap are all large-scale (Problems 21–24 partially), suggesting that DPC capacity catches up with RC output at larger network scales. This arc therefore represents a scale-dependent bottleneck: a critical infrastructure gap at small and medium scales that partially resolves at large scales.

Arc A1 (E $\rightarrow$ CC) has a positive gap in 5 of 24 instances, all in the large-scale class (Problems 20–24), where end-user waste generation volumes begin to exceed collection center capacity. The gaps are modest relative to A3 and A5, ranging from 424 to 1666 units, but signal an emerging capacity constraint at the network entry point as scale increases.

Arcs A2, A4, and A6 are fully balanced across all 24 instances (gap = 0 in all cases), indicating that the processing-to-sorting, treatment-to-recycling, and reverse waste feed-in interfaces are appropriately sized relative to their upstream sources across the entire benchmark.

5.6. Statistical Validation

A Kruskal–Wallis non-parametric test confirms that the three size groups exhibit statistically significant differences in bottleneck FGW cost (H = 20.48, $p < 0.001$), justifying the size-stratified analysis in Sections 5.2–5.4. Pairwise Mann–Whitney U tests confirm that all three group pairs differ significantly (Small vs. Medium: $p = 0.0002$; Medium vs. Large: $p = 0.0002$; Small vs. Large: $p = 0.0002$). Given the complete separation of group ranges reported in Section 5.3 (no overlap across the three size classes), the rank-biserial effect size for each pairwise comparison is approximately 1.0, indicating a maximal effect, not merely statistical significance. Applying a Bonferroni correction for the three pairwise tests ($\alpha = 0.05/3 = 0.0167$) leaves all three comparisons significant ($p = 0.0002$ in each case). We note that the 24 instances form a parameterized sweep of increasing facility counts rather than independent random draws (Appendix C), and the reported effect should be interpreted accordingly as a property of this monotonic construction rather than a fully general statistical inference over independent samples.

The Pearson correlation matrix of arc FGW costs across the 24 instances is shown in Figure 11, with Panel A presenting the full heatmap and Panel B illustrating six representative arc-pair scatter plots. A4 and A5 are the most strongly correlated pair (r = 0.959), indicating that the TC $\rightarrow$ RC and RC $\rightarrow$ DPC arcs respond almost identically to changes in network scale—consistent with their sequential position in the recovery sub-network. A2 is strongly correlated with both A4 (r = 0.906) and A6 (r = 0.906), suggesting that the collection-to-processing interface co-moves with the mid-chain recovery and reverse feed-in arcs. A3 is moderately correlated with A2 (r = 0.683) and more strongly with A1 (r = 0.860) and A4 (r = 0.850), reflecting its intermediate position in the chain. A6 is weakly correlated with A1 (r = 0.446) and A3 (r = 0.475), confirming that the reverse arc's cost behaviour is largely independent of the forward collection and processing interfaces. These patterns suggest that scale-driven cost increases propagate differently across the forward and reverse sub-networks, a hypothesis that could be explored via structural equation modelling in future work.

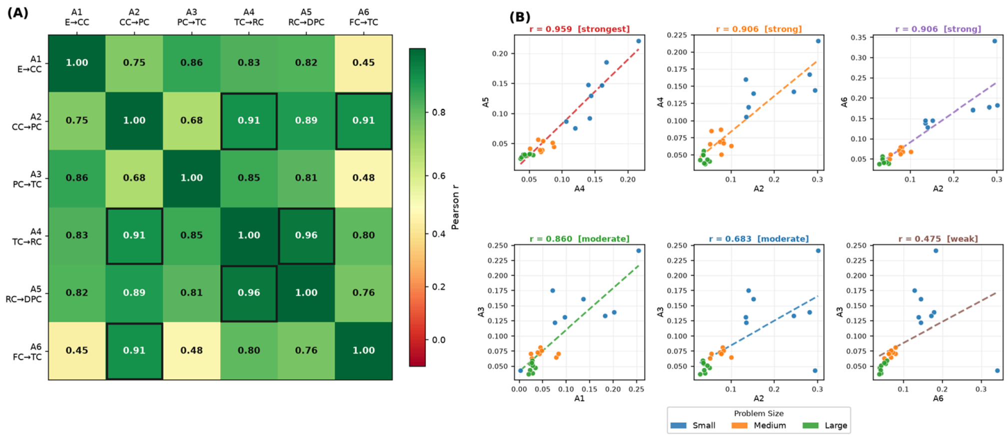

Figure 11. Arc FGW Cost Correlation Matrix and Selected Scatter Plots (α = 0.5, 24 instances). Panel (**A**): full Pearson r matrix with significance markers; boxes indicate r ≥ 0.90. Panel (**B**): six representative arc pairs from strongest (A4–A5, r = 0.959) to weakest (A6–A3, r = 0.475); points colored by problem size. Dashed lines indicate the fitted linear regression trend for each arc pair.

6. Conclusions, Policy Implications, and Future Research

6.1. Conclusions

This paper proposed a Fused Gromov–Wasserstein (FGW) framework for evaluating arc-level coupling efficiency in closed-loop supply chains (CLSCs). By embedding each CLSC node as a metric-measure space that encodes both spatial position and operational attributes, FGW provides a principled, decomposable measure of structural and attribute-level misalignment across the six arc interfaces of a filament manufacturing and PET bottle recovery network. Computational experiments on 24 problem instances across three scale classes yielded three robust findings.

First, the FGW bottleneck cost decreases sharply with network size (88% reduction from the highest-cost small to the lowest-cost large instance; Spearman $\varrho = -0.967$), confirming that larger CLSC networks achieve better coupling efficiency through richer matching opportunities. Second, the binding bottleneck arc shifts from the collection interface (A2: CC → PC) in small networks to the processing handoff (A3: PC→TC) in large networks, revealing a scale-dependent shift in the most frequent bottleneck arc that standard single-metric evaluations cannot detect. Third, attribute heterogeneity, not spatial misalignment, accounts for the majority of FGW cost across all arcs and instances (mean feature share = 57.9%), a finding that holds robustly across the full $\alpha \in [0, 1]$ frontier.

The framework has several limitations, detailed throughout this paper: validation relies on synthetic benchmark instances rather than operational data. The balanced OT formulation does not encode capacity imbalance as a hard constraint. Results are sensitive to feature engineering, weighting, and normalization choices. We emphasize that no comparison against nearest-facility assignment, cost-minimizing network flow, MILP-based, or multi-criteria assignment baselines has been conducted in this study, and such comparisons remain a priority for future work.

6.2. Policy Implications

To implement fully circular economy strategies, corporations must not only integrate sustainable goals in their planning but also develop the mindset and tools for cooperation with an evolving network of peers and third parties, considering the whole product supply chain while focusing on product recovery, reutilization, and remanufacturing. This network is evolving not only in size but, more importantly, in heterogeneity.

Closing the loop in supply chains generates a more complex process with many new risks, specifically due to the circular or multi-circular structure articulated in multi-echelons [40].

Risks are a decisive factor in shaping closed-loop supply chains, defined as the likelihood and impact of unexpected events at both macro and micro levels. The number of risks increases when companies decide to close the loop, driven by the uncertainty that the complex nature of CLSCs entails. Ref. [41] analyzed the risks of reverse logistics and prioritized them using hybrid multi-criteria decision making. A specific focus is on Green SCM, which includes a critical issue related to green supplier selection and categorization [42].

Uncertainty refers to situations that cannot be directly expressed. While describing various possibilities or scenarios, companies must understand what these risks are and how to deal with and hedge against them. A key challenge is how to integrate datasets from different sources: beyond the need to recognize their multi-source nature, it is critically important to acknowledge that data are inherently stochastic and that datasets live in different spaces. This generates the need for entirely new approaches to data representation and analysis. Traditional analytical tools introduce uncontrollable distortions and biases when applied to such complex structures as CLSCs. The relevance and pervasiveness of these risks are the driving factors for the introduction in this manuscript of a novel risk representation model—mm-spaces and probability distributions—and optimal transport-based tools such as Gromov–Wasserstein distance.

Focusing these considerations into direct actionable implications, we propose the following action points for designers and operations managers.

Attribute-first redesign. The analysis indicates that attribute standardization is likely to deliver greater efficiency gains than facility relocation in this setting. Since feature cost dominates in all six arcs (mean share 57.9%, reaching 62.0% in the aggregate bottleneck arc A2), harmonizing processing capacities, recovery yields, and cost structures across echelons delivers greater efficiency gains than spatial optimization of facility locations. This challenges the conventional focus on transportation cost minimization in reverse logistics.

Small networks: focus on A2 and A6. For small-scale CLSC operators, the collection-center interface (A2: CC $\rightarrow$ PC) and the reverse waste feed-in arc (A6: DPC/FC $\rightarrow$ TC) are the highest-priority redesign targets. Investment in standardizing collection-center capacity profiles or improving the consistency of reverse material quality will yield the largest FGW cost reductions at this scale.

Large networks: focus on A3. As networks scale, the processing-to-treatment handoff (A3: PC $\rightarrow$ TC) becomes the dominant bottleneck. Large-scale operators should invest in capacity alignment and material-flow coordination protocols between processing and treatment centers, where the infrastructure gap is also most persistent, present in 23 of 24 instances.

Results suggest limited returns from investment in A1 under the present benchmark. The end-user to collection-center arc (A1: E $\rightarrow$ CC) is never the binding constraint across any of the 24 tested instances. Managerial resources allocated to optimizing A1, such as collection-point density campaigns, should be deprioritized relative to mid-chain interventions.

α as a strategic tuning parameter. The α parameter can be used as a managerial dial: setting α close to 0 emphasizes attribute fit, relevant when standardization is the strategic priority, while α close to 1 captures spatial efficiency, relevant when facility location is under review. The $\alpha = 0.5$ reference provides a balanced assessment suitable for routine network audits, and Section Sensitivity Analysis and Justification of $\alpha = 0.5$ demonstrates that the bottleneck identification is stable across $\alpha \in [0.0, 0.7]$, providing confidence that the diagnostic is not sensitive to the precise choice of α.

6.3. Directions for Future Research

This study opens several promising avenues for future investigation. The framework has been validated only on synthetic benchmark instances derived from a prior MILP study [2]; no operational data from an active closed-loop supply chain has been used. The reported scale effect, bottleneck migration pattern, and feature-dominance finding should be read as properties of this benchmark construction rather than established empirical facts about real CLSC networks.

A full empirical comparison against nearest-facility assignment, cost-minimizing network flow, MILP-based CLSC models, and multi-criteria assignment baselines would further strengthen the empirical case for the framework and is identified as a priority direction for future benchmarking.

Dynamic CLSCs. Extending the framework to dynamic, multi-period CLSCs where node attributes and spatial positions evolve over time would capture the temporal dimension of network coupling efficiency, particularly relevant for circular economy transitions where facility capacities and material recovery rates change as markets mature.

General α-selection procedure. The $\alpha = 0.5$ reference established here is specific to the present benchmark network. Developing a systematic procedure for selecting α across different CLSC problem classes, via scenario-based or stakeholder-preference elicitation, remains an open question for future work.

Joint multi-arc coupling. The current framework evaluates each arc independently. Future work could develop a joint multi-arc FGW formulation that captures inter-arc dependencies and propagates upstream misalignment to downstream arcs, better reflecting the sequential nature of material flow in circular networks.

Uncertainty and robustness. The FGW framework could be integrated with stochastic programming or distributionally robust optimization to account for demand uncertainty, return-rate variability, and disruption scenarios, a critical gap for real-world CLSC resilience planning. All network parameters in the present study are treated as deterministic. Extending the FGW formulation to incorporate input uncertainty, e.g., via stochastic programming or distributionally robust optimal transport over uncertain feature vectors and distance matrices, is a natural next step toward operational deployment under real-world variability.

Latent structure analysis. The arc-to-arc correlation analysis in Section 5.6 reveals a differentiated dependency structure, with A4–A5 most tightly coupled ($r = 0.959$) and A6 largely independent of the forward processing arcs ($r = 0.446$ with A1, $r = 0.475$ with A3). Structural equation modelling or factor analysis could decompose these correlations and identify root-cause drivers of co-movement across the forward and reverse sub-networks.

Empirical validation. Validating the framework on real-world CLSC data—for instance, from electronics take-back programs or automotive remanufacturing networks—would test the generalizability of the scale effect and bottleneck migration findings reported here, and establish whether the 57.9% feature dominance result holds beyond synthetic benchmark instances.

The static formulation presented here does not capture disruption scenarios arising from geopolitical instability or armed conflict, which represent an increasingly important stressor on global supply and transport networks [43,44]. Extending the FGW framework to incorporate disruption-aware or scenario-based formulations, informed by this growing body of conflict-related transport and logistics literature, is a relevant direction for future work.

Author Contributions: Conceptualization, I.S., A.C. and F.A.; methodology, A.C. and F.A.; software, I.S.; validation, I.S., A.C. and F.A.; formal analysis, I.S., A.C. and F.A.; investigation, I.S. and A.C.; resources, I.S.; data curation, I.S. and A.C.; writing—original draft, I.S. and F.A.; writing—review and editing, I.S., A.C. and F.A.; visualization, I.S.; supervision, A.C. and F.A.; project administration, A.C. and F.A.; funding acquisition, F.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The data presented in this study are openly available at https://github.com/iman-ie/fgw-clsc-network (accessed on 25 May 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Metric Measure Spaces

Appendix A.1. Metric Measure Spaces (mm-Spaces)

In the discrete graph setting, mm-spaces provide a natural representation of weighted graphs with node distributions. Specifically, a graph $G = (V, E, w)$ with node set $V = \{v_1, \ldots, v_n\}$, edge set E, and edge weights $w : E \to \mathbb{R}_+$ can be encoded as an mm-space as follows:

Space X: The node set V serves as the underlying space. Each node represents a point in the discrete metric space.

Metric d: The metric structure is induced by the graph connectivity. Common choices include:

Shortest-path distance: $d(v_i, v_j)$ equals the length of the shortest path between nodes v_i and v_j in the graph, where edge lengths are determined by weights or are uniform (unweighted case).

Resistance distance: Based on the effective resistance between nodes when the graph is viewed as an electrical network [45,46].

Diffusion distance: Measures similarity based on how heat diffuses between nodes via the graph Laplacian [47,48].

For weighted graphs, the shortest-path metric is typically computed using algorithms such as Dijkstra's or Floyd-Warshall, yielding a symmetric distance matrix $C \in \mathbb{R}^{n \times n}$ where $C_{ij} = d(v_i, v_j)$.

Measure μ: The probability measure μ assigns a mass to each node, encoding the importance or distribution over the graph. In the discrete case, the measure has the form $\mu = \sum_{i=1}^{n} \mu_i \, \delta_{v_i}, \mu_i \geq 0, \sum_{i=1}^{n} \mu_i \ = 1$.

Thus, $\mu_i = \mu(\{v_i\})$ represents the probability mass associated with the node v_i. Common choices include:

Uniform distribution: $\mu_i = \frac{1}{n}$ for all i, treating all nodes equally.

Degree-based distribution: $\mu_i \propto deg(v_i)$, where $deg(v_i)$ is the degree of the node v_i, emphasizing well-connected nodes.

Feature-based distribution: μ_i derived from node attributes or importance scores (e.g., PageRank, centrality measures).

Thus, a graph G with n nodes is represented as the mm-space (V, C, μ), where V is the node set, C is the $n \times n$ distance matrix encoding graph connectivity (e.g., shortest-path distances between nodes), and μ is the probability vector over nodes representing node importance or mass distribution. This representation enables the application of optimal transport theory to graph comparison. The matrix C captures the internal structure (topology) of the graph—how nodes are connected and how far apart they are within the graph. In contrast, if nodes carry feature vectors (e.g., node attributes like age, income, or embeddings in $\mathbb{R}^{d}$), these are handled separately as node features ψ_X, leading to attributed networks (Section 2.4).

The Wasserstein distance between two mm-spaces (V_s, C_s, μ_s) and (V_t, C_t, μ_t) quantifies how much "effort" is required to transform one graph's node distribution into the other's, accounting for both the graph geometry (via C_s, C_t, which encode connectivity patterns) and node masses (via μ_s, μ_t, which encode importance distributions). This framework compares graphs based on their structural properties rather than node features.

To illustrate the mm-space framework concretely, Figure A1 presents a complete decomposition of a simple 8-node graph into its three fundamental components. Panel (A) shows the graph structure with nodes $V = \{v_1, \ldots, v_8\}$ connected by edges, where node sizes visually encode the probability mass assigned to each node. Panel (B) displays the metric structure through the shortest-path distance matrix C, revealing the pairwise geodesic distances between all nodes—for instance, nodes v_1 and v_8 are separated by a distance of 4, requiring traversal of 4 edges along the shortest path. Panel (C) presents the probability measure μ as a discrete distribution $\sum_{i=1}^{8} \mu_i = 1$. In this example, the measure is non-uniform, with higher mass concentrated on central nodes (v_2 and v_3 receive $\mu_2 = 0.20$ and $\mu_3 = 0.25$, respectively) while peripheral nodes receive less mass. This triple (V, C, μ) completely specifies the mm-space representation of the graph, providing the foundation for computing optimal transport distances between graphs with potentially different structures and distributions.

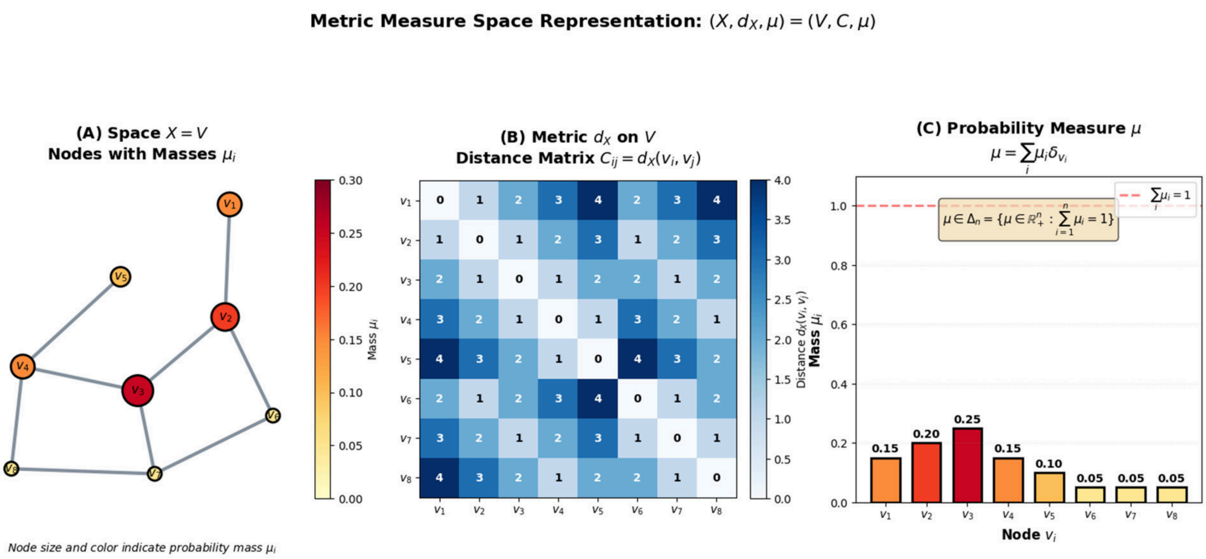

Figure A1. Mm-space (V, C, μ) representation of graphs. (**A**) Node set V with probability masses p_i (indicated by size and color). (**B**) Shortest-path distance matrix C defining the metric structure. (**C**) Probability distribution p forming a valid measure $(\Sigma\mu_i = 1)$. Together, these components encode both graph topology and node importance, enabling optimal transport-based graph comparison via Gromov–Wasserstein distances.

Appendix A.2. Measure Networks (m-Net)

The concept of an m-net generalizes that of an mm-space by relaxing the requirement that the pairwise interaction function be a distance metric. This broadens the applicability of the GW framework to complex relational data structures. A m-net space is a triple $\mathcal{X} = (X, \omega_X, \mu)$ where:

X is a separable and completely metrizable topological space (a Polish space);

μ_X is a fully supported Borel probability measure on X;

$\omega_X : X \times X \to R$ is a measurable function, called the network function.

The collection of all m-net spaces is denoted by $\mathcal{N}$. When ω_X satisfies the metric axioms and generates the topology of X, it is denoted d_X, and the corresponding triple (X, d_X, μ) is a mm-space belonging to the class $\mathcal{M}$. Figure A2 visualizes a m-net space, generalizing the concept of an mm-space.

Nodes represent elements of the set X.

Node sizes indicate the probability measure μ_X, highlighting regions of higher or lower importance.

Edges represent the network function ω_X, which may encode asymmetric or non-metric interactions between points.

Edge thickness corresponds to the magnitude of ω_X, making stronger interactions visually prominent.

Unlike mm-spaces, the pairwise interactions in m-net-spaces are not constrained by distance axioms, allowing this framework to capture complex relational structures such as weighted graphs or networks. Every mm-space is a special case of an m-net-space where ω_X satisfies metric axioms.

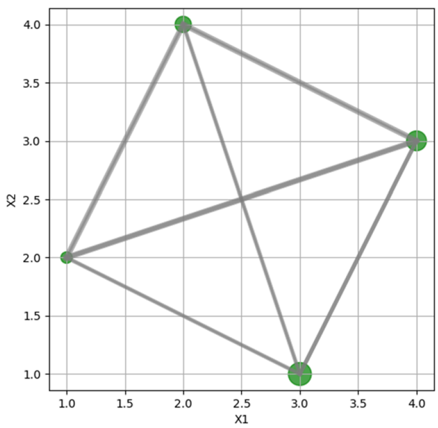

Figure A2. Illustration of an m-net space (X, ω_X, μ), where node sizes encode the probability measure μ over elements of X, and edge thickness represents the strength of pairwise interactions $\omega_X\left(x_i, x_j\right)$, which are not required to satisfy metric axioms.

M-nets include numerous important structures [49]:

Pseudo-mm-spaces.

A pseudo-metric $\omega_X : X \times X \to R$ satisfies all the metric axioms besides allowing that $\omega_X(x, x') = 0$ for $x \neq x'$, may occur for distinct points. Pseudo-mm-spaces arise naturally as elements in the completion of $\mathcal{M}$ under Gromov–Wasserstein distances [50].

Graphs as m-nets.

A finite graph $G = (V, E)$ can be viewed as an m-net by setting $X = V$, letting μ_X be the uniform measure, and defining:

$$\omega_X(x, x') = \begin{cases} 1 & if\{x, x'\} \in E \\ 0 & otherwise \end{cases} \tag{A1}$$

This provides a discrete example of an m-net [51].

Generalized graph kernels.

More generally, one can define ω_X through graph-based similarity measures. A prominent example is the heat kernel (diffusion kernel) at time t: [52].

$$\omega_X(x, x') = [exp(-tL)]_{xx'} \tag{A2}$$

where L is the graph Laplacian and $exp(-tL)$ is the matrix exponential. This kernel measures the amount of heat diffused from node x to node x' after time t, encoding both local and global graph structure [53,54].

The GW framework applies naturally to these generalized spaces. Indeed, it has become a key concept in the analysis of graphs and complex data representable as networks or relational structures, effectively extending optimal transport to settings where distributions cannot be compared pointwise [24].

Attributed Network in m-nets spaces.

The analysis of network structures can be categorized based on the types of data associated with their nodes and edges. A basic representation is a graph with edge weights and node weights, visualized through size variations, which is encoded as an m-net and compared using the GW distance [23]. A more complex structure involves a weighted graph with additional node features, where each node is assigned a vector in $\mathbb{R}^n$ (visualized as a column vector), allowing for comparison via the Fused Gromov–Wasserstein (FGW) distance [6]. Furthermore, graphs can be enriched with edge features, assigning a point in a fixed metric space Z to each edge—such as a 1-dimensional probability distribution or time-varying distance in this case—modeled as Z-networks, allowing comparison through a proposed distance tailored to the chosen target space Z.

Appendix B. Review of the Computational Issues in Gromov–Wasserstein

This section provides an overview of existing GW solvers, both unregularized and entropic approach algorithms, and is designed for GW extensions such as Fused [6] and Unbalanced [38]. Since GW is closely related to the quadratic assignment problem (QAP), which is known to be NP-hard [55]. Finding a global minimum is computationally expensive and theoretically intractable given the non-convexity of the objective function. Recent works have proposed relaxations to make this global search tractable: ref. [56] formulate GW as a generalized moment problem that can be truncated at the desired degree of precision, while ref. [57] propose a semi-definite relaxation solvable in polynomial time. Alternatively, ref. [58] exploit the structure of squared Euclidean costs to design a cutting plane algorithm that efficiently explores the set of admissible solutions. Despite these advances, global methods remain slow and are limited to inputs containing at most a few hundred points.

Many practical applications only require a quantification of similarity between α and β rather than an exact matching. In such cases, cheaper alternatives to the full GW problem may suffice. This category includes SLICED GW [59], QUANTIZED-GW [60], MINIBATCH-GW [61] and LINEAR-GW [62], as well as GW-inspired distances for Gaussian mixture models [63]. While these heuristics can be computed in quadratic or even linear time, they rely on heavy approximations and are less suitable for tasks requiring precise point-wise correspondences, such as shape matching.

Exact Entropic Solvers. Entropic regularization, introduced by [26,64], allows GW to be solved using the efficient Sinkhorn algorithm. The standard solver, ENTROPIC-GW, is a Projected Gradient Descent (PGD) in the KL-divergence geometry. When costs are squared norms, ref. [65] further reduces this computation to $O(NMD)$. Similarly, the dual solver DUAL-GW of [66] achieves quadratic complexity and is guaranteed to converge.

Accelerated Approximations. Apart from the squared norm case, solving EGW remains expensive. The algorithm complexity can still be improved thanks to several approximation strategies introduced in the last few years. SAMPLED-GW [67] and SPARSE-GW [68] apply a sparsity mask to fasten both cost matrix computations and Sinkhorn iterations. LOWRANK-GW [65] approximates the input cost matrices via low-rank factorization. These three algorithms reach a quadratic running time, improving significantly the scalability of the original EGW solvers. With LINEARLOWRANK-GW [65] combine the low-rank cost reduction with a low-rank constraint on the transport plan, further reducing computation time to $O(N+M)$. Finally, ULOT (Unsupervised Learning of Optimal Transport) [69] trains neural networks to predict optimal couplings, removing the optimization burden at inference time—although its training phase keeps a cubic complexity.

Appendix C. Benchmark Instances and FGW Results

Table A1 reports the full set of 24 benchmark problem instances originally introduced by [2] for a circular CLSC network combining 3D-printing filament manufacturing and post-consumer PET bottle recovery. The instances span three scale classes—small (1–8), medium (9–16), and large (17–24)—and are parameterized according to the node counts and cost distributions reported in that study (see Table 2 therein). For each instance, the FGW cost at $\alpha = 0.5$ is reported for each of the six arc interfaces. Red-shaded cells indicate the binding bottleneck arc per instance.

Table A1. Benchmark dimensions and FGW cost ($\alpha = 0.5$) per arc.

#	Sz	DPC	TC	RC	PC	CC	FC	A1	A2	A3	A4	A5	A6
1	S	2	2	2	2	2	3	0.0027	0.2942	0.0428	0.1441	0.1298	0.3414
2	S	2	3	5	4	3	8	0.2016	0.2816	0.1394	0.1674	0.1854	0.1781
3	S	3	3	6	5	3	12	0.2544	0.3009	0.2412	0.2163	0.2210	0.1827
4	S	4	4	8	6	4	16	0.1361	0.1518	0.1609	0.1395	0.1476	0.1450
5	S	5	4	10	7	5	19	0.1825	0.2447	0.1334	0.1420	0.0927	0.1712
6	S	6	5	10	8	6	22	0.0969	0.1343	0.1311	0.1603	0.1467	0.1393
7	S	8	5	12	9	7	28	0.0707	0.1397	0.1752	0.1199	0.0762	0.1285
8	S	10	6	12	10	8	33	0.0754	0.1347	0.1220	0.1058	0.0875	0.1450
9	M	12	10	18	13	10	50	0.0784	0.1010	0.0652	0.0636	0.0568	0.0680
10	M	14	11	20	14	12	55	0.0451	0.0803	0.0814	0.0711	0.0548	0.0798
11	M	16	12	20	15	14	60	0.0829	0.0767	0.0706	0.0872	0.0452	0.0794
12	M	18	13	22	17	15	65	0.0431	0.0541	0.0707	0.0851	0.0513	0.0620
13	M	20	14	22	19	17	70	0.0388	0.0744	0.0730	0.0694	0.0399	0.0719
14	M	22	15	24	20	18	75	0.0477	0.0824	0.0765	0.0673	0.0367	0.0683
15	M	25	16	24	22	19	79	0.0263	0.0515	0.0632	0.0662	0.0394	0.0500
16	M	30	17	26	25	20	83	0.0243	0.0776	0.0710	0.0507	0.0419	0.0624
17	L	35	20	34	30	25	103	0.0345	0.0501	0.0475	0.0409	0.0323	0.0386
18	L	38	21	35	32	28	107	0.0233	0.0363	0.0543	0.0560	0.0313	0.0535
19	L	40	22	37	33	29	111	0.0272	0.0349	0.0598	0.0509	0.0337	0.0567
20	L	45	24	39	35	30	115	0.0281	0.0446	0.0516	0.0438	0.0298	0.0421
21	L	50	25	40	37	32	119	0.0283	0.0399	0.0387	0.0367	0.0250	0.0419
22	L	53	28	42	40	33	123	0.0264	0.0352	0.0555	0.0504	0.0328	0.0505
23	L	58	29	42	42	34	127	0.0223	0.0382	0.0436	0.0381	0.0278	0.0414
24	L	60	30	43	43	35	131	0.0201	0.0280	0.0374	0.0429	0.0319	0.0376

Notes: DPC = 3D printing centres; TC = treatment centres; RC = recovery centres; PC = processing centres; CC = collection centres; FC = filament customers. S = Small (1–8); M = Medium (9–16); L = Large (17–24). Arc interfaces: A1 (E $\rightarrow$ CC), A2 (CC $\rightarrow$ PC), A3 (PC $\rightarrow$ TC), A4 (TC $\rightarrow$ RC), A5 (RC $\rightarrow$ DPC), A6 (DPC/FC $\rightarrow$ TC). Red-shaded entries indicate the binding bottleneck arc. $\alpha = 0.5$ for all entries.

Facility counts, costs, capacities, recovery efficiencies, and CO_2 intensities are generated by sampling from parameter ranges informed by [2]. The small, medium, and large size classes correspond to increasing facility counts across all eight echelons, as shown in Table A1.

Appendix D. Supplementary Results for Section 5

Appendix D.1. Full Per-Instance FGW Cost Results (α = 0.5)

Table A2 reports the complete FGW cost for each arc across all 24 instances at α = 0.5. Bottleneck values are highlighted in red bold. Row shading distinguishes size groups: orange = Small, green = Medium, blue = Large. Arc A1 never appears as a bottleneck.

Table A2. FGW cost per arc for all 24 instances (α = 0.5). Red bold = bottleneck value.

#	Size	B.Arc	B.FGW	A1	A2	A3	A4	A5	A6
1	Small	A6	**0.3414**	0.0027	0.2942	0.0428	0.1441	0.1298	**0.3414**
2	Small	A2	**0.2815**	0.2016	**0.2815**	0.1394	0.1674	0.1854	0.1781
3	Small	A2	**0.3009**	0.2544	**0.3009**	0.2412	0.2163	0.2210	0.1827
4	Small	A3	**0.1609**	0.1361	0.1518	**0.1609**	0.1395	0.1476	0.1450
5	Small	A2	**0.2447**	0.1825	**0.2447**	0.1334	0.1420	0.0927	0.1712
6	Small	A4	**0.1603**	0.0969	0.1343	0.1311	**0.1603**	0.1467	0.1393
7	Small	A3	**0.1752**	0.0707	0.1397	**0.1752**	0.1199	0.0762	0.1285
8	Small	A6	**0.1450**	0.0754	0.1347	0.1220	0.1058	0.0875	**0.1450**
9	Medium	A2	**0.1010**	0.0784	**0.1010**	0.0652	0.0636	0.0568	0.0680
10	Medium	A3	**0.0814**	0.0451	0.0803	**0.0814**	0.0711	0.0548	0.0798
11	Medium	A4	**0.0872**	0.0829	0.0767	0.0706	**0.0872**	0.0452	0.0794
12	Medium	A4	**0.0851**	0.0431	0.0541	0.0707	**0.0851**	0.0513	0.0620
13	Medium	A2	**0.0744**	0.0388	**0.0744**	0.0730	0.0694	0.0399	0.0719
14	Medium	A2	**0.0824**	0.0477	**0.0824**	0.0765	0.0673	0.0367	0.0683
15	Medium	A4	**0.0662**	0.0263	0.0515	0.0632	**0.0662**	0.0394	0.0500
16	Medium	A2	**0.0776**	0.0243	**0.0776**	0.0710	0.0507	0.0419	0.0624
17	Large	A2	**0.0501**	0.0345	**0.0501**	0.0475	0.0409	0.0323	0.0386
18	Large	A4	**0.0560**	0.0233	0.0363	0.0543	**0.0560**	0.0313	0.0535
19	Large	A3	**0.0598**	0.0272	0.0349	**0.0598**	0.0509	0.0337	0.0567
20	Large	A3	**0.0516**	0.0281	0.0446	**0.0516**	0.0438	0.0298	0.0421
21	Large	A6	**0.0419**	0.0283	0.0399	0.0387	0.0367	0.0250	**0.0419**
22	Large	A3	**0.0555**	0.0264	0.0352	**0.0555**	0.0504	0.0328	0.0505
23	Large	A3	**0.0436**	0.0223	0.0382	**0.0436**	0.0381	0.0278	0.0414
24	Large	A4	**0.0429**	0.0201	0.0280	0.0374	**0.0429**	0.0319	0.0376

Appendix D.2. Feature Cost Share per Arc (α = 0.5)

Table A3 reports feature cost share (%) per arc and instance. Cells with feature cost share $\geq$ 65%: attribute-dominated. Cells with feature cost share $\leq$ 40%: structure-dominated.

Table A3. Feature cost share (%) per arc for all 24 instances (α = 0.5).

#	Size	A1 (%)	A2 (%)	A3 (%)	A4 (%)	A5 (%)	A6 (%)
1	Small	33.4	90.0	34.0	90.4	83.7	65.4
2	Small	46.3	77.4	34.3	73.0	70.5	68.2
3	Small	76.0	60.9	68.5	61.2	50.8	60.2
4	Small	43.5	36.1	75.8	54.6	38.2	60.2
5	Small	70.4	69.3	51.8	49.0	33.6	59.0
6	Small	72.7	45.6	34.7	52.5	70.1	40.6
7	Small	70.4	74.3	66.9	65.4	52.3	57.9
8	Small	65.1	59.8	68.2	46.9	41.4	40.5
9	Medium	29.4	45.6	58.0	62.1	70.2	54.7
10	Medium	60.6	49.3	64.4	59.0	53.0	46.3
11	Medium	61.0	52.8	49.4	52.8	58.4	53.1

https://doi.org/10.3390/su18136906

Table A3. *Cont.*

#	Size	A1 (%)	A2 (%)	A3 (%)	A4 (%)	A5 (%)	A6 (%)
12	Medium	56.6	61.8	72.5	72.5	73.0	65.2
13	Medium	60.3	63.0	62.2	68.7	52.4	57.5
14	Medium	48.8	65.9	53.8	44.8	73.3	45.6
15	Medium	44.9	63.9	58.5	71.5	58.7	53.0
16	Medium	42.3	56.2	66.8	56.2	68.5	57.9
17	Large	44.1	67.0	45.3	57.1	55.8	53.7
18	Large	45.2	64.9	62.9	60.7	47.3	59.2
19	Large	58.1	67.8	50.6	45.4	51.9	48.0
20	Large	57.9	73.2	66.0	64.0	60.7	63.5
21	Large	62.4	60.7	59.7	56.5	62.9	72.3
22	Large	47.8	58.2	43.1	38.2	53.9	41.1
23	Large	43.2	67.3	65.3	53.2	51.9	64.5
24	Large	53.2	57.6	73.5	66.9	55.0	75.1

Appendix D.3. Supplementary Figures

Figure A3 shows mean FGW cost per arc by size group. A2 dominates small/medium; A3 rises at large scale. A1 is consistently lowest across all sizes.

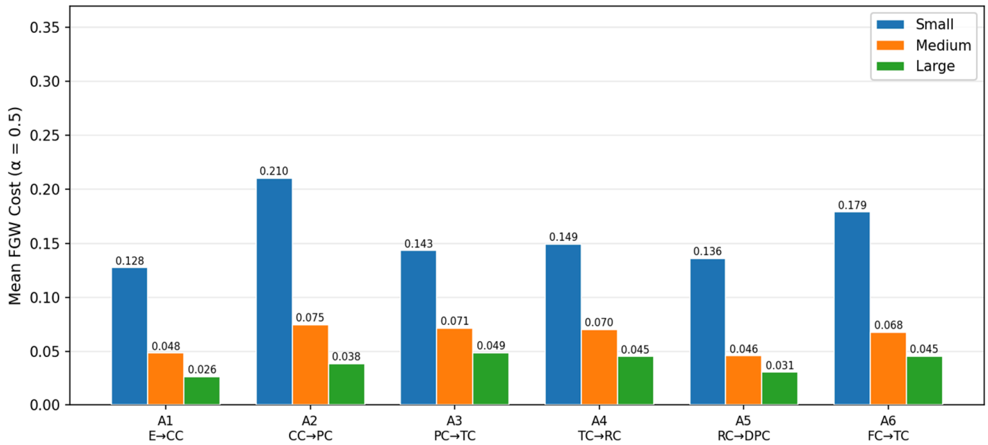

Figure A3. Mean arc FGW cost by problem size group (α = 0.5).

Figure A4 (left) plots bottleneck cost vs. N, confirming Spearman $\varrho = -0.967$. The right panel proves A1 lies strictly below the bottleneck y = x line for every instance.

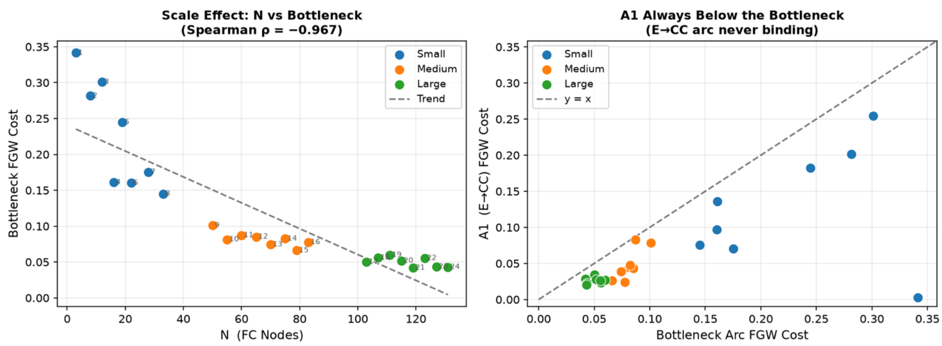

Figure A4. (**Left**): Scale effect (N vs. bottleneck FGW). (**Right**): A1 is never the bottleneck arc.

Figure A5 shows the feature cost share distribution per arc across all 24 instances. All arcs have median feature share above 50%, confirming attribute-heterogeneity dominance. The 50% threshold line (dashed red) is exceeded by the median in every arc.

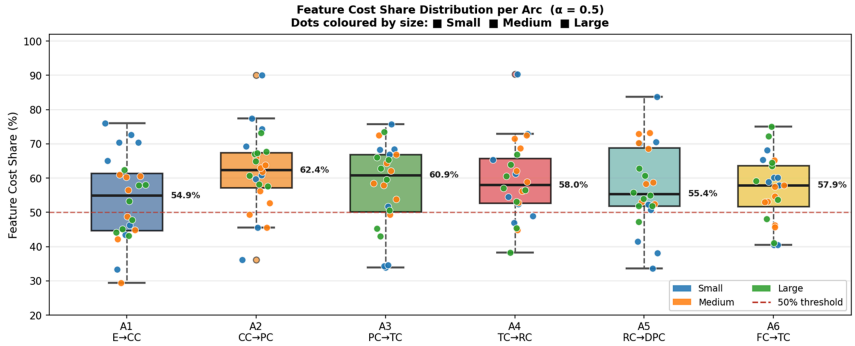

Figure A5. Feature cost share (%) per arc—all 24 instances. Red dashed line = 50% threshold. Dots colored by problem size.

Figure A6 (left) shows the FGW cost frontier per arc as α varies from 0 to 1, illustrating how structural cost rises gradually. The right panel decomposes A2 cost into feature and structural components, showing the crossover near α = 0.7.

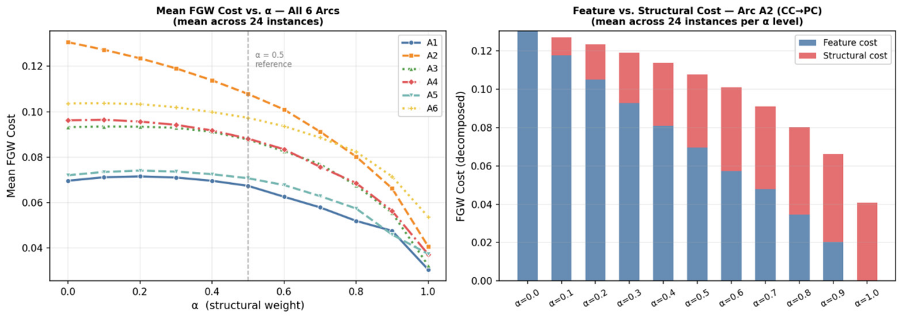

Figure A6. α-Sensitivity Analysis: FGW cost frontier per arc (**left**) and feature/structural decomposition for A2 (**right**).

Figure A7 presents the Pearson correlation matrix of arc FGW costs across 24 instances. Mid-chain arcs A2, A3, A4 are strongly correlated (r ≥ 0.85), while A1 and A5 exhibit weaker correlations with the bottleneck group (r < 0.55).

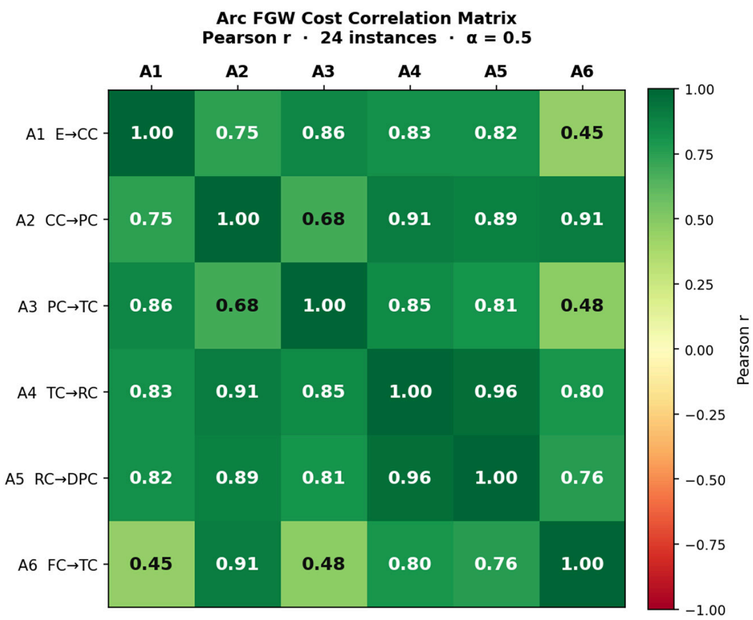

Figure A7. Pearson correlation matrix of arc FGW costs across 24 instances (α = 0.5).

Appendix D.4. Weight Sensitivity Analysis

To validate the arc-level sustainability weight vectors specified in Section 4.4, a sensitivity analysis was conducted across four weighting schemes applied to all 24 problem instances (8 small, 8 medium, 8 large):

W0 (Baseline): paper weights as defined in Section 4.4.

W1 (Equal): uniform weights across all arc attributes.

W2 (Cost-Heavy): cost attributes set to 0.60, remainder shared equally.

W3 (Sustainability-Heavy): CO_2 and recovery attributes set to 0.80 combined.

Figure A8 reports the mean bottleneck FGW cost per scheme. W3 achieves the lowest mean (0.102), followed by W2 (0.110) and W0 (0.119). However, a lower mean bottleneck FGW under W3 reflects cost compression rather than improved matching quality: concentrating 80% of weight on CO_2 and recovery attributes suppresses arc-level differentiation, causing 13 of 24 bottlenecks to collapse onto A4 alone (Table A4). W0 achieves a higher mean bottleneck FGW precisely because it preserves meaningful cost variation across arcs, which is the intended behaviour of a feature-cost-driven OT formulation.

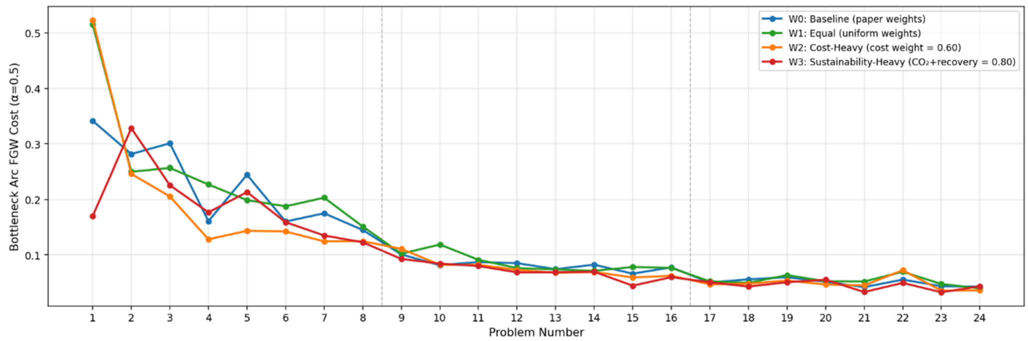

Figure A8. Bottleneck FGW cost across 24 instances (weight sensitivity).

Figure A9 shows the per-arc feature cost share, i.e., the fraction of total FGW cost attributable to feature attributes. Under W0, the overall mean feature cost share is 0.579, indicating that node attributes drive approximately 58% of transport cost across all arcs. This is higher than the sustainability-heavy scheme W3 (0.520), which paradoxically reduces feature influence by concentrating mass on fewer attribute dimensions. Arcs A2 and A4 consistently exhibit the highest feature cost share ($\geq$0.59) under W0, consistent with their richer attribute sets (3 features per echelon).

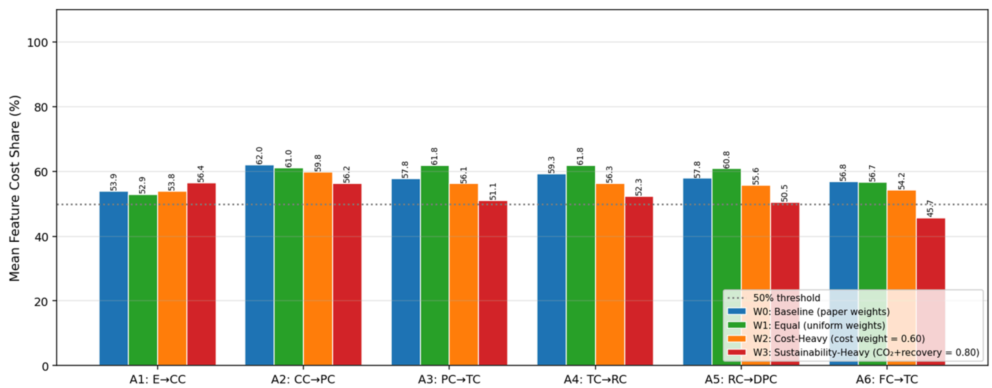

Figure A9. Mean feature cost share per arc, weight sensitivity (α = 0.5, 24 instances).

Figure A10 displays the bottleneck arc frequency heatmap across all 24 instances. Under W0, bottlenecks concentrate on arcs A2, A3, and A4, the intermediate processing arcs with the greatest operational heterogeneity, while A1 and A5 record zero bottlenecks. This distribution is theoretically consistent: end-user (A1) and final-stage (A5) arcs have lower node-level variability, and the paper weights correctly avoid inflating their transport cost. By contrast, the cost-heavy scheme W2 shifts 9 of 24 bottlenecks to A6 (reverse logistics arc), an artifact of over-weighting cost on a low-variability arc. Table A4 summarizes bottleneck frequencies across schemes.

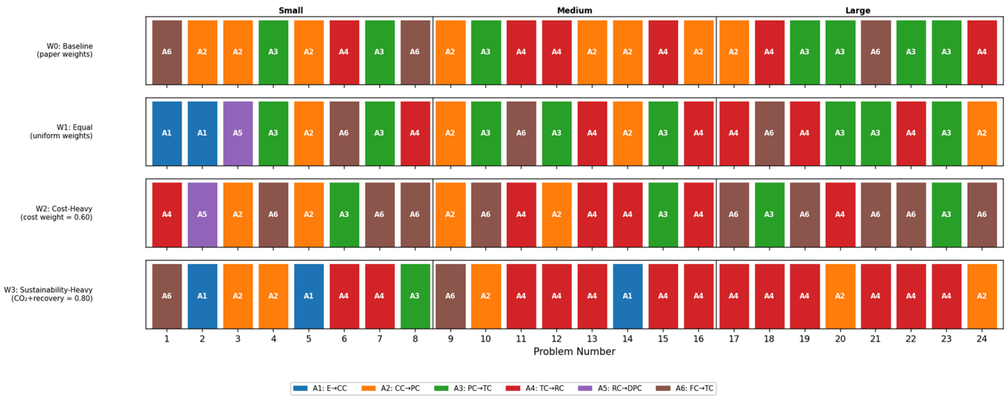

Figure A10. Bottleneck arc identity by weight scheme (α = 0.5).

Table A4. Bottleneck arc frequency and mean bottleneck FGW cost by weighting scheme (24 instances).

Scheme	A1	A2	A3	A4	A5	A6	Mean BN FGW
W0 Baseline	0	8	7	6	0	3	0.119
W1 Equal	2	4	8	6	1	3	0.129
W2 Cost-Heavy	0	4	4	6	1	9	0.110
W3 Sustainability	3	5	1	13	0	2	0.102

Collectively, these results confirm that the baseline weights (W0) achieve a well-calibrated balance: bottleneck concentration on operationally variable arcs, stable feature cost contribution (~58%), and lower mean transport cost than uniform weighting. The sustainability-heavy scheme W3 confirms that further shifting weight towards CO_2 and recovery attributes does not improve transport efficiency and reduces arc-level feature influence, supporting the paper's choice not to adopt extreme sustainability weights.

References

1. Akbari-Aghghaleh, Z.; Mozdgir, A.; Seyedi, I.; Messina, E. Designing a perishable closed-loop poultry supply chain: Metaheuristic approaches and model evaluation. *Environ. Dev. Sustain.* **2025**, 1–43. [CrossRef]
2. Rajabi-Kafshgar, A.; Seyedi, I.; Tirkolaee, E.B. Circular closed-loop supply chain network design considering 3D printing and PET bottle waste. *Environ. Dev. Sustain.* **2024**, *27*, 20345–20381. [CrossRef]
3. Villani, C. Optimal Transport. In *Grundlehren der Mathematischen Wissenschaften*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 338. [CrossRef]
4. Peyré, G.; Cuturi, M. Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **2019**, *11*, 355–607. [CrossRef]
5. Mémoli, F. A spectral notion of Gromov–Wasserstein distance and related methods. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 363–401. [CrossRef]
6. Vayer, T.; Chapel, L.; Flamary, R.; Tavenard, R.; Courty, N. Fused Gromov-Wasserstein Distance for Structured Objects. *Algorithms* **2019**, *13*, 212. [CrossRef]
7. Mishra, A.; Dutta, P.; Jayasankar, S.; Jain, P.; Mathiyazhagan, K. A review of reverse logistics and closed-loop supply chains in the perspective of circular economy. *Benchmarking Int. J.* **2023**, *30*, 975–1020.
8. Seyedi, I.; Hamedi, M.; Tavakkoli-Moghaddam, R. Optimization for a truck scheduling problem in multi-door cross dokcing with learning effect and deteriorating jobs. *J. Transp. Res.* **2022**, *19*, 183–206. [CrossRef]
9. Shokouhifar, M.; Naderi, R.; Goli, A.; Gultom, P.; Nikabadi, M.S.; Weber, G.-W. Metaheuristic-driven extended exergy accounting for sustainable closed-loop food supply chain management. *Comput. Ind. Eng.* **2024**, *191*, 110148. [CrossRef]

10. Devika, K.; Jafarian, A.; Nourbakhsh, V. Designing a sustainable closed-loop supply chain network based on triple bottom line approach: A comparison of metaheuristics hybridization techniques. *Eur. J. Oper. Res.* **2014**, *235*, 594–615. [CrossRef]
11. Esmaeilian, S.; Mohamadi, D.; Esmaelian, M.; Ebrahimpour, M. A multi-objective model for sustainable closed-loop supply chain of perishable products under two carbon emission regulations. *J. Model. Manag.* **2023**, *18*, 285–317.
12. Govindan, K.; Soleimani, H. A review of reverse logistics and closed-loop supply chains. *J. Clean. Prod.* **2017**, *142*, 371–384. [CrossRef]
13. Monge, G. Mémoire sur le calcul intégral des équations aux différences partielles. *Mém. Acad. R. Sci. Paris* **1787**, *1784*, 118–192.
14. Kantorovich, L. On the transfer of masses (in Russian). In *Doklady Akademii Nauk*; Russian Academy of Sciences (RAS): Moscow, Russia, 1942; p. 227. Available online: https://cir.nii.ac.jp/crid/1370565168575910170 (accessed on 17 July 2025).
15. Cuturi, M.; Doucet, A. Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2014; pp. 685–693. Available online: https://proceedings.mlr.press/v32/cuturi14.html (accessed on 9 June 2025).
16. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2017; pp. 214–223. Available online: https://proceedings.mlr.press/v70/arjovsky17a.html (accessed on 17 July 2025).
17. Seyedi, I.; Candelieri, A.; Archetti, F. Distributionally Robust Bayesian Optimization via Sinkhorn-Based Wasserstein Barycenter. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 90. [CrossRef]
18. Solomon, J.; de Goes, F.; Peyré, G.; Cuturi, M.; Butscher, A.; Nguyen, A.; Du, T.; Guibas, L. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.* **2015**, *34*, 1–11. [CrossRef]
19. Flamary, R.; Courty, N.; Gramfort, A.; Alaya, M.Z.; Boisbunon, A.; Chambon, S.; Chapel, L.; Corenflos, A.; Fatras, K.; Fournier, N.; et al. Pot: Python optimal transport. *J. Mach. Learn. Res.* **2021**, *22*, 1–8.
20. Carlsson, J.G.; Behroozi, M.; Mihic, K. Wasserstein Distance and the Distributionally Robust TSP. *Oper. Res.* **2018**, *66*, 1603–1624. [CrossRef]
21. Singh, D.R. Data-Driven Distributionally Robust Stochastic Optimization via Wasserstein Distance with Applications to Portfolio Risk Management and Inventory Control. University of Minnesota. 2020. Available online: https://search.proquest.com/openview/e07ea32622dc17c7154041cb826d7d2e/1?pq-origsite=gscholar&cbl=18750&diss=y (accessed on 29 April 2026).
22. Shan, Z.; Lu, X.; Yang, J. Distributionally robust resource allocation using Wasserstein distance. *J. Glob. Optim.* **2025**. *online first*. [CrossRef]
23. Chowdhury, S.; Mémoli, F. The Gromov–Wasserstein distance between networks and stable network invariants. *Inf. Inference J. IMA* **2019**, *8*, 757–787. [CrossRef]
24. Mémoli, F. Gromov–Wasserstein Distances and the Metric Approach to Object Matching. *Found. Comput. Math.* **2011**, *11*, 417–487. [CrossRef]
25. De Plaen, H.; Fanuel, M.; Suykens, J.A. Wasserstein exponential kernels. In *2020 International Joint Conference on Neural Networks (IJCNN)*; IEEE: New York, NY, USA, 2020; pp. 1–6. Available online: https://ieeexplore.ieee.org/abstract/document/9207630/ (accessed on 27 February 2026).
26. Peyré, G.; Cuturi, M.; Solomon, J. Gromov-wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2016; pp. 2664–2672.
27. Ma, X.; Chu, X.; Wang, Y.; Lin, Y.; Zhao, J.; Ma, L.; Zhu, W. Fused gromov-wasserstein graph mixup for graph-level classifications. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 15252–15276.
28. Brogat-Motte, L.; Flamary, R.; Brouard, C.; Rousu, J.; D'Alché-Buc, F. Learning to predict graphs with fused Gromov-Wasserstein barycenters. In *International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2022; pp. 2321–2335. Available online: http://proceedings.mlr.press/v162/brogat-motte22a.html (accessed on 23 April 2026).
29. Seyedi, I.; Candelieri, A.; Archetti, F. Fused Unbalanced Gromov–Wasserstein-Based Network Distributional Resilience Analysis for Critical Infrastructure Assessment. *Mathematics* **2026**, *14*, 417. [CrossRef]
30. Wasi, A.T.; Islam, M.S.; Akib, A.R.; Bappy, M.M. Graph Neural Networks in Supply Chain Analytics and Optimization: Concepts, Perspectives, Dataset and Benchmarks. *arXiv* **2025**, arXiv:2411.08550. [CrossRef]
31. Goh, Q.H.; Tan, W.-S.; Ho, Y.K.; Chew, I.M.L. Integrated optimisation of biowaste-based green hydrogen supply chains from economic, environmental, and safety perspectives. *Comput. Chem. Eng.* **2025**, *199*, 109120. [CrossRef]
32. Kchaou-Boujelben, M.; Bensalem, M.; Jemai, Z. Bi-objective stochastic closed-loop supply chain network design under uncertain quantity and quality of returns. *Comput. Ind. Eng.* **2023**, *181*, 109308. [CrossRef]
33. Seyedi, I.; Candelieri, A.; Messina, E.; Archetti, F. Gromov–Wasserstein Meets Combinatorial Optimization: A Scalable Solver for the Capacitated Quadratic Assignment Problem. *Mathematics* **2026**, *14*, 1972. [CrossRef]
34. Dekker, R.; Fleischmann, M.; Inderfurth, K.; van Wassenhove, L.N. *Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains*; Springer Science & Business Media: Berlin, Germany, 2013.
35. Ivanov, D.; Dolgui, A. A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Prod. Plan. Control* **2021**, *32*, 775–788. [CrossRef]

36. Bauer, M.; Mémoli, F.; Needham, T.; Nishino, M. The Z-Gromov-Wasserstein Distance. *arXiv* **2025**, arXiv:2408.08233. [CrossRef]
37. Yang, J.; Labeau, M.; d'Alché-Buc, F. Exploiting Edge Features in Graphs with Fused Network Gromov-Wasserstein Distance. *arXiv* **2023**, arXiv:2309.16604. [CrossRef]
38. Séjourné, T.; Vialard, F.-X.; Peyré, G. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8766–8779.
39. Seyedi, I.; Mirzazadeh, S.; Maleki-Daronkolaei, A.; Mukhtar, M.; Sahran, S. An inventory model with reworking and setup time to consider effect of inflation and time value of money. *J. Eng. Sci. Technol.* **2016**, *11*, 416–430.
40. Simonetto, M.; Sgarbossa, F.; Battini, D.; Govindan, K. Closed loop supply chains 4.0: From risks to benefits through advanced technologies. A literature review and research agenda. *Int. J. Prod. Econ.* **2022**, *253*, 108582. [CrossRef]
41. Prajapati, H.; Kant, R.; Shankar, R. A modelling and management approach to risks in reverse logistics implementation. *IJBCRM* **2023**, *13*, 19. [CrossRef]
42. Ahmadi-Gh, Z.; Bello-Pintado, A. Sustainability isomorphism in buyer–supplier relationships: The impact of supply chain leadership. *Bus. Strat. Environ.* **2024**, *33*, 3635–3653. [CrossRef]
43. Rudyk, Y.; Bubela, T.; Maciuk, K. Russia-Ukraine War: Transport and Logistics Support for Grain Supply Chain in Regional Food Safety. 2023. Available online: https://sci.ldubgd.edu.ua/jspui/handle/123456789/11847 (accessed on 20 June 2026).
44. Azzouz, L. At a crossroads: Post-war Syria's transport future between sustainability and car-centric regression. *J. Transp. Geogr.* **2026**, *134*, 104711.
45. Klein, D.J. Resistance-distance sum rules. *Croat. Chem. Acta* **2002**, *75*, 633–649.
46. Ghosh, A.; Boyd, S.; Saberi, A. Minimizing Effective Resistance of a Graph. *SIAM Rev.* **2008**, *50*, 37–66. [CrossRef]
47. Hammond, D.K.; Gur, Y.; Johnson, C.R. Graph diffusion distance: A difference measure for weighted graphs based on the graph Laplacian exponential kernel. In *2013 IEEE Global Conference on Signal and Information Processing*; IEEE: New York, NY, USA, 2013; pp. 419–422. Available online: https://ieeexplore.ieee.org/abstract/document/6736904/ (accessed on 15 January 2026).
48. Barbe, A.; Sebban, M.; Gonçalves, P.; Borgnat, P.; Gribonval, R. Graph Diffusion Wasserstein Distances In Joint European conference on machine learning and knowledge discovery in databases. In *Lecture Notes in Computer Science*; Springer International Publishing: Cham, Switzerland, 2020; pp. 577–592.
49. Mémoli, F.; Needham, T. Comparison results for Gromov–Wasserstein and Gromov–Monge distances. *ESAIM Control Optim. Calc. Var.* **2024**, *30*, 78. [CrossRef]
50. Sturm, K.-T. *The Space of Spaces: Curvature Bounds and Gradient Flows on the Space of Metric Measure Spaces*; American Mathematical Society: Washington, DC, USA, 2023; Volume 290. Available online: https://www.ams.org/memo/1443/ (accessed on 30 July 2025).
51. Xu, H.; Luo, D.; Carin, L. Scalable Gromov-Wasserstein learning for graph partitioning and matching. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 3052–3062.
52. Chowdhury, S.; Needham, T. Generalized spectral clustering via Gromov-Wasserstein learning. In *International Conference on Artificial Intelligence and Statistics*; PMLR: Cambridge, MA, USA, 2021; pp. 712–720. Available online: https://proceedings.mlr.press/v130/chowdhury21a.html (accessed on 24 October 2025).
53. Chung, F.R. *Spectral Graph Theory*; American Mathematical Society: Washington, DC, USA, 1997; Volume 92.
54. Coifman, R.R.; Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 5–30. [CrossRef]
55. Burkard, R.; Dell'Amico, M.; Martello, S. Assignment problems: Revised reprint. In *Assignment Problems, Society for Industrial and Applied Mathematics*; Cambridge University Press: Cambridge, UK, 2012; pp. 327–393. [CrossRef]
56. Mula, O.; Nouy, A. Moment-SoS methods for optimal transport problems. *Numer. Math.* **2024**, *156*, 1541–1578. [CrossRef]
57. Chen, J.; Nguyen, B.T.; Koh, S.H.; Soh, Y.S. Semidefinite relaxations of the Gromov-Wasserstein distance. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 69814–69839. [CrossRef]
58. Ryner, M.; Kronqvist, J.; Karlsson, J. Globally solving the Gromov-Wasserstein problem for point clouds in low dimensional Euclidean spaces. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 7930–7946. [CrossRef]
59. Vayer, T.; Flamary, R.; Courty, N.; Tavenard, R.; Chapel, L. Sliced gromov-wasserstein. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 14726–14736.
60. Chowdhury, S.; Miller, D.; Needham, T. Quantized Gromov-Wasserstein. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 12977, pp. 811–827. [CrossRef]
61. Fatras, K.; Zine, Y.; Majewski, S.; Flamary, R.; Gribonval, R.; Courty, N. Minibatch optimal transport distances; analysis and applications. *arXiv* **2021**, arXiv:2101.01792. [CrossRef]
62. Beier, F.; Beinert, R.; Steidl, G. On a linear Gromov–Wasserstein distance. *IEEE Trans. Image Process.* **2022**, *31*, 7292–7305. [CrossRef] [PubMed]
63. Salmona, A.; Delon, J.; Desolneux, A. Gromov-Wasserstein-like Distances in the Gaussian Mixture Models Space. *arXiv* **2024**, arXiv:2310.11256. [CrossRef]

64. Solomon, J.; Peyré, G.; Kim, V.G.; Sra, S. Entropic metric alignment for correspondence problems. *ACM Trans. Graph.* **2016**, *35*, 1–13. [CrossRef]
65. Scetbon, M.; Peyré, G.; Cuturi, M. Linear-time gromov wasserstein distances using low rank couplings and costs. In *International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2022; pp. 19347–19365. Available online: https://proceedings.mlr.press/v162/scetbon22b (accessed on 21 August 2025).
66. Rioux, G.; Goldfeld, Z.; Kato, K. Entropic gromov-wasserstein distances: Stability and algorithms. *J. Mach. Learn. Res.* **2024**, *25*, 1–52.
67. Kerdoncuff, T.; Emonet, R.; Sebban, M. Sampled Gromov Wasserstein. *Mach. Learn.* **2021**, *110*, 2151–2186. [CrossRef]
68. Li, M.; Yu, J.; Xu, H.; Meng, C. Efficient Approximation of Gromov-Wasserstein Distance Using Importance Sparsification. *J. Comput. Graph. Stat.* **2023**, *32*, 1512–1523. [CrossRef]
69. Mazelet, S.; Flamary, R.; Thirion, B. Unsupervised Learning for Optimal Transport plan prediction between unbalanced graphs. *arXiv* **2025**, arXiv:2506.12025. [CrossRef]