Journal of Machine Learning Research 26 (2025) 1-57 Submitted 12/24; Revised 11/25; Published 12/25

The Z-Gromov-Wasserstein Distance

Martin Bauer MBAUER2@FSU.EDU
Tom Needham TNEEDHAM@FSU.EDU
Mao Nishino MNISHINO@FSU.EDU
Department of Mathematics
Florida State University
Tallahassee, FL 32306-4510, USA

Facundo Mémoli FACUNDO.MEMOLI@RUTGERS.EDU
Department of Mathematics
Rutgers University
Piscataway, NJ 08854-8019, USA

Editor: Qiang Liu

Abstract

The Gromov-Wasserstein (GW) distance is a powerful tool for comparing metric measure spaces which has found broad applications in data science and machine learning. Driven by the need to analyze data sets whose objects have increasingly complex structure (such as node and edge-attributed graphs), several variants of GW distance have been introduced in the recent literature. With a view toward establishing a general framework for the theory of GW-like distances, this paper considers a vast generalization of the notion of a metric measure space: for an arbitrary metric space Z, we define a Z-network to be a measure space endowed with a kernel valued in Z. We introduce a method for comparing Z-networks by defining a generalization of GW distance, which we refer to as Z-Gromov-Wasserstein (Z-GW) distance. This construction subsumes many previously known metrics and offers a unified approach to understanding their shared properties. This paper demonstrates that the Z-GW distance defines a metric on the space of Z-networks which retains desirable properties of Z, such as separability, completeness, and geodesicity. Many of these properties were unknown for existing variants of GW distance that fall under our framework. Our focus is on foundational theory, but our results also include computable lower bounds and approximations of the distance which will be useful for practical applications.

Keywords: Gromov-Wasserstein distance, metric measure spaces, optimal transport, attributed networks, geodesics

Contents

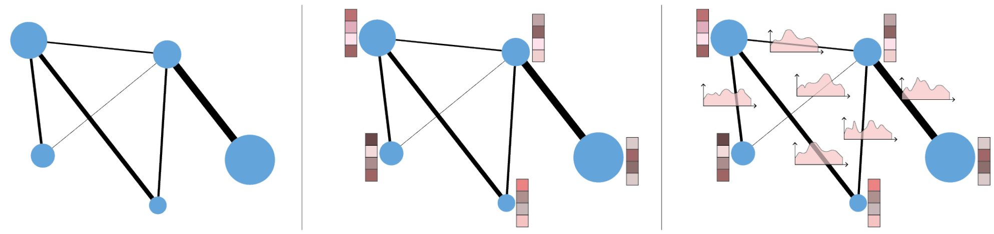
1. Introduction

Frequently in pure and applied mathematics, one requires a mechanism for measuring the dissimilarity of objects which are a priori incomparable. For example, the Gromov-Hausdorff distance provides such a mechanism for comparing abstract Riemannian manifolds and famously provides a notion of convergence for sequences of manifolds, under which certain geometric features are preserved (Cheeger and Colding, 1997; Gromov, 2006). The Gromov-Hausdorff distance, in fact, defines a metric on the space of arbitrary compact metric spaces. This is useful in applied areas such as shape analysis and data science, where it is increasingly common to encounter analysis tasks involving ensembles of complex data objects which naturally carry metric structures, such as point clouds or graphs—the ability to compute distances between these objects opens these problems to metric-based techniques (Mémoli and Sapiro, 2004, 2005; Chazal et al., 2009; Carlsson and Mémoli, 2010; Mémoli, 2012). Introducing measures on the metric spaces allows for the application of ideas from the field of optimal transport, leading to the following construction, introduced and studied in (Mémoli, 2007; Mémoli, 2011a): given a pair of metric spaces endowed with probability measures (X, d_X, μ_X) and (Y, d_Y, μ_Y), the *Gromov-Wasserstein p-distance* (for $p \geqslant 1$) between them is given by the quantity

$$\mathrm{GW}_p(X,Y) = \frac{1}{2}\inf_{\pi}\left(\iint\limits_{(X\times Y)^2} |d_X(x,x') - d_Y(y,y')|^p \pi(dx\times dy)\pi(dx'\times dy')\right)^{1/p}, \tag{1}$$

where the infimum is over probability measures π on $X \times Y$ whose left and right marginals are μ_X and μ_Y, respectively. Intuitively, such a measure describes a probabilistic correspondence between the sets X and Y, and the integral in (1) measures the extent to which such a correspondence distorts the metric structures of the spaces. The optimization problem (1) therefore seeks a probabilistic correspondence which minimizes total metric distortion. It was shown by Mémoli (2007) that the Gromov-Wasserstein p-distance defines a metric on the space of all triples (X, d_X, μ_X), considered up to measure-preserving isometries, when (X, d_X) is compact and μ_X is fully supported (the compactness assumption can be relaxed; see Mémoli and Needham 2022a; Sturm 2023).

The Gromov-Wasserstein (GW) framework has become a popular tool in data science and machine learning—see (Demetci et al., 2022; Chowdhury and Needham, 2021; Chapel et al., 2020; Xu, 2020; Chowdhury and Needham, 2020; Alvarez-Melis and Jaakkola, 2018), among many others—due to its flexibility in handling diverse data types, its robustness to changes in metric or measure structures (Mémoli, 2011a, Theorem 5.1), and recent advances in scalable and empirically accurate computational schemes for its estimation (Peyré et al., 2016; Xu et al., 2019; Chowdhury et al., 2021; Scetbon et al., 2022; Li et al., 2023; Vedula



(a) Measure Network (b) Node-Attributed Network (c) Z-Network

Figure 1: Schematic illustration of types of networks. (a) A graph with edge weights and node weights (each visualized by size variations). This structure is encoded as a measure network, and two such structures can be compared through the Gromov-Wasserstein distance (Mémoli, 2007; Chowdhury and Mémoli, 2019). (b) A weighted graph with additional node features, consisting of an assignment of a vector in $\mathbb{R}^n$ to each node (visualized as a column vector). These objects can be compared via the Fused Gromov-Wasserstein distance (Vayer et al., 2019, 2020a). (c) Additionally, a graph can be endowed with edge features, assigning a point in some fixed metric space Z to each edge (here, we visualize a 1-dimensional probability distribution attached to each edge). These complex objects are modeled as Z-networks, in the language of this paper, and two such objects can be compared through our proposed framework. By choosing an appropriate target space Z, one recovers many notions of distance between structured objects that have appeared previously in the literature—see Table 1.

et al., 2024). It was observed by Peyré et al. (2016) that the formula for GW distance in (1) still makes sense when the condition that d_X and d_Y are metrics is relaxed; that is, the formula gives a meaningful comparison between structures of the form (X, ω_X, μ_X), where $\omega_X : X \times X \to \mathbb{R}$ is an arbitrary measurable function. This point of view was later studied formally by Chowdhury and Mémoli (2019), where it was shown that the GW distance defines a metric on the space of such triples (X, ω_X, μ_X), considered up to a natural notion of equivalence. This is convenient, for example, when handling graph data sets, where it is natural to represent a graph's structure through its adjacency function, Laplacian, or heat kernel.

Driven by applications to machine learning on increasingly complex data types, several variants of GW distance have been introduced in the literature (Mémoli, 2009; Vayer et al., 2020a; Kim, 2020; Mémoli et al., 2023; Arya et al., 2024; Yang et al., 2023). For example, Vayer et al. (2020a) introduces an adaptation of GW distance which is equipped to handle graphs endowed with *node features*; that is, graphs whose nodes are endowed with values in some auxiliary metric space (in fact, this idea goes back further to, at least, Chazal et al. 2009, Section 5, where a similar construction provided variants of both the Gromov-Hausdorff distance and the $p = \infty$ version of GW distance). Each time a new variant of the GW distance is introduced in the literature, its metric properties have been re-established. Reviewing these proofs reveals that they tend to follow a common template, suggesting the existence of a higher-level explanation of their shared properties.

The main goal of this paper is to formally develop a high-level structure which encompasses the GW variants described in the previous paragraph. This is accomplished by studying triples (X, ω_X, μ_X) where $\omega_X : X \times X \to Z$ is now a function valued in some fixed but arbitrary metric space (Z, d_Z); we call such a triple a *Z-network* (see Figure 1). Two Z-networks can be compared via a natural generalization of the GW distance (1): the *Z-Gromov-Wasserstein p-distance* between Z-networks (X, ω_X, μ_X) and (Y, ω_Y, μ_Y) is

$$\mathrm{GW}_p^Z(X,Y) = \frac{1}{2} \inf_\pi \left(\iint_{(X\times Y)^2} d_Z\big(\omega_X(x,x'), \omega_Y(y,y')\big)^p \pi(dx \times dy)\pi(dx' \times dy') \right)^{1/p}. \quad (2)$$

One should immediately observe that if $Z = \mathbb{R}$ (with its standard metric) then GW_p^Z recovers $\mathrm{GW}_p(X,Y)$. Moreover, we will show that the GW variants described above correspond to Z-GW distances for other appropriate choices of Z. The broad goal of the paper is to establish a general theoretical framework for GW-like distances, as understanding geometric properties of this general structure eliminates the need to re-derive the properties for GW variants which fall into our framework. Besides enabling the avoidance of such redundancies for future variants of GW distance, this general perspective leads to novel insights about existing metrics: several of our theoretical results were previously unknown or only shown in weaker forms for distances already studied in the literature.

1.1 Main Contributions and Outline

Let us now outline our main results, which are stated here somewhat informally.

- We show that several GW-like metrics appearing in the literature can be realized as Z-GW distances. In particular, Theorem 12 states that the Wasserstein distance, (standard) GW distance (Mémoli, 2007), ultrametric GW distance (Mémoli et al., 2023), (p,q)-GW distance (Arya et al., 2024), Fused GW distance (Vayer et al., 2020a), Fused Network GW distance (Yang et al., 2023), spectral GW distance (Mémoli, 2009) and GW distance between weighted dynamic metric spaces (Kim, 2020) can all be realized as Z-GW distances. We also explain how the graphon cut metric (Borgs et al., 2008) fits into our framework. Finally, we show that Z-GW distances define natural metrics on spaces of shape graphs (Guo et al., 2022), connection graphs (Robertson et al., 2023) and probabilistic metric spaces (Menger, 1942). These results are summarized in Table 1.

- Theorem 29 says that, when (Z, d_Z) is separable, the Z-GW distance GW_p^Z defines a metric on the space $\mathcal{M}_\sim^{Z,p}$ consisting of Z-networks, considered up to a natural notion of equivalence (Definition 28). As a consequence, Corollary 30 shows that Fused GW and Fused Network GW distances are metrics; these were previously only shown to satisfy a certain *weak triangle inequality*. The proof of Theorem 29 relies on a technical result, which says that the solution of the optimization problem (2) is always realized (Theorem 26).

- Several geometric and topological properties of the metric space $(\mathcal{M}_\sim^{Z,p}, \mathrm{GW}_p^Z)$ are established: it is separable (Proposition 36), complete if and only if Z is (Theorem 39), contractible, regardless of the topology of Z (Theorem 42), and geodesic if Z is (Theorem 45). Many of these results are novel when restricted to the examples of Z-GW distances covered in Theorem 12.

- Approximations of Z-GW distances are established via certain polynomial-time computable lower bounds in Theorem 50. Moreover, Theorem 52 provides a quantitative approximation of general Z-GW distances by $\mathbb{R}^n$-GW distances, the latter of which should be efficiently estimable through mild adaptations of existing GW algorithms.

The structure of the paper is as follows. After concluding the introduction section with a discussion of related work, the main definitions of the objects of interest are given in Section 2. Section 3 describes examples of Z-GW distances, including some distances which have already appeared in the literature and some which are novel to the present paper. The basic geometric properties of Z-GW distances are established in Section 4. Finally, our results on approximations of Z-GW distances are presented in Section 5. The main body of the paper concludes with a discussion of open questions and future research directions in Section 6. Proofs of our main results are included in the main body of the paper, but we relegate proofs of a few technical results to Section A.

1.2 Related Work

Concepts related to the Z-Gromov-Wasserstein distance have been considered previously, see, e.g., (Jain and Obermayer, 2009; Peyré et al., 2016; Yang et al., 2023; Kawano et al., 2024). We now give precise comparisons of these previous works to the setting of the present article.

Peyré et al. (2016) already considers variants of GW distance between $\mathbb{R}$-networks (i.e., objects of the form (X, ω_X, μ_X) with ω_X valued in $\mathbb{R}$) where the integrand of (1) is replaced with a more general *loss function* of the form

$$X \times Y \times X \times Y \ni (x, y, x', y') \mapsto L(\omega_X(x, x'), \omega_Y(y, y')) \in \mathbb{R},$$

with $L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ an arbitrary function. The article places a specific focus on the case where L is the Kullback-Leibler divergence, which is not a metric, and therefore does not fall within our framework. The geometric properties that we establish for Z-GW distances in this paper largely depend on the assumption that (Z, d_Z) is a metric space, so we restrict our attention to this setting.

While we were in the process of preparing the article, Yang et al. (2023) independently introduced the idea of a Z-network, although in the slightly less general context that network kernels are assumed to be bounded and continuous. A version of the Z-GW distance is also formulated therein, and their framework is applied to the analysis of attributed graph data. The work of Yang et al. (2023) is primarily geared toward computational applications, while the present paper is focused on developing the mathematical properties of the Z-GW distance; all of our results are novel, with connections to previous results clarified as necessary below. We also remark here that the work of Yang et al. gives a notion of Z-GW distance which appears at first glance to be more general, in that it includes additional terms handling node features and edge weights; we show, however, in Section 3.1.2 that these terms are, in fact, not necessary. A similar framework to that of Yang et al. was

Metric	Data Type	Target space Z	References
Wasserstein Distance	Distributions over (Z, d_Z)	Z, d_Z (any metric space)	Villani (2003)
Gromov-Wasserstein (GW) Distance	Metric measure spaces	$\mathbb{R}$, standard metric	Mémoli (2007); Mémoli (2011a); Chowdhury and Mémoli (2019)
Ultrametric GW Distance	Ultrametric measure spaces	$\mathbb{R}$, max metric	Mémoli et al. (2023)
(p, q)-GW Distance	Metric measure spaces	$\mathbb{R}$, Λ_q (see Equation 4)	Arya et al. (2024)
Fused GW Distance	Node-attributed graphs	$\mathbb{R} \times Y$ for a metric space Y	Vayer et al. (2019, 2020a)
Fused Network GW Distance	Node- and edge-attributed graphs	Product of node and edge spaces	Yang et al. (2023); Kawano et al. (2024)
Spectral GW Distance	Riemannian manifolds	Continuous functions, sup metric	Mémoli (2009)
GW Distance for Weighted DMSs	Weighted Dynamic Metric Spaces	Cts. functions, interleaving distance	Kim (2020)
Shape Graph Distance	Embedded 1-d stratified spaces	Space of curves, any metric	Sukurdeep et al. (2022); Srivastava et al. (2020); Bal et al. (2024)
Connection Graph Distance	$\mathrm{O}(n)$-attributed graphs	$\mathrm{O}(n)$, Frobenius distance	Bhamre et al. (2015); Singer (2011); Robertson et al. (2023)
Probabilistic Metric Space Distance	Probabilistic metric spaces	Wasserstein space over $\mathbb{R}$	Menger (1942); Wald (1943); Kramosil and Michálek (1975)

Table 1: Summary of Gromov-Wasserstein-like distances which fall under the framework described in this paper. In each case, we provide the name of the metric, the type of data objects which it is able to compare, the choice (Z, d_Z) which realizes the metric in our framework, and references to where the metric was first studied. The last three distances are novel to this paper, but handle data objects which have been studied in prior work.

also recently studied by Kawano et al. (2024), where the focus was restricted to $\mathbb{R}^n$-valued kernels on finite sets, also with a view toward applications to attributed graph data sets. In summary, Yang et al. (2023) and Kawano et al. (2024) focus on extending the notion of fused Gromov-Wasserstein distance to do machine learning on attributed networks, whereas our goal is to develop a unifying theory on a more general class of GW-like distances.

The first work to consider structures similar to Z-networks, to our knowledge, is that of Jain and Obermayer (2009), where the objects under consideration are kernels over a finite set of fixed size n, valued in a fixed metric space. These structures are compared using a distance similar to the Z-GW distance, with a key difference being that the formulation does not involve measures: the optimization is performed over the conjugation action by the permutation group on n letters, rather than on the space of couplings. Kernels on sets of different cardinalities are compared by "padding" the smaller kernel with some fixed value to bring them to the same size—e.g., if the target metric space is a vector space, then the kernel is padded with zero vectors. The formulations of Yang et al. (2023); Kawano et al. (2024) and the present paper can then be viewed as OT-based extensions of the ideas of Jain and Obermayer, which are able to avoid the arguably unnatural padding operation.

Another recent approach to a general framework for Gromov-Wasserstein-like distances is developed by Zhang et al. (2024a). In that paper, the generalization is given by allowing optimization over spaces of measure couplings with a certain additional structure, rather than varying the target of the kernel function. That framework encompasses a different collection of variants of GW distances (namely, those of the form presented by Chowdhury et al. 2023; Vayer et al. 2020b) than those considered here. Moreover, Zhang et al. 2024a is mainly focused on curvature bounds and on a computational framework, rather than the structural results presented in this paper.

2. Z-Gromov-Wasserstein Distances

We now present the main definitions and constructions of the paper.

2.1 Background

We begin by recalling some background terminology and notation on various notions of distance between probability measures.

2.1.1 Basic Terminology and Notation

We emphasize here that the term **metric space** is reserved for a pair (Z, d_Z), where $d_Z : Z \times Z \to \mathbb{R}$ satisfies the usual axioms of symmetry, positive-definiteness and the triangle inequality. If $d_Z(z, z') = 0$ for some $z \neq z'$ (i.e., d_Z does not satisfy the positive-definiteness axiom), then we refer to d_Z as a **pseudometric** and (Z, d_Z) as a **pseudometric space**. When the existence of a preferred (pseudo)metric is clear from context, we may abuse notation and use only Z to denote the metric space.

Recall that a topological space Z is called a **Polish space** if it is completely metrizable and separable (i.e., contains a countable dense subset). When metrized by a particular complete metric d_Z, we call (Z, d_Z) a **Polish metric space**. A standard assumption for many results in abstract measure theory is that the underlying space is Polish—for example, see (Srivastava, 2008).

Let (X, μ) be a measure space, Y a measurable space and $\phi : X \to Y$ a measurable map. The **pushforward measure** on Y is denoted $\phi_* \mu$. We recall that this is defined on a measurable subset $A \subset Y$ by $\phi_* \mu(A) = \mu(\phi^{-1}(A))$.

Given a point $z \in Z$, the associated **Dirac measure** is the Borel measure δ_z on Z defined on a Borel set $A \subset Z$ by

$$\delta_z(A) = \begin{cases} 1 & \text{if } z \in A \\ 0 & \text{if } z \notin A. \end{cases}$$

Dirac measures will appear in several constructions below (e.g., Section 3.2.5, the proofs of Proposition 38 and Theorem 39, etc.).

2.1.2 Optimal Transport Distances

Our constructions are largely inspired by concepts from the field of optimal transport (OT), which studies certain metrics on spaces of probability distributions, called *Wasserstein distances.* The following concepts are well-established—see (Villani, 2003) as a general reference on OT.

Definition 1 (Coupling)**.** *Let (X, μ) and (Y, ν) be probability spaces. A* ***coupling*** *of μ and ν is a probability measure π on $X \times Y$ with marginals μ and ν, respectively. We denote the set of couplings between μ and ν as $\mathcal{C}(\mu, \nu)$.*

Definition 2 (Wasserstein Distance)**.** *Let (Z, d_Z) be a Polish metric space and let μ and ν be Borel probability measures on Z with finite p-th moments, for $p \in [1, \infty]$. The* ***Wasserstein p-distance*** *between μ and ν is*

$$\mathrm{W}_p(\mu, \nu) = \inf_{\pi \in \mathcal{C}(\mu,\nu)} \|d_Z\|_{L^p(\pi)}.$$

This is written more explicitly, for $p < \infty$, as

$$\mathrm{W}_p(\mu, \nu) = \inf_{\pi \in \mathcal{C}(\mu,\nu)} \left(\int_{Z \times Z} d_Z(x, y)^p \pi(dx \times dy) \right)^{\frac{1}{p}},$$

and the integral expression is replaced by an essential supremum in the $p = \infty$ case.

The classical notion of Wasserstein distance between probability measures on the *same* metric space can be adapted to compare probability measures on *distinct* metric spaces—this was the setting of the initial work on *Gromov-Wasserstein* distances (Mémoli, 2007; Mémoli, 2011a). The concept extends further to define metrics on the space of more general objects, called *measure networks*, as was first formalized in (Chowdhury and Mémoli, 2019). We recall the relevant definitions below.

Definition 3 (Measure Network)**.** *A* ***measure network*** *is a triple (X, ω_X, μ_X) such that X is a Polish space, μ_X is a Borel probability measure and*

$$\omega_X : X \times X \to \mathbb{R}$$

is a measurable function which we refer to as the ***network kernel****. We frequently abuse notation and write X in place of (X, ω_X, μ_X), when the existence of a preferred network kernel and measure are clear from context.*

Examples 4. Two of the most common sources of examples of measure networks are:

1. **Metric measure spaces.** A **metric measure space** is a measure network such that the network kernel is a metric (which induces the given topology on X).

2. **Graph representations.** Given a (possibly weighted) graph, one can construct a measure network representation by taking X to be the vertex set, μ_X some probability measure on the vertices (e.g., uniform or degree-weighted), and ω_X a graph kernel which encodes connectivity information. Natural choices of graph kernel include a (weighted) adjacency function (Xu et al., 2019) or a graph heat kernel (Chowdhury and Needham, 2021).

Measure networks can be compared via the pseudometric defined by Mémoli (2007); Chowdhury and Mémoli (2019), which we now recall.

Definition 5 (Gromov-Wasserstein p-Distance)**.** *Consider measure networks* (X, ω_X, μ_X) *and* (Y, ω_Y, μ_Y)*, and let* $p \in [1, \infty]$*. The* ***p-distortion*** *of a coupling* $\pi \in \mathcal{C}(\mu_X, \mu_Y)$ *is*

$$\mathrm{dis}_p(\pi) = \|\omega_X - \omega_Y\|_{L^p(\pi \otimes \pi)},$$

where $\pi \otimes \pi$ *is the product measure on* $(X \times Y)^2$*, and where we consider* $(x, y, x', y') \mapsto \omega_X(x, x') - \omega_Y(y, y')$ *as a function on this space. Explicitly, for* $p < \infty$ *this is given by*

$$\mathrm{dis}_p(\pi) = \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p \pi(dx \times dy) \pi(dx' \times dy') \right)^{1/p},$$

and for $p = \infty$ *it is given by the essential supremum*

$$\mathrm{dis}_\infty(\pi) = \mathrm{esssup}_{\pi \otimes \pi} |\omega_X - \omega_Y| = \sup_{(x,y),(x',y') \in \mathrm{supp}(\pi)} |\omega_X(x, x') - \omega_Y(y, y')|,$$

where $\mathrm{supp}(\pi)$ *denotes the support of the measure. Note that we suppress the dependence on* X *and* Y *from our notation for the distortion function. The* ***Gromov-Wasserstein (GW) p-distance*** *between* X *and* Y *is*

$$\mathrm{GW}_p(X, Y) = \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X, \mu_Y)} \mathrm{dis}_p(\pi).$$

In this article we will introduce a generalization of the GW distance and when clarity is needed, we refer to GW_p as the ***standard* Gromov-Wasserstein (GW) p-distance**. The standard GW p-distance is finite when restricted to the space of measure networks whose network kernels have finite p-th moment. On this subspace, it defines a pseudometric whose distance-zero equivalence classes can be precisely characterized: see (Chowdhury and Mémoli, 2019, Theorem 2.4) or the discussion in Section 4 below.

2.1.3 Metric Space-Valued L^p-Spaces

The main object of study in this paper generalizes the above definitions to the case where the network kernel ω takes values in a general metric space. To ensure that the distance is finite, we will assume that the network kernel is an L^p function, in the sense of Korevaar and Schoen (1993, Section 1.1). This parallels the definition of the Wasserstein distance, where we assume that the measures have finite p-th moments.

We will now recall the definition of L^p functions valued in a metric space. For measurable functions f and g, the natural definition of a distance between f and g is the integral of $d(f(x), g(x))$, but this function is not always measurable due to Nedoma's pathology (Schilling and Kühn, 2021, Section 15.9). However, if we assume that the target metric space is separable, the distance is measurable, and the integral is well-defined. With this in mind, we present the following definition, from Korevaar and Schoen (1993, Section 1.1).

Definition 6 (Metric Space-Valued L^p Spaces)**.** *Let X be a measure space equipped with a measure μ_X, and suppose (Y, d_Y) is a separable metric space. Fix $y_0 \in Y$. We define the* ***space of L^p-functions*** *$X \to Y$ by*

$$L^p(X, \mu_X; Y) = \left\{ f : X \to Y \middle| \int_X d_Y(f(x), y_0)^p \mu_X(dx) < \infty \right\}$$

for $p \in [1, \infty)$; similarly, for $p = \infty$, the space is defined by

$$L^\infty(X, \mu_X; Y) = \{ f : X \to Y | \inf\{C \geqslant 0 \mid d_Y(f(x), y_0) \leqslant C \text{ for } \mu_X\text{-a.e. } x \in X\} < \infty \}.$$

These can be expressed somewhat more concisely, for all $p \in [1, \infty]$, as

$$L^p(X, \mu_X; Y) = \{ f : X \to Y \mid \|d_Y(f(\cdot), y_0)\|_{L^p(\mu_X)} < \infty \}.$$

The L^p space is then equipped with the distance

$$D_p(f, g) = \|d_Y(f(x), g(x))\|_{L^p(\mu_X)} = \begin{cases} \left(\int_X d_Y(f(x), g(x))^p \mu_X(dx)\right)^{1/p} & 1 \leqslant p < \infty \\ \operatorname{esssup}_{\mu_X} d_Y(f(x), g(x)) & p = \infty. \end{cases}$$

Remark 7. *If $f \in L^p(X, \mu_X; Y)$, then $\|d_Y(f(\cdot), y)\|_{L^p(\mu_X)} < \infty$ holds for every $y \in Y$. This is an easy consequence of the triangle inequalities of d_Y and the $L^p(\mu_X)$-norm. Therefore, the above definition of L^p spaces is independent of the choice of y_0.*

The distance D_p is a pseudometric which assigns distance zero to functions that agree almost everywhere. We abuse notation and consider $L^p(X, \mu_X; Y)$ as a metric space by implicitly identifying functions that differ only on a measure zero set. The following properties of L^p spaces will be useful later to establish various properties of the metrics introduced in this paper—see Proposition 36, Proposition 38 and Theorem 39. The proof is provided in Section A.1.

Proposition 8. *Let X be a measure space equipped with a measure μ_X, and suppose (Y, d_Y) is a separable metric space. Then:*

1. *$L^p(X, \mu_X; Y)$ is a complete metric space if Y is complete.*
2. *$L^p(X, \mu_X; Y)$ is separable.*
3. *Any $f \in L^p([0,1]^d, \mathscr{L}^d; Y)$ can be approximated by piecewise constant functions that are constant on a grid of a fixed step size. Here, $\mathscr{L}^d$ is the Lebesgue measure on $[0,1]^d$.*

Remark 9. *Item 2 and Item 3 in Proposition 8 are classical results when Y is a Banach space. However, to the best of our knowledge, the case when Y is a general separable metric space has not been studied in the prior literature.*

2.2 Z-Valued Measure Networks

In this subsection, we introduce our main objects of study, which generalize the notion of a measure network (Definition 3) and an extension of GW distance for comparing them. The basic idea is to allow network kernels to take values in some fixed, but arbitrary metric space. Similar ideas were considered recently in the machine learning literature (Yang et al., 2023; Kawano et al., 2024), with applications to analysis of attributed graph data sets, but the idea goes back at least to Jain and Obermayer (2009)—see the discussion in Section 1.2. The definitions given below generalize those which have appeared in the previous literature.

2.2.1 Main Definitions

Throughout this section, and much of the rest of the paper, we fix a complete and separable metric space (Z, d_Z), which we frequently refer to only as Z.

Definition 10 (Z-Network)**.** *A Z**-valued** p**-measure network** is a triple (X, ω_X, μ_X), where X is a Polish space, μ_X is a Borel probability measure and*

$$\omega_X : X \times X \to Z$$

*is an element of $L^p(X \times X, \mu_X \otimes \mu_X; Z)$. We refer to ω_X as a **network kernel**, or Z**-valued** p**-network kernel**, when additional clarity is necessary. We frequently abuse notation and write X in place of (X, ω_X, μ_X). For short, we refer to X as a (Z, p)**-network**; if the particular value of p is not important to the discussion, we omit p and refer to X as a Z**-network**.*

We now introduce a notion of distance between Z-networks.

Definition 11 (Gromov-Wasserstein Distance for (Z, p)-Networks)**.** *Let $p \in [1, \infty]$ and let $X = (X, \omega_X, \mu_X)$ and $Y = (Y, \omega_Y, \mu_Y)$ be Z-valued p-networks. The associated p**-distortion** of a coupling $\pi \in \mathcal{C}(\mu_X, \mu_Y)$ is*

$$\mathrm{dis}_p^Z(\pi) = \|d_Z \circ (\omega_X \times \omega_Y)\|_{L^p(\pi \otimes \pi)}.$$

Explicitly, for $p < \infty$,

$$\mathrm{dis}_p^Z(\pi) = \left(\int_{X \times Y} \int_{X \times Y} d_Z(\omega_X(x, x'), \omega_Y(y, y'))^p \pi(dx \times dy) \pi(dx' \times dy') \right)^{1/p}$$

and, for $p = \infty$,

$$\mathrm{dis}_\infty^Z(\pi) = \mathrm{esssup}_{\pi \otimes \pi} d_Z \circ (\omega_X \times \omega_Y).$$

*The Z**-Gromov-Wasserstein (Z-GW)** p**-distance** between X and Y is*

$$\mathrm{GW}_p^Z(X, Y) = \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X, \mu_Y)} \mathrm{dis}_p^Z(\pi).$$

When the metric on Z needs to be emphasized, we write this as $\mathrm{GW}_p^{(Z, d_Z)}$.

Important examples of Z-networks are provided below in Section 3. Sections 4 and 5 will be dedicated to rigorously developing and unveiling the fundamental properties of the Z-GW distance.

3. Examples of Z-Networks and Z-Gromov-Wasserstein Distances

We show in Section 4 below that the Z-Gromov-Wasserstein distance from Definition 11 induces a metric on a certain quotient of the space of Z-networks and establish some of its geometric properties. To motivate this, we first provide many examples of Z-GW distances; these include metrics which have already appeared in the literature (Section 3.1) and other natural metrics that appear to be novel (Section 3.2). An important takeaway message is that it is not always necessary to re-establish metric properties of variants of GW distances from scratch—a practice which is common in the recent literature—since these properties frequently follow immediately from high-level principles.

3.1 Examples of Existing Z-Gromov-Wasserstein Distances in the Literature

In this subsection, we show that the Z-GW distance, as defined above, generalizes several optimal transport distances which have previously appeared in the literature. Our main results are summarized in the following theorem (see also Table 1):

Theorem 12. *For appropriate choices of Z, the following distances can be realized as Z-Gromov-Wasserstein distances: Wasserstein distance, standard GW distance (Mémoli, 2007), ultrametric GW distance (Mémoli et al., 2023), (p,q)-GW distance (Arya et al., 2024), Fused GW distance (Vayer et al., 2020a), Fused Network GW distance (Yang et al., 2023), spectral GW distance (Mémoli, 2009), and GW distance between weighted dynamic metric spaces (Kim, 2020).*

The definitions of these optimal transport-type distances are provided below, as necessary, and the proof of the theorem is split among Propositions 13, 14, 15, 16, 23, 24, and 25, which treat the various metrics independently.

3.1.1 Wasserstein Distance and Gromov-Wasserstein Distances

The first result is obvious from the definitions.

Proposition 13. *The standard Gromov-Wasserstein distance between measure networks (Definition 5) is a Z-GW distance with (Z, d_Z) equal to $\mathbb{R}$ with its standard metric.*

Consider measures μ and ν on a metric space (Z, d_Z). It is well-known that the standard GW distance is not a generalization of the Wasserstein distance, in the sense that, for $X = (Z, d_Z, \mu)$ and $Y = (Z, d_Z, \nu)$,

$$\mathrm{GW}_p(X,Y) \leqslant \mathrm{W}_p(\mu,\nu),$$

(see Mémoli 2011a, Theorem 5.1(c)), but equality does not hold in general—as a simple example, consider the case where $Z = \mathbb{R}$ and μ and ν are translates of one another, yielding $\mathrm{GW}_p(X,Y) = 0$ and $\mathrm{W}_p(\mu,\nu) > 0$. Next, we show that Z-GW distance does give such a generalization.

Proposition 14. *The Wasserstein distance over an arbitrary metric space (Z, d_Z) (Definition 2) can be realized as a Z-GW distance.*

Proof Let (Z, d_Z) be a metric space and let μ and ν be Borel probability measures on Z with finite p-th moments. To account for different scaling conventions in the definitions of Wasserstein and GW distances, we replace d_Z with $\hat{d}_Z = 2 \cdot d_Z$ as the metric on the target for our Z-networks. Define Z-valued p-measure networks $X = (Z, \omega_X, \mu)$ and $Y = (Z, \omega_Y, \nu)$ with $\omega_X, \omega_Y : Z \times Z \to Z$ both denoting projection onto the first coordinate, i.e., $\omega_X(z, z') = \omega_Y(z, z') = z$. Then, for any coupling $\pi \in \mathcal{C}(\mu, \nu)$, we have, for $p < \infty$,

$$\begin{aligned}\frac{1}{2^p}\mathrm{dis}_p^Z(\pi)^p &= \frac{1}{2^p}\int_{X\times Y}\int_{X\times Y} \hat{d}_Z(\omega_X(x,x'),\omega_Y(y,y'))^p \pi(dx \times dy)\pi(dx' \times dy') \\ &= \frac{1}{2^p}\int_{X\times Y}\int_{X\times Y} 2^p \cdot d_Z(x,y)^p \pi(dx \times dy)\pi(dx' \times dy') \\ &= \int_{X\times Y} d_Z(x,y)^p \pi(dx \times dy).\end{aligned}$$

The cost of any coupling is therefore the same for the Z-GW and Wasserstein distances, and it follows that $\mathrm{GW}_p^Z(X, Y) = \mathrm{W}_p(\mu, \nu)$. We wrote out the calculation using integrals for the sake of clarity, but re-writing in terms of norms shows that it extends to the $p = \infty$ case without change. ∎

The *ultrametric Gromov-Wasserstein* distances were introduced by Mémoli et al. (2023) as a way to compare **ultrametric measure spaces**—that is, measure networks (X, ω_X, μ_X) such that the network kernel is a metric on X which additionally satisfies the *strong triangle inequality* $\omega_X(x, x'') \leqslant \max\{\omega_X(x, x'), \omega_X(x', x'')\}$. For $p \in [1, \infty)$ (or with the obvious extension if $p = \infty$), the **ultrametric Gromov-Wasserstein** distance between ultrametric measure spaces X and Y is

$$\mathrm{GW}_{p,\infty}(X,Y) = \frac{1}{2}\left(\iint_{(X\times Y)^2} \Lambda_\infty(\omega_X(x,x'),\omega_Y(y,y'))^p \pi(dx \times dy)\pi(dx' \times dy')\right)^{1/p},$$

where $\Lambda_\infty : \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ is the function,

$$\Lambda_\infty(a,b) = \begin{cases} \max\{a,b\} & a \neq b \\ 0 & a = b. \end{cases} \tag{3}$$

It is easy to check that Λ_∞ is a metric, so that $\mathrm{GW}_{p,\infty} = \mathrm{GW}_p^{(\mathbb{R}_{\geqslant 0}, \Lambda_\infty)}$. We have proved the following:

Proposition 15. *The ultrametric Gromov-Wasserstein distance introduced by Mémoli et al. (2023) is a Z-GW distance with* $(Z, d_Z) = (\mathbb{R}_{\geqslant 0}, \Lambda_\infty)$.

Analogously to (3), one can define a family of functions $\Lambda_q : \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}$, for $q \in [1, \infty)$, by

$$\Lambda_q(a,b) = |a^q - b^q|^{1/q}. \tag{4}$$

This leads to the (p,q)**-Gromov-Wasserstein distance** of Arya et al. (2024):

$$\mathrm{GW}_{p,q}(X,Y) = \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \left(\iint_{(X\times Y)^2} \Lambda_q(\omega_X(x,x'),\omega_Y(y,y'))^p \pi(dx\times dy)\pi(dx'\times dy') \right)^{1/p} \tag{5}$$

(defined here for $p < \infty$, with the $p = \infty$ version defined similarly). The functions Λ_q are metrics on $\mathbb{R}_{\geqslant 0}$ (Arya et al., 2024, Proposition 1.13) and we clearly have $\mathrm{GW}_{p,q} = \mathrm{GW}_p^{(\mathbb{R}_{\geqslant 0},\Lambda_q)}$. Thus we have shown:

Proposition 16. *The (p,q)-Gromov-Wasserstein distance of Arya et al. (2024) is a Z-GW distance with $(Z,d_Z) = (\mathbb{R}_{\geqslant 0},\Lambda_q)$.*

Remark 17. *Sturm (2023) considers a distance very similar to the (p,q)-GW distance defined in* (5)*. The distinction is that the integrand in Sturm's version is raised to a qth power (i.e., it integrates the function $\Lambda_q(\omega_X(\cdot,\cdot),\omega_Y(\cdot,\cdot))^{qp}$). As was observed by Arya et al. (2024), this version does not enjoy the same homogeneity under scaling that* (5) *does. Moreover, we prefer the formulation of* (5)*, since Λ_q defines a metric, but Λ_q^q does not.*

3.1.2 Attributed Graphs

In practice, graph data sets are commonly endowed with additional attributes. As a concrete example (from Kawano et al. 2024), consider a graphical representation of a molecule, where nodes represent atoms and edges represent bonds. Each node and edge is then naturally endowed with a categorical label—the atom type and bond type, respectively. This information can be one-hot encoded, leading to an $\mathbb{R}^n$-valued label on each node and an $\mathbb{R}^m$-valued label on each edge (n and m being the number of atom and bond types present in the data set, respectively). Attributed graphs are also ubiquitous in modeling social networks (see, e.g., Bothorel et al., 2015), where node attributes encode statistics about members of the network and edge attributes can record diverse information about member interactions—in this setting, data typically includes (non-binary) attributes in Euclidean spaces. Graphs with attributes in more exotic metric spaces are also of interest—we describe examples from Bal et al. (2022) and Robertson et al. (2023) in detail below, in Section 3.2.

A general model for graphs with metric space-attributed nodes and edges is introduced in the work of Yang et al. (2023). Following their ideas, we consider the following structure.

Definition 18 (Attributed Network)**.** *We work with the following collection of hyperparameters $\mathcal{H} = (p,\Omega,d_\Omega,\Psi,d_\Psi)$, where $p \in [1,\infty]$, and (Ω,d_Ω), (Ψ,d_Ψ) are separable metric spaces. An* ***attributed network with hyperparameter data*** $\mathcal{H}$*, or simply* $\mathcal{H}$***-network****, is a 5-tuple of the form $(X,\psi_X,\phi_X,\omega_X,\mu_X)$, where*

- *the triple (X,ϕ_X,μ_X) is a measure network (valued in $\mathbb{R}$), with $\phi_X \in L^p(\mu_X \otimes \mu_X)$, which models the underlying graph of an attributed graph,*
- *$\psi_X \in L^p(X,\mu_X;\Psi)$ models node features attributed in Ψ, and*
- *$\omega_X \in L^p(X\times X,\mu_X\otimes\mu_X;\Omega)$ models edge features attributed in Ω.*

Remark 19. *The definition above is more general than the one given by Yang et al. (2023, Definition 2.1): therein, all functions are assumed to be bounded and continuous. Moreover, Yang et al. (2023) does not assume separability of the target metric spaces, which can potentially lead to technical issues of well-definedness, following the discussion in Section 2.1.3.*

A notion of distance between $\mathcal{H}$-networks is also given by Yang et al. (2023), as we now recall.

Definition 20 (Fused Network Gromov-Wasserstein Distance, Yang et al. (2023))**.** *For fixed hyperparameter data $\mathcal{H} = (p, \Omega, d_\Omega, \Psi, d_\Psi)$, let*

$$X = (X, \psi_X, \phi_X, \omega_X, \mu_X) \quad \text{and} \quad Y = (Y, \psi_Y, \phi_Y, \omega_Y, \mu_Y)$$

be $\mathcal{H}$-networks. Given additional hyperparameters $q \in [1, \infty)$ and $\alpha, \beta \in [0, 1]$ with $\alpha+\beta \leqslant 1$, the associated ***fused network Gromov-Wasserstein (FNGW) distance*** *is*

$$\mathrm{FNGW}^{\mathcal{H}}_{q,\alpha,\beta}(X, Y) = \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X, \mu_Y)} \bigg(\int_{X\times Y} \int_{X\times Y} ((1 - \alpha - \beta)\, d_\Psi(\psi_X(x), \psi_Y(y))^q \\ + \alpha\, d_\Omega(\omega_X(x, x'), \omega_Y(y, y'))^q + \beta\, |\phi_X(x, x') - \phi_Y(y, y')|^q)^{p/q}\, \pi(dx \times dy)\pi(dx' \times dy') \bigg)^{1/p} \tag{6}$$

The definition extends to the case $p = \infty$ or $q = \infty$ cases straightforwardly.

Remark 21. *The formulation given in* (6) *is actually subtly different than the one given by Yang et al. (2023). In analogy with Remark 17, we have added an extra power of* $1/q$ *to the integrand. We will see below that this leads to improved theoretical properties of the distance.*

Observe that the distance (6) specializes to other notions of distance in the literature through appropriate choices of the balance parameters α and β:

1. Taking $\alpha = \beta = 0$, one obtains the Wasserstein p-distance on (Ψ, d_Ψ) between the pushforward measures $(\psi_X)_*\mu_X$ and $(\psi_Y)_*\mu_Y$.

2. Taking $\alpha = 1$, $\beta = 0$, we recover the Z-GW p-distance (Definition 11) between the underlying Z-networks, with $(Z, d_Z) = (\Omega, d_\Omega)$. Conversely, we show below that FNGW can be formulated as a special case of a Z-GW distance (Proposition 23).

3. Taking $\alpha = 0$, $\beta = 1$, recovers the standard GW distance between the underlying measure networks.

4. Finally, taking general $\alpha = 0$ and $\beta \in [0, 1]$, we get the **Fused Gromov-Wasserstein (FGW) distance**, which was introduced by Vayer et al. (2020a) as a framework for comparing graphs with only node attributes.

Remark 22. *The original formulation of FGW distance also omitted the q-th root in the integrand that was discussed in Remark 21, but we argue for its inclusion for theoretical reasons discussed below. The FGW framework introduced by Vayer et al. (2020a) is also slightly more general, in that it considers a joint distribution on $X \times \Psi$, rather than a function $\psi_X : X \to \Psi$; the formulation considered here is the one most commonly used in practice—see (Vayer et al., 2019; Flamary et al., 2021).*

We now show that the rather complicated formulation of (6) is somewhat redundant.

Proposition 23. *An attributed network with hyperparameter data $\mathcal{H}$ has a natural representation as a (Z,p)-network, for an appropriate choice of (Z,d_Z). Moreover, the Fused Network Gromov-Wasserstein distance* (6) *can be realized as a Z-Gromov-Wasserstein distance.*

Proof Let hyperparameters $\mathcal{H} = (p,\Omega,d_\Omega,\Psi,d_\Psi)$, as well as $q \in [1,\infty)$ and $\alpha,\beta \in [0,1]$ with $\alpha+\beta \leqslant 1$ be given. Suppose that $X = (X,\psi_X,\phi_X,\omega_X,\mu_X)$ and $Y = (Y,\psi_Y,\phi_Y,\omega_Y,\mu_Y)$ are $\mathcal{H}$-networks. We construct a metric space Z as follows. Let $Z = \Psi \times \Omega \times \mathbb{R}$ and let d_Z be the weighted ℓ^q metric

$$d_Z((a,b,c),(a',b',c')) = \left((1-\alpha-\beta)d_\Psi(a,a')^q + \alpha d_\Omega(b,b')^q + \beta|c-c'|^q\right)^{1/q}.$$

Now, we define a Z-network $\overline{X} = (X,\overline{\omega}_X,\mu_X)$ by setting

$$\overline{\omega}_X(x,x') = \left(\psi_X(x),\omega_X(x,x'),\phi_X(x,x')\right) \in \Psi \times \Omega \times \mathbb{R} = Z.$$

We define $\overline{Y} = (Y,\overline{\omega}_Y,\mu_Y)$ similarly. Then, for any $\pi \in \mathcal{C}(\mu_X,\mu_Y)$, we have

$$\begin{aligned}\iint & d_Z(\overline{\omega}_X(x,x'),\overline{\omega}_Y(y,y'))^p \pi(dx\times dy)\pi(dx'\times dy') \\ &= \iint \big((1-\alpha-\beta)d_\Psi(\psi_X(x),\psi_Y(y))^q + \alpha d_\Omega(\omega_X(x,x'),\omega_Y(y,y'))^q \\ &\qquad + \beta|\phi_X(x,x')-\phi_Y(y,y')|^q\big)^{p/q}\,\pi(dx\times dy)\pi(dx'\times dy'),\end{aligned}$$

and it follows that $\mathrm{GW}_p^Z(\overline{X},\overline{Y}) = \mathrm{FNGW}^{\mathcal{H}}_{q,\alpha,\beta}(X,Y)$. ∎

3.1.3 Diffusions and Stochastic Processes

Inspired by a notion of spectral convergence for Riemannian manifolds introduced by Kasue and Kumura (1994), Mémoli (2009, 2011b) considers a certain *spectral* version of the GW distance between Riemannian manifolds (see S. Lim's PhD thesis, Lim 2021, Chapter 5, for a generalization to Markov processes and Chen et al. 2022, 2023 for related ideas). For a given compact Riemannian manifold (M,g_M) let $k_M : M \times M \times \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ denote its normalized heat kernel[1] and μ_M its normalized volume measure (we use $\mathbb{R}_{>0}$ to denote the set of positive real numbers). Then, if N is another compact Riemannian manifold and $p \geqslant 1$, the **spectral Gromov-Wasserstein p-distance** between M and N is

$$\mathrm{GW}_p^{\mathrm{spec}}(M,N) := \frac{1}{2}\inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)}\sup_{t>0} c^2(t)\cdot\|\Gamma^{\mathrm{spec}}_{M,N,t}\|_{L^p(\pi\otimes\pi)}$$

where $c(t) := e^{-t^{-1}}$ and $\Gamma^{\mathrm{spec}}_{M,N,t}(x,y,x',y') := \left|k_M(x,x',t) - k_N(y,y',t)\right|$ for $x,x' \in M$ and $y,y' \in Y$. The reason for the use of the dampening function $c(t)$ is to tame the blow up of the

1. i.e., $\lim_{t\to\infty} k_M(x,x',t) = 1$ for all $x,x' \in M$.

heat kernel as $t \to 0$: $k_M(x, x', t) \sim t^{-\frac{m}{2}}$ where m is the dimension of M; see (Rosenberg, 1997).

To recast $\mathrm{GW}_p^{\mathrm{spec}}$ (or, rather, a variant thereof—see below) as a Z-GW distance, we make the following choices:

$$Z = \left\{ f : \mathbb{R}_{>0} \to \mathbb{R}_{>0} \mid \sup_{t>0} c^2(t)|f(t) - 1| < \infty \right\} \tag{7}$$

and, for $f_1, f_2 \in Z$,

$$d_Z(f_1, f_2) = \sup_{t>0} c^2(t)\big|f_1(t) - f_2(t)\big|. \tag{8}$$

We then represent a compact Riemannian manifold (M, g_M) as a Z-measure network (M, ω_M, μ_M) with μ_M as above and $\omega_M : M \times M \to Z$ defined by

$$\omega_M(x, x') = k_M(x, x', \cdot) \in Z, \quad \text{where} \quad k_M(x, x', \cdot) : t \mapsto k_M(x, x', t).$$

Then

$$\begin{aligned} d_Z \circ (\omega_M \times \omega_N)(x, y, x', y') &= \sup_{t>0} c^2(t)\big|k_M(x, x', t) - k_N(y, y', t)\big| \\ &= \sup_{t>0} c^2(t)\Gamma^{\mathrm{spec}}_{M,N,t}(x, y, x', y'), \end{aligned}$$

so that, for an arbitrary coupling $\pi \in \mathcal{C}(\mu_X, \mu_Y)$,

$$\|d_Z \circ (\omega_M \times \omega_N)\|_{L^p(\pi\times\pi)} = \left\| \sup_{t>0} c^2(t)\Gamma^{\mathrm{spec}}_{M,N,t} \right\|_{L^p(\pi\otimes\pi)}.$$

Accordingly, we define a variant of the spectral GW p-distance:

$$\widetilde{\mathrm{GW}}_p^{\mathrm{spec}}(M, N) := \frac{1}{2} \inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)} \left\| \sup_{t>0} c^2(t)\Gamma^{\mathrm{spec}}_{M,N,t} \right\|_{L^p(\pi\otimes\pi)}.$$

We have the following result.

Proposition 24. *The spectral GW p-distance $\widetilde{\mathrm{GW}}_p^{\mathrm{spec}}$ is a Z-GW distance, with (Z, d_Z) as in* (7) *and* (8)*, which upper bounds the spectral Gromov-Wasserstein distance introduced by Mémoli (2009).*

Proof The fact that $\widetilde{\mathrm{GW}}_p^{\mathrm{spec}}$ is a Z-GW distance follows immediately by definition. To verify the upper bound claim, let M and N be compact Riemannian manifolds and choose an arbitrary coupling π between their normalized volume measures. We then consider the map

$$\begin{aligned} F : (X \times Y \times X \times Y) \times \mathbb{R}_{>0} &\to \mathbb{R} \\ \big((x, y, x', y'), t\big) &\mapsto c^2(t)\Gamma^{\mathrm{spec}}_{M,N,t}(x, y, x', y') \end{aligned}$$

as a function on the product of measure spaces $(X \times Y \times X \times Y, \pi \otimes \pi)$ and $(\mathbb{R}, \mathcal{L})$, where we use $\mathcal{L}$ to denote Lebesgue measure on $\mathbb{R}_{>0}$ through the duration of the proof. We now compare the objective functions of the optimization problems associated to the distances $\widetilde{\mathrm{GW}}_p^{\mathrm{spec}}(M,N)$ and $\mathrm{GW}_p^{\mathrm{spec}}(M,N)$ as

$$\begin{aligned}\left\|\sup_{t>0} c^2(t)\Gamma^{\mathrm{spec}}_{M,N,t}\right\|_{L^p(\pi\otimes\pi)} &= \left\|\|F\|_{L^\infty(\mathcal{L})}\right\|_{L^p(\pi\otimes\pi)} \\ &\geqslant \left\|\|F\|_{L^p(\pi\otimes\pi)}\right\|_{L^\infty(\mathcal{L})} \\ &= \sup_{t>0}\|c^2(t)\Gamma^{\mathrm{spec}}_{M,N,t}\|_{L^p(\pi\otimes\pi)} \\ &= \sup_{t>0} c^2(t)\cdot\|\Gamma^{\mathrm{spec}}_{M,N,t}\|_{L^p(\pi\otimes\pi)},\end{aligned}$$

where we have interchanged suprema and L^∞ norms via continuity in t, and we have applied a generalized version of Minkowski's inequality (Bahouri, 2011, Proposition 1.3). Since the estimate holds for an arbitrary coupling π, this shows that $\widetilde{\mathrm{GW}}_p^{\mathrm{spec}}(M,N) \geqslant \mathrm{GW}_p^{\mathrm{spec}}(M,N)$. ■

3.1.4 The Gromov-Wasserstein Distance Between Dynamic Metric Spaces

A **dynamic metric space** (or DMS, for short) is a pair (X, d_X), where X is a finite set and $d_X : \mathbb{R}\times X\times X \to \mathbb{R}_{\geqslant 0}$ is such that:

- for every $t\in\mathbb{R}$, $(X, d_X(t))$ is a pseudometric space, where $d_X(t) : X\times X\to\mathbb{R}$ is the map $d_X(t)(x,x') = d_X(t,x,x')$;
- for any $x,x'\in X$ with $x\neq x'$ the function $d_X(\cdot)(x,x') : \mathbb{R}\to\mathbb{R}_{\geqslant 0}$ defined by $t\mapsto d_X(t)(x,x')$ is continuous and not identically zero.

A **weighted dynamic metric space** (or wDMS) is a triple (X, d_X, μ_X) such that (X, d_X) is a DMS and μ_X is a fully supported probability measure on X.

Dynamic metric spaces provide a mathematical structure suitable for modeling natural phenomena such as flocking and swarming behaviors (Sumpter, 2010). Variants of the Gromov-Hausdorff and Gromov-Wasserstein distances were proposed by Kim et al. (2020); Kim and Mémoli (2021); Kim (2020) in order to metrize the collection of all DMSs and wDMSs, respectively. We now show how wDMSs and these distances fit into the framework of Z-GW distances.

Let $C(\mathbb{R},\mathbb{R}_{\geqslant 0})$ denote the collection of all continuous functions $f:\mathbb{R}\to\mathbb{R}_{\geqslant 0}$. For $\lambda\geqslant 0$ one defines (see Kim, 2020, Definition 2.7.1) the **λ-slack interleaving distance** between $f_1, f_2 \in C(\mathbb{R},\mathbb{R}_{\geqslant 0})$ by

$$d_\lambda(f_1,f_2) := \inf\left\{\varepsilon\in[0,\infty] | \forall t\in\mathbb{R},\ \min_{s\in[t-\varepsilon,t+\varepsilon]} f_i(s)\leqslant f_j(t)+\lambda\varepsilon,\ i,j=1,2\right\}.$$

According to Kim (2020, Definition 2.9.5), the (p,λ)-Gromov-Wasserstein distance between two wDMSs (X,d_X,μ_X) and (Y,d_Y,μ_Y) is defined as

$$\frac{1}{2}\inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)}\left\|d_\lambda\circ(d_X\times d_Y)\right\|_{L^p(\pi\otimes\pi)}$$

(the original definition did not include the 1/2-scaling factor, but we include it here for consistency). We consider a wDMS (X, d_X, μ_X) as a Z-network with $Z = C(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ by defining a network kernel $\omega_X : X \times X \to Z$ as $\omega_X(x, x') = d_X(\cdot)(x, x')$. This immediately yields the following.

Proposition 25. *The (p, λ)-Gromov-Wasserstein distance between wDMSs (Kim, 2020) is a Z-GW distance with $(Z, d_Z) = (C(\mathbb{R}, \mathbb{R}_{\geqslant 0}), d_\lambda)$.*

3.1.5 The Cut Distance Between Graphons

Recall that in the work of Borgs, Chayes, Lovász, Sós and Vesztergombi (Borgs et al., 2008) (see also Lovász 2012; Janson 2010), a *kernel* on a probability space (Ω, μ) is any symmetric integrable function $U : \Omega \times \Omega \to \mathbb{R}_{\geqslant 0}$. A *graphon* on (Ω, μ) is any kernel W with codomain $[0, 1]$. The *cut norm* of $W \in L^1(\Omega, \mu)$ is the number

$$\|W\|_{\square,\Omega,\mu} := \sup_{S,T \subset \Omega} \left| \iint_{S \times T} W(x, x')\, \mu(dx)\, \mu(dx') \right|,$$

where the supremum is over measurable sets S, T.

Now, given two probability spaces (Ω, μ) and (Ω', μ'), and kernels U on Ω and U' on Ω' one defines the *cut distance* between (Ω, U, μ) and (Ω', U', μ) as (Janson, 2010, Theorem 6.9)

$$d_\square(\Omega, \Omega') := \inf_{\pi \in \mathcal{C}(\mu, \mu')} \left\| U - U' \right\|_{\square, \Omega \times \Omega', \pi}.$$

We note that the cut distance defines a compact metric space, while GW-type metrics do not in general exhibit completeness (see Mémoli 2011a, Remark 5.18), and therefore do not lead to compactness. In particular, the δ_1 metric considered by Janson in (Janson, 2010) is precisely a Z-Gromov-Wasserstein distance for $Z = [0, 1]$ and it is not compact, as mentioned in Janson (2010, p15). In fact, the topology induced by δ_1 is strictly finer than the one induced by the cut distance. The structural discrepancy lies in how kernel values are compared: the Z-Gromov-Wasserstein distance aggregates kernel values ω_X, ω_Y via the distance function $d_Z(\omega_X, \omega_Y)$ and then considers the L^p norm of this quantity. On the other hand, the cut distance aggregates kernel values by taking their differences and then computing the cut norm.

While not exactly fitting into our framework, it seems interesting to explore the possibility of formulating an even more general setting than the one in Definition 11 which could (precisely) encompass the cut distance. For example, considering both an abstract "aggregator" of kernel values and an arbitrary norm for the resulting *aggregated* quantity could lead to a framework subsuming the Z-Gromov-Wasserstein distance as well as the cut distance.

3.2 Further Examples of Z-Valued Measure Networks

Next, we give several more examples of naturally occurring Z-valued measure networks. We believe that the resulting Z-GW distances are novel, but point to related notions in the literature when appropriate.

3.2.1 An Alternative Approach to Edge-Attributed Graphs

In Definition 18, we gave a flexible model for graphs with node and edge attributions, which we referred to as *attributed networks*, following Yang et al. (2023). However, the formalism developed there may be unsatisfactory in practice, as we now explain. Consider a directed graph (X, E), where X is a set of nodes and $E \subset X \times X$ is a set of directed edges (this also captures the notion of an undirected graph, which we consider as a directed graph with a symmetric edge set). Realistic graphical data frequently comes with edge attributes (see the discussion at the beginning of Section 3.1.2)—in practice, this is given by a function from E into some attribute space (Ω, d_Ω). Now observe that the formalism described in Definition 18 (which originates from Yang et al. 2023) models edge attributes as functions of the form $\omega_X : X \times X \to \Omega$; that is, the function is not only defined on the edge set, but on the full product space $X \times X$. This difference can be handled in an ad hoc manner: for example, if $\Omega = \mathbb{R}^d$, one could assign pairs $(x, x') \in (X \times X) \backslash E$ to the zero vector. Such a choice is essentially arbitrary, and does not naturally extend to a more general metric space Ω. We now describe a more principled approach to handling this issue via *cone metrics*.

Let (Ω, d_Ω) be a metric space, which we will later consider as an edge attribute space. The **cone space** (Burago et al., 2022, Section 3.6.2) of Ω is

$$\mathrm{Con}(\Omega) := (\Omega \times \mathbb{R}_{\geqslant 0})/(\Omega \times \{0\}).$$

For $(u, r) \in \Omega \times \mathbb{R}_{\geqslant 0}$, its equivalence class $[u, r]$ consists of points (v, s) such that $(u, r) = (v, s)$ or $r = s = 0$. The **cone metric** on $\mathrm{Con}(\Omega)$ is

$$d_{\mathrm{Con}(\Omega)}([u, r], [v, s]) := \left(r^2 + s^2 - 2rs\cos(\bar{d}_\Omega(u, v))\right)^{1/2},$$

where $\bar{d}_\Omega(u, v) = \min\{d_\Omega(u, v), \pi\}$.

Now let (X, E) be a directed graph endowed with an edge attribute function $\hat{\omega}_X : E \to \Omega$ and a network kernel $\phi_X : X \times X \to \mathbb{R}_{\geqslant 0}$ such that $\phi_X(x, x') > 0$ if and only if $(x, x') \in E$; this can either be data-driven, or derived directly from the combinatorial structure of the graph by taking ϕ_X to be the directed binary adjacency function induced by E. Let $(Z, d_Z) = (\mathrm{Con}(\Omega), d_{\mathrm{Con}(\Omega)})$. For any choice of probability measure μ_X, the associated **Z-network induced by** (X, E) is (X, ω_X, μ_X), with $\omega_X : X \times X \to \mathrm{Con}(\Omega)$ defined by

$$\omega_X(x, x') := \begin{cases} [\hat{\omega}(x, x'), \phi_X(x, x')] & \text{if } (x, x') \in E \\ [p_0, 0] & \text{otherwise,} \end{cases}$$

where $p_0 \in \Omega$ is a fixed but arbitrary point. The idea is that this extends the original attribute function $\hat{\omega}_X$ (whose domain is E) to the full product space $X \times X$, but the cone construction results in the choice of assignment for $(x, x') \notin E$ being inconsequential. Node attributes for the graph can be handled as in Definition 18, so that the structure defined here gives a complete refinement of the previous attributed network structure.

The next two subsections provide examples in the literature of Z-networks arising from attributed graph structures. In particular, a similar idea was used by Bal et al. (2024) to model *shape graphs*, described in detail below, where this setup was implemented computationally.

3.2.2 Shape Graphs

There is a growing body of work on statistical analysis of *shape graphs* (Sukurdeep et al., 2022; Liang et al., 2023; Guo et al., 2022; Srivastava et al., 2020; Bal et al., 2024)—roughly, these are 1-dimensional stratified spaces embedded in some Euclidean space, relevant for modeling filamentary structures such as arterial systems or road networks (see Figure 2). We now explain how some models of shape graphs fit into the Z-GW framework.

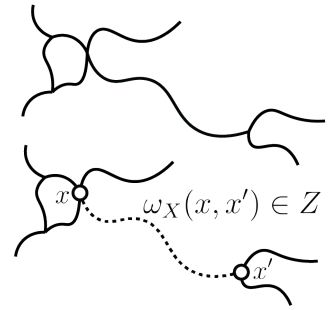


Figure 2: Top: Simple example of a shape graph in $\mathbb{R}^2$. Bottom: Representation as an attributed network. The intersection points x and x' (circles) are nodes in X, $\omega_X(x, x')$ (dashed) is an element of the space of curves Z.

Let Z denote the *space of unparameterized curves* in $\mathbb{R}^d$, where $d \in \{2, 3\}$ in typical applications; that is, $Z = C^\infty([0,1], \mathbb{R}^d)/\mathrm{Diff}([0,1])$, the space of smooth maps of the interval into $\mathbb{R}^d$, denoted $C^\infty([0,1], \mathbb{R}^d)$, considered up to the reparameterization action of the diffeomorphism group of the interval, denoted $\mathrm{Diff}([0,1])$. Let d_Z be some fixed metric on Z; in the references considered here, this is typically induced by a diffeomorphism-invariant *elastic metric* (Srivastava and Klassen, 2016; Bauer et al., 2024), but the specific choice of metric is not important in the discussion to follow.

Guo et al. (2022) modeled a shape graph as a function of the form $\omega_X : X \times X \to Z$, where X is a finite set of nodes. In practice, given a filamentary structure such as an arterial system, the nodes are the endpoints and intersection points of filaments and the value of $\omega_X(x, x') \in Z$ is the (unparameterized) filament joining the endpoints, as is illustrated in Figure 2. For pairs (x, x') which are not joined by a curve, the authors choose to assign the constant curve taking the value $0 \in \mathbb{R}^d$. In an effort to overcome the shortcomings of this ad hoc assignment, the followup work (Bal et al., 2024) introduces a formalism similar to that of Section 3.2.1 for representing shape graphs, where the particular point in Z assigned to missing edges is made irrelevant to distance computations via a quotient construction. Either method for modeling shape graphs results in a Z-network by, say, assigning the node set the uniform measure.

In both Guo et al. (2022) and Bal et al. (2024), comparisons of shape graphs were done via the Jain and Obermayer framework (Jain and Obermayer, 2009) described earlier. Specifically, for shape graphs with the same number of nodes, the distance between them is computed by aligning the Z-valued kernels over permutations. When comparing two shape graphs with node sets of different sizes, the smaller of the two is "padded" with additional disconnected nodes. This padding step is arguably unnatural, and can be avoided by instead working within the Z-GW framework, where differently sized node sets are not an issue.

3.2.3 Connection Graphs

Let (X, E) be a directed graph such that $(x, x') \in E \Leftrightarrow (x', x) \in E$ (i.e., (X, E) uses the directed graph formalism to encode an inherently undirected graph). A **connection** on (X, E) (Robertson et al., 2023) is a map $\hat{\omega}_X : E \to \mathrm{O}(d)$ such that $\hat{\omega}(x, x') = \hat{\omega}(x', x)^{-1}$, where $\mathrm{O}(d)$ is the orthogonal group of Euclidean space $\mathbb{R}^d$. These structures appear, for example, in alignment problems coming from cryo-electron microscopy (Bhamre et al., 2015; Singer, 2011). Let $d_{\mathrm{O}(d)}$ be the metric on $\mathrm{O}(d)$ induced by identifying it with a subspace of the matrix space $\mathbb{R}^{d\times d}$, endowed with Frobenius distance. For a choice of measure μ_X, one obtains a Z-network (X, ω_X, μ_X), with $(Z, d_Z) = (\mathrm{O}(d), d_{\mathrm{O}(d)})$ and with $\omega_X : X \times X \to Z$ defined as in Section 3.2.1.

3.2.4 Binary Operations

Let (Z, d_Z) be a metric space endowed with a binary operation $\bullet : Z \times Z \to Z$. For any subset $X \subset Z$, let $\omega_X : X \times X \to Z$ be defined by $\omega_X(x, x') = x \bullet x'$. A choice of probability measure μ_X results in a Z-network (X, ω_X, μ_X).

As an example of the above construction, suppose that Z is a compact Lie group endowed with a bi-invariant Riemannian metric, and let $\bullet$ denote its group law. Given a compact Lie subgroup $X \subset Z$, let μ_X denote its normalized Haar measure. The Z-GW distance then defines a natural distance on the space of compact Lie subgroups of Z.

3.2.5 Probabilistic Metric Spaces

In his 1942 paper (Menger, 1942), Menger initiated the study of *probabilistic metric spaces*; see also (Wald, 1943; Kramosil and Michálek, 1975; Schweizer and Sklar, 1960). A **probabilistic metric space** is a pair (X, p_X) in which the "distance" $p_X(x, x')$ between any two points $x, x' \in X$ is a Borel probability measure on $\mathbb{R}_{\geqslant 0}$ satisfying the following axioms:

1. $x = x'$ if and only if $p_X(x, x') = \delta_0$;
2. $p_X(x, x') = p_X(x', x)$ for all $x, x' \in X$;
3. $\min\left(p_X(x, x')([0, s]), p_X(x', x'')([0, t])\right) \leqslant p_X(x, x'')([0, s + t])$ for all $x, x', x'' \in X$ and all $s, t \in \mathbb{R}_{\geqslant 0}$.

Note that axiom 3 is a generalization of the triangle inequality: if d_X is a metric on X and distances are *deterministic*, i.e., $p_X(x, x') = \delta_{d_X(x,x')}$ for all x, x', then this condition is equivalent to the triangle inequality for d_X.

We can recast probabilistic metric spaces as Z-networks for the choice

$$Z = \{\text{Borel probability measures on } \mathbb{R}_{\geqslant 0} \text{ satisfying axioms 1,2, and 3}\}$$

and for $d_Z = \mathrm{W}_p$, the Wasserstein distance (of order $p \geqslant 1$) on $\mathbb{R}_{\geqslant 0}$. Probabilistic metric spaces can therefore be compared via the associated Z-GW distance.

4. Properties of Z-Gromov-Wasserstein Distances

This section establishes basic metric properties of the Z-GW distances, as well as the induced topological properties of the space of (Z, p)-networks. As a result, we obtain a unified proof that these properties hold for the distances from the literature described in Section 3.1 (summarized in Theorem 12 and Table 1), and a simultaneous proof for the distances in the novel settings described in Section 3.2.

4.1 Basic Metric Structure

We begin by showing that Z-GW distances are metrics when the space of (Z, p)-networks is considered up to a natural notion of equivalence.

4.1.1 Existence of Optimal Couplings

In this section, we will study the important question of whether the optimization problem defining the Z-GW distance always has a solution. Our main result of this part is the following theorem, which gives an affirmative answer in the general setting of the present article:

Theorem 26. *For any Z-networks X, Y and any $p \in [1, \infty]$, there exists $\pi \in \mathcal{C}(\mu_X, \mu_Y)$ such that $\mathrm{GW}_p^Z(X, Y) = \frac{1}{2}\mathrm{dis}_p^Z(\pi)$. That is, optimal couplings always exist.*

Our proof strategy is as follows: since $\mathcal{C}(\mu_X, \mu_Y)$ is sequentially compact with respect to the weak convergence of probability measures (Villani, 2003, p. 32), it is enough to prove that the distortion functional dis_p^Z is continuous in this topology. This approach was also considered in previous works such as Chowdhury and Mémoli (2019) and Mémoli (2011a), but the same results do not apply in our general setting. For example, Mémoli (2011a) heavily utilizes the fact that the kernels are metrics, and Chowdhury and Mémoli (2019) use the fact that the bounded continuous $\mathbb{R}$-valued functions are dense in L^p spaces, which is not true for a general target metric space (consider $Z = \{0, 1\}$). The corresponding result for the fused network GW distances (Definition 20) was established by Yang et al. (2023). In light of Proposition 23, and the fact that Yang et al. (2023) considered more restrictive structures with bounded and continuous kernels, our result further generalizes the result of Yang et al. (2023, Theorem 2.4).

To deal with our case, we utilize the following technical lemma, which generalizes the result by Santambrogio (2015, Lemma 1.8) to the Gromov-Wasserstein setting. The proof appears in Section A.2.

Lemma 27. *Let X and Y be Polish spaces endowed with probability measures μ and ν, respectively, and suppose $\gamma_n, \gamma \in \mathcal{C}(\mu, \nu)$. Let $a : X \times X \to \tilde{X}$ and $b : Y \times Y \to \tilde{Y}$ be measurable maps valued in separable metric spaces $\tilde{X}$ and $\tilde{Y}$. Finally, let $c : \tilde{X} \times \tilde{Y} \to [0, \infty)$ be a continuous function such that $c(a, b) \leqslant f(a) + g(b)$ for some continuous maps f and g satisfying $\int (f \circ a) d(\mu \otimes \mu), \int (g \circ b) d(\nu \otimes \nu) < +\infty$. Then, $\gamma_n \to \gamma$ weakly implies*

$$\int_{X\times Y} \int_{X\times Y} c(a(x, x'), b(y, y'))\gamma_n(dx \times dy)\gamma_n(dx' \times dy')$$
$$\to \int_{X\times Y} \int_{X\times Y} c(a(x, x'), b(y, y'))\gamma(dx \times dy)\gamma(dx' \times dy')$$

Proof of Theorem 26. As mentioned earlier, it is enough to prove that $\mathcal{C}(\mu_X, \mu_Y) \ni \pi \mapsto \mathrm{dis}_p^Z(\pi)$ is continuous with respect to the topology of weak convergence. We can apply Lemma 27 by setting

$$\tilde{X} = \tilde{Y} = Z, \quad a = \omega_X, \quad b = \omega_Y, \quad \mu = \mu_X, \quad \nu = \mu_Y,$$
$$c(a,b) = d(\omega_X, \omega_Y)^p, \quad f(a) = 2^{p-1} d_Z(a,z)^p, \quad g(b) = 2^{p-1} d_Z(b,z)^p,$$

for some fixed (but arbitrary) point $z \in Z$. The general estimate $(s+t)^p \leqslant 2^{p-1}(s^p + t^p)$, for $s, t \geqslant 0$, combined with the triangle inequality, shows that

$$d(\omega_X, \omega_Y)^p \leqslant 2^{p-1} d_Z(\omega_X, z)^p + 2^{p-1} d_Z(\omega_Y, z)^p,$$

so the assumption $c(a,b) \leqslant f(a) + g(b)$ is satisfied. The integrability of $f \circ a$ and $g \circ b$ is satisfied by the L^p assumption on $a = \omega_X$ and $b = \omega_Y$, so all assumptions are clear. Thus, dis_p is continuous, and an optimal coupling exists for $1 \leqslant p < \infty$. Since the $p = \infty$ case can be seen as the supremum of dis_p for $p < \infty$, we have lower semicontinuity, and an optimal coupling exists for $p = \infty$. ∎

4.1.2 Z-Gromov-Wasserstein Distance Is a Metric

It turns out that the Z-network Gromov-Wasserstein distances are only pseudometrics; that is, a zero distance does not imply that two (Z,p)-networks are equal. Similar to the argument by Chowdhury and Mémoli (2019), the following condition characterizes the equivalence class of the set of networks with zero distance, as we will prove below. Given functions $\omega : Y \times Y \to Z$ and $\phi : X \to Y$, we define the **pullback of ω by ϕ** to be the function $\phi^*\omega : X \times X \to Z$ defined by

$$\phi^*\omega(x, x') = \omega(\phi(x), \phi(x')).$$

Definition 28 (Weak Isomorphism)**.** *A pair of (Z,p)-networks $X = (X, \omega_X, \mu_X)$ and $Y = (Y, \omega_Y, \mu_Y)$ is defined to be **weakly isomorphic** if there exists a (Z,p)-network $W = (W, \omega_W, \mu_W)$, and maps $\phi_X : W \to X$ and $\phi_Y : W \to Y$ such that*

- *ϕ_X and ϕ_Y are measure-preserving, and*
- *$\phi_X^*\omega_X = \phi_Y^*\omega_Y = \omega_W$ $\mu_W \otimes \mu_W$-almost everywhere. That is,*

$$\omega_W(w, w') = \omega_X(\phi_X(w), \phi_X(w')) = \omega_Y(\phi_Y(w), \phi_Y(w'))$$

  for $\mu_W \otimes \mu_W$-almost every pair $(w, w') \in W \times W$.

If X and Y are weakly isomorphic, write $X \sim Y$.

*Let $\mathcal{M}^{Z,p}$ denote the **space of** (Z,p)**-networks**, and let $\mathcal{M}^{Z,p}_\sim$ denote the **space of** (Z,p)**-networks considered up to weak isomorphism** (i.e., the quotient of $\mathcal{M}^{Z,p}$ by the weak isomorphism equivalence relation).*

In the following, we make no distinction between a Z-network (X, ω, μ) and its equivalence class in $\mathcal{M}_{\sim}^{Z,p}$, unless necessary.

We will now prove that the Gromov-Wasserstein distances define metrics up to weak isomorphism:

Theorem 29. *For any separable metric space (Z, d_Z), GW_p^Z induces a metric on $\mathcal{M}_{\sim}^{Z,p}$.*

For the Fused GW (Vayer et al., 2020a) and Fused Network GW (Yang et al., 2023) distances, it was previously only shown that they satisfy "relaxed" triangle inequalities, including an extra scaling term depending on the q parameter—that is, an inequality of the form

$$d_q(x,z) \leqslant 2^{q-1}(d_q(x,y) + d_q(y,z)),$$

where we momentarily use d_q as a placeholder for a generic metric depending on a parameter $q \in [1, \infty)$ (cf. Vayer et al., 2020a, Theorem 1). With our slight reformulation of these distances (see Remark 21 and Remark 22), we have the following corollary of Theorem 29. The result follows immediately by transporting the definition of weak isomorphism into the appropriate context and via Proposition 23 and Theorem 29. In particular, this result strengthens the previous results on relaxed triangle inequalities, promoting them to true triangle inequalities.

Corollary 30. *The Fused GW and Fused Network GW distances are metrics, up to weak isomorphism.*

The proof of Theorem 29 follows the standard strategy (cf. Mémoli 2011a), for which we will need the following well-known lemma:

Lemma 31 (Gluing Lemma, Villani 2003)**.** *Let μ_1, μ_2, μ_3 be three probability measures, supported on Polish spaces X_1, X_2, X_3 respectively, and let $\pi_{12} \in \mathcal{C}(\mu_1, \mu_2), \pi_{23} \in \mathcal{C}(\mu_2, \mu_3)$ be two couplings. Then there exists a probability measure π on $X_1 \times X_2 \times X_3$, with marginals π_{12} on $X_1 \times X_2$ and π_{23} on $X_2 \times X_3$.*

Proof of Theorem 29. First, it is clear that $\mathrm{GW}_p^Z(X, Y) \geqslant 0$ for any Z-networks X, Y. For symmetry, the symmetry of d proves that the mapping $f : X \times Y \to Y \times X, f(x, y) = (y, x)$ defines a bijection $\mathcal{C}(\mu_X, \mu_Y) \ni \pi \mapsto f_*\pi \in \mathcal{C}(\mu_Y, \mu_X)$ such that $\mathrm{dis}_p^Z(\pi) = \mathrm{dis}_p^Z(f_*\pi)$, and taking the infimum shows that $\mathrm{GW}_p^Z(X, Y) = \mathrm{GW}_p^Z(Y, X)$.

We will now prove the triangle inequality. We consider three Z-networks (X_1, ω_1, μ_1), (X_2, ω_2, μ_2) and (X_3, ω_3, μ_3) and take couplings $\pi_{12} \in \mathcal{C}(\mu_1, \mu_2)$ and $\pi_{23} \in \mathcal{C}(\mu_2, \mu_3)$ realizing $\mathrm{GW}_p^Z(X_1, X_2)$ and $\mathrm{GW}_p^Z(X_2, X_3)$, respectively (Theorem 26). By the Gluing Lemma, there exists a probability measure π on $X_1 \times X_2 \times X_3$ with marginals π_{12} and π_{23}. Denote the marginal of π on $X_1 \times X_3$ by π_{13}. Then, by definition, we have $\pi_{13} \in \mathcal{C}(\mu_1, \mu_3)$. Therefore, by Minkowski's inequality, we have

$$\begin{aligned}
2 \cdot \mathrm{GW}_p^Z(X_1, X_3) &\leqslant \|d_Z \circ (\omega_1 \times \omega_3)\|_{L^p(\pi_{13}\otimes\pi_{13})} && (9)\\
&= \|d_Z \circ (\omega_1 \times \omega_3)\|_{L^p(\pi\otimes\pi)} && (10)\\
&\leqslant \|d_Z \circ (\omega_1 \times \omega_2) + d_Z \circ (\omega_2 \times \omega_3)\|_{L^p(\pi\otimes\pi)} && (11)\\
&\leqslant \|d_Z \circ (\omega_1 \times \omega_2)\|_{L^p(\pi\otimes\pi)} + \|d_Z \circ (\omega_2 \times \omega_3)\|_{L^p(\pi\otimes\pi)} && (12)\\
&= \|d_Z \circ (\omega_1 \times \omega_2)\|_{L^p(\pi_{12}\otimes\pi_{12})} + \|d_Z \circ (\omega_2 \times \omega_3)\|_{L^p(\pi_{23}\otimes\pi_{23})} && (13)\\
&= \mathrm{dis}_p^Z(\pi_{12}) + \mathrm{dis}_p^Z(\pi_{23})\\
&= 2 \cdot \big(\mathrm{GW}_p^Z(X_1, X_2) + \mathrm{GW}_p^Z(X_2, X_3)\big), && (14)
\end{aligned}$$

where we have used suboptimality of π_{13} in (9), marginal conditions on π in (10) and (13), the triangle inequality for d_Z in (11), Minkowski's inequality in (12), and the optimality of π_{12} and π_{23} in (14).

We finally prove that, for Z-networks X and Y, $\mathrm{GW}_p^Z(X,Y) = 0$ if and only if X and Y are weakly isomorphic. We will restrict our attention to the case $1 \leqslant p < \infty$ since we can prove the $p = \infty$ case by simply replacing the integrals with essential suprema. Suppose that X and Y are weakly isomorphic. Then, there exists a Z-network (W, ω_W, μ_W) and measure-preserving maps $\phi_X : W \to X$ and $\phi_Y : W \to Y$ such that $\phi_X^* \omega_X = \phi_Y^* \omega_Y = \omega_W$ $\mu_W \otimes \mu_W$-almost everywhere. By definition of ϕ_X and ϕ_Y, $\pi_{XY} = (\phi_X, \phi_Y)_* \mu_W$ belongs to $\mathcal{C}(\mu_X, \mu_Y)$, and

$$\begin{aligned}
\mathrm{dis}_p^Z(\pi)^p &= \int_{X\times Y}\int_{X\times Y} d_Z(\omega_X(x,x'), \omega_Y(y,y'))^p \pi(dx \times dy)\pi(dx' \times dy')\\
&= \int_W\int_W d_Z(\omega_X(\phi_X(dw), \phi_X(dw')), \omega_Y(\phi_Y(dw), \phi_Y(dw')))^p \mu_W(dw)\mu_W(dw')\\
&= \int_W\int_W d_Z(\phi_X^*\omega_X(w,w'), \phi_Y^*\omega_Y(w,w'))^p \mu_W(dw)\mu_W(dw')\\
&= \int_W\int_W d_Z(\omega_W(w,w'), \omega_W(w,w'))^p \mu_W(dw)\mu_W(dw') = 0.
\end{aligned}$$

Therefore, $\mathrm{GW}_p^Z(X,Y) = 0$. To prove the other direction, suppose that $\mathrm{GW}_p^Z(X,Y) = 0$. By Theorem 26, there exists $\pi \in \mathcal{C}(X \times Y)$ such that $\mathrm{dis}_p^Z(\pi) = 0$. Now, define measure-preserving maps $\phi_X : X \times Y \to X, \phi_Y : X \times Y \to Y$ as the projection onto X and Y, respectively, and define a Z-network by $(W, \omega_W, \mu_W) = (X \times Y, \phi_X^*\omega_X, \pi)$. Now, we have

$$\begin{aligned}
0 = \mathrm{dis}_p^Z(\pi)^p &= \int_{X\times Y}\int_{X\times Y} d_Z(\omega_X(x,x'), \omega_Y(y,y'))^p \pi(dx \times dy)\pi(dx' \times dy')\\
&= \int_W\int_W d_Z(\phi_X^*\omega_X(w,w'), \phi_Y^*\omega_Y(w,w'))^p \pi(dw)\pi(dw')
\end{aligned}$$

Since $d_Z \geqslant 0$, we have $d_Z(\phi_X^*\omega_X(w,w'), \phi_Y^*\omega_Y(w,w')) = 0$ $\pi \otimes \pi$-almost everywhere, or equivalently, $\phi_W = \phi_X^*\omega_X = \phi_Y^*\omega_Y$ $\pi \otimes \pi$-almost everywhere. Therefore, X and Y are weakly isomorphic. ■

4.1.3 Z-Gromov-Wasserstein Space as a Quotient of the L^p Space

One of the classical results in metric measure geometry is that every metric measure space (X, d_X, μ_X) admits a **parameter**, i.e., a surjective Borel measurable map $\rho : [0,1] \to X$ such that $\rho_* \mathscr{L} = \mu_X$; see works by Gromov (2006, Section $3\frac{1}{2}.3$) or Shioya (2016, Lemma 4.2).

This result was applied by Chowdhury and Mémoli (2019) and Sturm (2023) to obtain convenient representations of ($\mathbb{R}$-valued) measure networks. Here, we consider a generalization of this construction to the setting of Z-networks. The proof of the following proposition is obvious from its formulation.

Proposition 32 (Parametrization)**.** *Any (Z,p)-network $X = (X, \omega_X, \mu_X)$ is weakly isomorphic to a (Z,p)-network $([0,1], \rho^*\omega_X, \mathscr{L})$ where $\mathscr{L}$ is the Lebesgue measure on $[0,1]$, $\rho : [0,1] \to X$ is a parameter of X, and $\rho^*\omega_X : [0,1]^2 \to Z$ is defined by $\rho^*\omega_X(s,t) = \omega_X(\rho(s), \rho(t))$.*

An interesting observation is that, given a Z-network $([0,1], \omega, \mathscr{L})$ already defined over $[0,1]$, by considering parameters $\rho : [0,1] \to [0,1]$, we can construct a family of Z-networks $([0,1], \rho^*\omega, \mathscr{L})$, all of which are weakly isomorphic to the original Z-network. Following this idea, Sturm (2023) established an instructive representation of the standard GW distance; here, we will generalize their construction to the Z-GW distance.

Definition 33 (Invariant Transforms, Sturm 2023)**.** *We say that a Borel measurable map $\phi : [0,1] \to [0,1]$ is **$\mathscr{L}$-invariant** if $\phi_*\mathscr{L} = \mathscr{L}$. The $\mathscr{L}$-invariant maps form a semigroup via function composition, which we denote by $\mathrm{Inv}([0,1], \mathscr{L})$.*

For $p \in [1, \infty]$, $\mathrm{Inv}([0,1], \mathscr{L})$ acts on the space of kernels $L^p([0,1]^2, \mathscr{L}^2; Z)$ via the pullback $\omega \mapsto \phi^*\omega$, and the action defines the following equivalence relation on the L^p space , which closely relates to weak isomorphism:

$$\omega \simeq \omega' \Leftrightarrow \exists \phi, \psi \in \mathrm{Inv}([0,1], \mathscr{L}) \text{ s.t. } \phi^*\omega = \psi^*\omega'.$$

Denote the equivalence class of $\omega \in L^p([0,1]^2, \mathscr{L}^2; Z)$ by $[\omega]$. The quotient space

$$\mathbb{L}^p := L^p([0,1]^2, \mathscr{L}^2; Z)/\mathrm{Inv}([0,1], \mathscr{L})$$

admits a metric

$$\mathbb{D}_p([\omega], [\omega']) := \frac{1}{2} \inf \left\{ D_p(\phi^*\omega, \psi^*\omega') | \phi, \psi \in \mathrm{Inv}([0,1], \mathscr{L}) \right\},$$

where D_p is the L^p distance (see Definition 6) and this family of metric spaces provides a representation of the Z-GW space:

Theorem 34. *For $p \in [1, \infty]$, $(\mathbb{L}^p, \mathbb{D}_p)$ is isometric to $(\mathcal{M}_{\sim}^{Z,p}, \mathrm{GW}_p^Z)$ by*

$$\Theta_p : \mathbb{L}^p \ni [\omega] \mapsto [([0,1], \omega, \mathscr{L})] \in \mathcal{M}_{\sim}^{Z,p}$$

where $[([0,1], \omega, \mathscr{L})]$ denotes the equivalence class of a (Z,p)-network $([0,1], \omega, \mathscr{L})$ by weak isomorphism.

Proof As the proof by Sturm (2023, Theorem 5.10) does not rely on the properties of the target metric space, trivial modifications allow us to apply the proof to our case. ■

4.2 Metric and Topological Properties

Having shown that the Z-GW distance is a metric, we now establish some of the properties of its induced topology. In particular, under mild assumptions, we will prove that:

- It is *separable*, i.e., it contains a countable, dense subset. This is crucial for approximation results, in that we can frequently pass to a relatively simple subspace.
- It is *complete*, i.e., its Cauchy sequences always converge. Completeness and separability together make the space of Z-networks a *Polish space*; this is generally considered to be the minimum requirement for the theory of measures over a space to be well-behaved (see Section 2.1.1).
- It is *path-connected* and, in fact, *contractible*. The existence of paths joining points in the space is required for the application of statistical methods. For example, it allows interpolations between points, which is a fundamental ingredient for the computation of geometric statistics such as Fréchet means. Contractibility (that the space of Z-networks is homotopy equivalent to a point) tells us that the space is quite simple, from a topological perspective.

We also show that the Z-Gromov-Wasserstein space is *geodesic*, provided that Z is. Recall that a metric space (X, d_X) is **geodesic** if, for any pair of points $x, x' \in X$, there exists a **geodesic path** joining them; this is a path $\gamma : [0, 1] \to X$ such that, for all $s, t \in [0, 1]$,

$$d_X(\gamma(s), \gamma(t)) = |t - s| d_X(x, x')$$

(see, e.g., Burago et al. (2022) for details). In fact, to check that a given curve γ is a geodesic path, it suffices to show that $d_X(\gamma(s), \gamma(t)) \leqslant |t - s| d_X(x, x')$ holds for all $s, t \in [0, 1]$ (Chowdhury and Mémoli, 2018a, Lemma 1.3). In a geodesic space, points are not only always connected by paths, but are connected by paths which behave like 'straight lines' in the space.

4.2.1 Separability

We first show that the Z-GW distance inherits separability from its target space Z. We will use the following lemma.

Lemma 35. *Let $X_0 = (X, \omega_0, \mu), X_1 = (X, \omega_1, \mu)$ be (Z, p)-networks with the same underlying Polish space and measure but different kernels. We have*

$$\mathrm{GW}_p^Z(X_0, X_1) \leqslant \frac{1}{2} D_p(\omega_0, \omega_1)$$

Here, D_p denotes the L^p distance of Definition 6.

Proof Let $\Delta : X \to X \times X, \Delta(x) = (x, x)$ be the mapping to the diagonal. We note that $\Delta_*\mu$ defines a coupling between μ and itself, so substituting $\pi = \Delta_*\mu$ to the distortion functional, we obtain, for $p \in [1, \infty)$ (the $p = \infty$ case is similar),

$$\begin{aligned}\mathrm{GW}_p^Z(X_0, X_1) \leqslant \frac{1}{2}\mathrm{dis}_p^Z(\pi) &= \frac{1}{2}\left(\int_X \int_X d_Z(\omega_0(x, x'), \omega_1(x, x'))^p \mu(dx)\mu(dx')\right)^{1/p} \\ &= \frac{1}{2} D_p(\omega_0, \omega_1).\end{aligned}$$

■

Proposition 36. *$\mathcal{M}_{\sim}^{Z,p}$ is separable.*

Proof Without loss of generality (by Proposition 32), we consider Z-networks of the form $([0,1],\omega,\mathscr{L})$ where $\mathscr{L}$ is the Lebesgue measure. Since $L^p([0,1]\times[0,1],\mathscr{L}\otimes\mathscr{L};Z)$ is separable by Proposition 8, we can take a countable dense subset $\{\omega_i\}_{i=1}^{\infty}$ of $L^p([0,1]\times[0,1],\mathscr{L}\otimes\mathscr{L};Z)$. Now, for any Z-network $X=([0,1],\omega,\mathscr{L})$ and any $\epsilon>0$, there exists ω_i such that $D_p(\omega,\omega_i)<\epsilon$. By constructing a Z-network by $X_i=([0,1],\omega_i,\mathscr{L})$, and by Lemma 35, we see that $\mathrm{GW}_p^Z(X,X_i)<\epsilon$. Therefore, $\{X_i\}_{i=1}^{\infty}$ is dense in $\mathcal{M}_{\sim}^{Z,p}$. ■

Although the proof above is sufficient to prove the separability, the argument remains high-level and does not provide an explicit countable dense subset in $\mathcal{M}_{\sim}^{Z,p}$. To construct such a subset, we introduce the following object.

Definition 37 (n-Point Z-Network)**.** *For each $n\in\mathbb{N}$, an **n-point Z-network** is the equivalence class of a Z-network (X,ω,μ) by weak isomorphism, where $X=\{1,\dots,n\}$, $\omega: X\times X\to Z$ is any function and $\mu=\frac{1}{n}\sum_{i=1}^{n}\delta_i$. Denote the set of n-point (Z,p)-networks by $\mathcal{M}_n^Z$.*

We note that n-point Z-networks are (Z,p)-networks for any $p\in[1,\infty]$ because the kernels are supported on finite sets. That is, we have $\mathcal{M}_n^Z\subset\mathcal{M}_{\sim}^{Z,p}$ for any $p\in[1,\infty]$. Sturm (2023) proved the density of n-point Z-networks in the special case where he only considers symmetric $\mathbb{R}$-valued kernels ω. However, the proof can be generalized to Z-networks.

Proposition 38. *The countable set $\bigcup_n \mathcal{M}_n^Z$ is dense in $\mathcal{M}_{\sim}^{Z,p}$.*

Proof We first take an arbitrary (Z,p)-network. By parameterization (Proposition 32), it is weakly isomorphic to a (Z,p)-network of the form $X=([0,1],\omega,\mathscr{L})$. The kernel ω belongs to $L^p([0,1]^2,\mathscr{L}^2;Z)$, so by Proposition 8, it can be approximated by a piecewise constant function ω_n that is constant on a grid of step size $1/n$ with n small enough. That is, we can partition $[0,1]$ into n pieces $R_i=[i/n,(i+1)/n)$ for $i=0,\dots,n-2$, $R_{n-1}=[(n-1)/n,n]$ and take ω_n to be constant on each $R_i\times R_j$. The Z-network $X_n=([0,1],\omega_n,\mathscr{L})$ is weakly isomorphic to $(\{1,\cdots,n\},\Omega_n,\frac{1}{n}\sum_{i=1}^{n}\delta_i)$ where Ω_n is the $n\times n$, Z-valued matrix with $\Omega_n(i,j)=\omega_n(i/n,j/n)$ because we can construct a measure preserving map $\phi: X\to\{1,\cdots,n\}$ by $\phi(x)=i$ whenever $x\in R_i$, and this map satisfies $\phi^*\Omega_n=\omega_n$ by definition. Finally, the claim follows by applying Lemma 35 to X and X_n. ■

4.2.2 Completeness

The Z-GW space also inherits completeness from its target space Z. Moreover, the opposite direction is also true.

Theorem 39. *$\mathcal{M}_{\sim}^{Z,p}$ is complete if and only if Z is complete.*

Proof Suppose Z is complete. Then, the proof that GW_p^Z is complete follows exactly as in the case $Z=\mathbb{R}$ (Sturm, 2023, Theorem 5.8), since $L^p(X\times X,\mu_X\otimes\mu_X;Z)$ is complete (Proposition 8). This proves one direction of the theorem. Let us now consider the converse.

We will generalize the proof for the Wasserstein space (Pinelis, 2024) to our setting. Suppose that GW_p^Z is complete. Let $\{z_n\}$ be a Cauchy sequence in Z. We will show that $\{z_n\}$ converges to some $z \in Z$. For each n, let $X_n = (\{0\}, \omega_n, \delta_0)$ where $\omega_n : \{0\} \times \{0\} \to Z$ is the function $\omega_n(0,0) = z_n$. Then, $\{X_n\}$ is a Cauchy sequence in $\mathcal{M}_\sim^{Z,p}$ with respect to GW_p^Z because $\mathrm{GW}_p^Z(X_n, X_m) = \frac{1}{2} d_Z(z_n, z_m)$. By completeness, there exists $X = (X, \omega, \mu) \in \mathcal{M}_\sim^{Z,p}$ such that $\mathrm{GW}_p^Z(X_n, X) \to 0$. Since the only coupling with a Dirac measure is the product measure, we can explicitly calculate $\mathrm{GW}_p^Z(X_n, X)$ as

$$2^p \mathrm{GW}_p^Z(X_n, X)^p = \int_X \int_X d_Z(z_n, \omega(x, x'))^p \mu(dx)\mu(dx') = \int_Z d_Z(z_n, z)^p \omega_*(\mu \otimes \mu)(dz)$$

for $p < \infty$. For $p = \infty$, we have

$$\mathrm{GW}_\infty^Z(X_n, X) = \frac{1}{2} \|d(z_n, z)\|_{L^\infty(\omega_*(\mu \otimes \mu))}$$

For brevity, we denote $\omega_*(\mu \otimes \mu)$ by μ_ω. We note that μ_ω is a probability measure on Z since $\mu \otimes \mu$ is a probability measure on $X \times X$. Therefore, if $p < \infty$, Markov's inequality is applicable to show that

$$\epsilon^p (1 - \mu_\omega(B_\epsilon(z_n))) = \epsilon^p \mu_\omega(\{z \in Z : d_Z(z_n, z) \geqslant \epsilon\}) \leqslant \int_Z d_Z(z_n, z)^p \mu_\omega(dz)$$

Moreover, if $p = \infty$, by the monotonicity of L^p norms,

$$\epsilon(1 - \mu_\omega(B_\epsilon(z_n)) = \epsilon \mu_\omega(\{z \in Z : d_Z(z_n, z) \geqslant \epsilon\}) \leqslant \|d_Z(z_n, z)\|_{L^1(\mu_\omega)} \leqslant \|d_Z(z_n, z)\|_{L^\infty(\mu_\omega)},$$

which shows that $\mu_\omega(B_\epsilon(z_n)) \to 1$ as $n \to \infty$ for any $\epsilon > 0$. Here, $B_\epsilon(z_n) = \{z \in Z : d_Z(z_n, z) < \epsilon\}$.

Intuitively, the above means that the mass of μ_ω is concentrated around z_n as $n \to \infty$, or in other words, z_n gets closer and closer to the support point of μ_ω. Indeed, for any $z \in \mathrm{supp}(\mu_\omega)$, we have $\mu_\omega(B_\epsilon(z)) > 0$ for any $\epsilon > 0$ by definition. This means that, if n is large enough, $B_\epsilon(z)$ and $B_\epsilon(z_n)$ should have a nonempty intersection for any $\epsilon > 0$ because $\mu_\omega(B_\epsilon(z)) > 0$ is a constant and $\mu_\omega(B_\epsilon(z_n)) \to 1$, so the sum of two measures will be larger than 1 at some point. Thus, for any $\epsilon > 0$, there exists a natural number N such that if $n \geqslant N$, we can take $v \in \mathrm{supp}(\mu_\omega) \cap B_\epsilon(z_n)$ so that $d_Z(z, z_n) \leqslant d_Z(z, v) + d_Z(v, z_n) < 2\epsilon$. Since $\epsilon > 0$ is arbitrary, this shows that $z_n \to z$ as $n \to \infty$. As any Cauchy sequence converges, we have shown that Z is complete. ■

4.2.3 Path-Connectedness

One of the distinct properties of $\mathcal{M}_\sim^{Z,p}$ is that it is always contractible, regardless of the topology of Z. We prove this below, but first show that $\mathcal{M}_\sim^{Z,p}$ is always path-connected. Intuitively, the result holds because a Z-network is equipped with a measure, which is a continuous object taking values on the real line. This intuition plays a fundamental role in the following proof. An important feature of the proof is that it provides an explicit formula for a continuous path between any pair of Z-networks. This formula will be used later in the proof of contractibility (see Lemma 43).

Proposition 40. *For any space Z, $\mathcal{M}_{\sim}^{Z,p}$ is path-connected for all $p \in [1, \infty)$.*

The $p = \infty$ case is more subtle, and is not completely resolved, as we describe below.

Proof For any Z-networks $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$, define a path of Z-networks by

$$X_t = (X \amalg Y, \omega_X \amalg \omega_Y, (1-t)\mu_X + t\mu_Y) \quad \text{for} \quad t \in [0, 1]. \tag{15}$$

Here, $\omega_X \amalg \omega_Y : X \amalg Y \to Z$ is any function (independent of t) that satisfies $\omega_X \amalg \omega_Y(x, x') = \omega_X(x, x')$ for $x, x' \in X$ and $\omega_X \amalg \omega_Y(y, y') = \omega_Y(y, y')$ for $y, y' \in Y$, and the measure is defined by

$$((1-t)\mu_X + t\mu_Y)(A) = (1-t)\mu_X(A \cap X) + t\mu_Y(A \cap Y)$$

for any measurable set $A \subset X \amalg Y$.

First, we note that X_0 and X_1 are weakly isomorphic to X and Y, respectively. For X_0 and X, this can be seen by pulling back $X \amalg Y$ to X by the inclusion map. In other words, for two Z-networks $X = (X, \omega_X, \mu_X)$ and $X_0 = (X \amalg Y, \omega_X \amalg \omega_Y, \mu_X)$, we consider $W = X$, the identity map $\phi_X : W \to X$ and the inclusion map $\phi_{X_0} : X \to X \amalg Y$. The triple W, ϕ_X, ϕ_{X_0} satisfies the definition of weak isomorphism. The X_1 case is similar.

We will now prove that the path X_t is continuous with respect to the GW distance. To see this, take any coupling π between X, Y and define a coupling $\pi_{s,t}$ between X_s and X_t for $s < t$ by

$$\pi_{s,t} = (1-t)\Delta_*^X \mu_X + (t-s)\pi + s\Delta_*^Y \mu_Y$$

where

$$\Delta^X : X \to (X \amalg Y)^2 \quad \text{and} \quad \Delta^Y : Y \to (X \amalg Y)^2$$

are mappings into the diagonal of X and Y, respectively. Moreover, π here as a measure on $(X \amalg Y)^2$ is defined as $\pi(A \cap (X \times Y))$ for measurable $A \subset (X \amalg Y)^2$. The coupling $\pi_{s,t}$ represents a transport plan where we keep $1-t$ and s units of mass at X and Y, respectively, and send $t - s$ units from X to Y. By definition, we have

$$\begin{aligned} 2^p \mathrm{GW}_p^Z(X_s, X_t)^p \leqslant &(1-t)^2 I_{XX} + (t-s)^2 I_{\pi\pi} + t^2 I_{YY} + 2(1-t)(t-s) I_{X\pi} \\ &+ 2(t-s)s I_{\pi Y} + 2s(1-t) I_{YX}. \end{aligned}$$

Here, $I_{\mu\nu}$, where $\mu, \nu \in \{X, Y, \pi\}$ is the integral

$$\int_{(X \amalg Y)^2} \int_{(X \amalg Y)^2} d_Z(\omega_X \amalg \omega_Y(u, u'), \omega_X \amalg \omega_Y(v, v'))^p \mu(du \times dv)\nu(du' \times dv'),$$

where X and Y are abusively being used as shorthand for $\Delta_*^X \mu_X$ and $\Delta_*^Y \mu_Y$, respectively. We note that I_{XX}, I_{YX} and I_{YY} vanish since the integrand is identically zero by the definition of $\Delta_*^X \mu_X$ and $\Delta_*^Y \mu_Y$. Therefore,

$$\begin{aligned} 2^p \mathrm{GW}_p^Z(X_s, X_t)^p &\leqslant (t-s)^2 I_{\pi\pi} + 2(1-t)(t-s) I_{X\pi} + 2(t-s)s I_{\pi Y} \\ &\leqslant (t-s)(I_{\pi\pi} + 2I_{X\pi} + 2I_{\pi Y}) \end{aligned} \tag{16}$$

Since $I_{\pi\pi}, I_{X\pi}$ and $I_{\pi Y}$ are independent of time, this proves that our path X_t is $1/p$-Hölder continuous. Therefore, $\mathcal{M}_{\sim}^{Z,p}$ is path-connected. ■

Let us now discuss the path-connectivity properties of $\mathcal{M}_\sim^{Z,\infty}$. To do so, it will be useful to introduce an invariant of a Z-network $X = (X, \omega_X, \mu_X)$. For a fixed point $z \in Z$, we define the **p-size of X, relative to** z to be

$$\mathrm{size}_{p,z}(X) := \|d_Z(\omega_X(\cdot,\cdot), z)\|_{L^p(\mu_X \otimes \mu_X)} \tag{17}$$

(see also Definition 49 below). It is a fact that, for any Z-networks X and Y,

$$\mathrm{GW}_p^Z(X,Y) \geqslant \frac{1}{2}|\mathrm{size}_{p,z}(X) - \mathrm{size}_{p,z}(Y)|. \tag{18}$$

Indeed, we prove this below in Theorem 50, in the context of several lower bounds on the Z-GW distance.

We now show that Proposition 40 fails in the $p = \infty$ case. Intuitively, the proof strategy does not apply because it is based on manipulating "weights" in the measures, whereas GW_∞^Z is insensitive to these weights and only optimizes a quantity which depends on supports of couplings.

Example 1. *Let $Z = \{0,1\} \subset \mathbb{R}$, with the restriction of Euclidean distance. We claim that $\mathcal{M}_\sim^{Z,\infty}$ is not path connected. To see this, observe that the size function* (17) *has the property that*

$$\mathrm{size}_{\infty,0}(X) \in \{0,1\}$$

for all $X \in \mathcal{M}_\sim^{Z,\infty}$. By Theorem 50 (or (18)*), the function $\mathrm{size}_{\infty,0} : \mathcal{M}_\sim^{Z,\infty} \to \mathbb{R}$ is Lipschitz continuous. This means that, if X_t is a continuous path in $\mathcal{M}_\sim^{Z,\infty}$, then the composition*

$$t \mapsto \mathrm{size}_{\infty,0}(X_t) : [0,1] \to \{0,1\}$$

must be continuous, and therefore constant. It follows that there is no path joining a (Z,∞)-network of size 0 to one of size 1. For a specific example, there is no continuous path from

$$X = (\{0\}, \omega_X(0,0) = 0, \delta_0) \quad \textit{to} \quad Y = (\{1\}, \omega_Y(1,1) = 1, \delta_1)$$

in $\mathcal{M}_\sim^{Z,\infty}$.

The idea of the justification in Example 1 can be extended to show that path-continuity of Z is a necessary condition for path-continuity of $\mathcal{M}_\sim^{Z,\infty}$ (we omit the details of this extension here). However, we were unable to determine whether this condition is sufficient. On the other hand, if Z is geodesic then so is $\mathcal{M}_\sim^{Z,\infty}$, as we show below in Theorem 45. This observation illustrates the subtlety of the following question:

Question 41. Does path-connectivity of Z imply path-connectivity of $\mathcal{M}_\sim^{Z,\infty}$?

4.2.4 Contractibility

We now take the path-connectivity result (Proposition 40) a step further and show that $\mathcal{M}_\sim^{Z,p}$ is always contractible.

Theorem 42. *For any space Z, $\mathcal{M}_\sim^{Z,p}$ is contractible for all $p \in [1,\infty)$.*

The proof will use the following lemma, which specializes some features of the proof of Proposition 40 to the case where one of the Z-networks is a one-point space. The result uses the size functions introduced in (17). The proof of the lemma is provided in Section A.3.

Lemma 43. *Let $X = (X, \omega_X, \mu_X)$ be an arbitrary Z-network, and let $Y = (\{\star\}, \omega_Y, \delta_\star)$ be a one-point Z-network, where $\omega_Y(\star, \star) = z$ for some fixed $z \in Z$, and $\delta_\star$ denotes the Dirac measure. Let X_t denote the path defined in* (15) *between X and Y (considered up to weak isomorphism), where we specifically define*

$$\omega_X \amalg \omega_Y(u, v) := \begin{cases} \omega_X(u, v) & \text{if } u, v \in X \\ z & \text{otherwise.} \end{cases}$$

Then the Hölder estimate (16) *simplifies to*

$$\mathrm{GW}_p^Z(X_s, X_t) \leqslant \frac{(3|t-s|)^{1/p}}{2} \cdot \mathrm{size}_{p,z}(X). \tag{19}$$

We now proceed with the proof of Theorem 42. The strategy is to define an explicit contraction to a one-point space which follows paths as defined in the proof of Proposition 40. Namely, fix an arbitrary point $z \in Z$ and, for any Z-network $X = (X, \omega_X, \mu_X)$ and $t \in [0, 1]$, let X_t denote the Z-network

$$X_t = (X \amalg \{\star\}, \hat{\omega}_X, (1-t)\mu_X + t\delta_\star), \tag{20}$$

where $\star$ is an abstract point,

$$\hat{\omega}_X(x, x') = \begin{cases} \omega_X(x, x') & x, x' \in X \\ z & \text{otherwise,} \end{cases}$$

and $\delta_\star$ is the Dirac measure. This is then the path from X to a one-point space, as in Proposition 40 and Lemma 43, where we are using the notation $\hat{\omega}_X$ rather than $\omega_X \amalg \omega_Y$ to condense notation. This will be used to construct the contraction.

Proof of Theorem 42. We define a map $\Phi : \mathcal{M}_\sim^{Z,p} \times [0, 1] \to \mathcal{M}_\sim^{Z,p}$ by

$$\Phi([X], t) = [X_t],$$

where we are using $[X]$ to denote the weak isomorphism class of a Z-network X, and X_t is as in (20). Arguments similar to those used in the proof of Proposition 40 show that Φ is well defined, and that, for any Z-network X, $X_0 \sim X$, and X_1 is weakly isomorphic to the Z-network

$$(\{\star\}, \omega_{\{\star\}}, \delta_\star), \quad \text{with} \quad \omega_{\{\star\}}(\star, \star) = z,$$

so that $\Phi(\cdot, 1)$ is a constant map. It remains to show that Φ is continuous. This will be done in two steps: **Step 1.** We show that the component function $\Phi(\cdot, t)$ is Lipschitz continuous for each fixed $t \in [0, 1]$. **Step 2.** We then combine this with various estimates to derive continuity in general.

Step 1 (Lipschitz Continuity for Fixed t). Following the plan described above, fix $t \in [0, 1]$ and consider the restricted map

$$\Phi(\cdot, t) : \mathcal{M}_\sim^{Z,p} \to \mathcal{M}_\sim^{Z,p}.$$

Toward establishing its continuity, let X and Y be Z-networks and choose an optimal coupling $\pi \in \mathcal{C}(\mu_X, \mu_Y)$. Let X_t and Y_t be as in (20). We extend π to a measure on

$$W = (X \amalg \{\star\}) \times (Y \amalg \{\star\}) = (X \times Y) \amalg (X \times \{\star\}) \amalg (\{\star\} \times Y) \amalg \{(\star, \star)\} \tag{21}$$

as $\overline{\pi} = (1-t)\iota_* \pi + t\delta_{(\star,\star)}$, where $\iota : X \times Y \hookrightarrow W$ denotes the inclusion map. Observe that, amongst the terms in the decomposition (21), $\overline{\pi}$ is supported only on $X \times Y$ and $\{(\star, \star)\}$. The Z-GW distance between X_t and Y_t is bounded above by the cost of the coupling $\overline{\pi}$; in the case that $p < \infty$ (with the $p = \infty$ case being similar), this implies

$$\begin{aligned}
2^p \mathrm{GW}_p^Z(X_t, Y_t)^p &\leqslant \iint_{W \times W} d_Z(\hat{\omega}_X(x, x'), \hat{\omega}_Y(y, y'))^p \overline{\pi}(dx \times dy)\overline{\pi}(dx' \times dy') \\
&= \left(\iint_{(X \times Y)^2} + \iint_{\{(\star,\star)\}^2} \right) d_Z(\hat{\omega}_X(x, x'), \hat{\omega}_Y(y, y'))^p \overline{\pi}(dx \times dy)\overline{\pi}(dx' \times dy') \quad (22) \\
&= (1-t)^2 \iint_{(X \times Y)^2} d_Z(\omega_X(x, x'), \omega_Y(y, y'))^p \pi(dx \times dy)\pi(dx' \times dy') \quad (23) \\
&= (1-t)^2 2^p \mathrm{GW}_p^Z(X, Y)^p, \quad (24)
\end{aligned}$$

with the various equalities justified as follows: (22) uses the observation above on the support of $\overline{\pi}$, together with the fact that

$$d_Z(\hat{\omega}_X(x, \star), \hat{\omega}_Y(y, \star)) = d_Z(\hat{\omega}_X(\star, x), \hat{\omega}_Y(\star, y)) = 0, \qquad \forall\ (x, y) \in X \times Y,$$

so that the remaining "cross-term" integrals $\iint_{(X\times Y)\times\{(\star,\star)\}}$ and $\iint_{\{(\star,\star)\}\times(X\times Y)}$ vanish; (23) follows by the definitions of $\hat{\omega}_X$, $\hat{\omega}_Y$ and $\overline{\pi}$; finally, (24) is given by the optimality of π. This shows that $\Phi(\cdot, t)$ is Lipschitz continuous.

Step 2 (Continuity in General). We now show that the map Φ is continuous. Fix a Z-network Y and $t \in [0, 1]$, and let $\epsilon > 0$. For the moment, let us assume that $\mathrm{size}_{p,z}(Y) > 0$ and $t < 1$ (the remaining special case will be addressed later), so that we may assume without loss of generality (for technical reasons) that

$$\epsilon < \frac{3}{2} \mathrm{size}_{p,z}(Y)(1-t)^{2/p}, \quad \text{or} \quad \frac{2\epsilon}{3\,\mathrm{size}_{p,z}(Y)(1-t)^{2/p}} < 1. \tag{25}$$

We will determine $\delta > 0$ such that

$$\max\{\mathrm{GW}_p^Z(X, Y), |t - s|\} < \delta \Rightarrow \mathrm{GW}_p^Z(X_s, Y_t) < \epsilon,$$

for an arbitrary Z-network X and $s \in [0, 1]$. We have

$$\begin{aligned}
\mathrm{GW}_p^Z(X_s, Y_t) &\leqslant \mathrm{GW}_p^Z(X_s, X_t) + \mathrm{GW}_p^Z(X_t, Y_t) \\
&\leqslant \frac{(3|t-s|)^{1/p}}{2} \mathrm{size}_{p,z}(X) + (1-t)^{2/p} \mathrm{GW}_p^Z(X, Y) \quad (26) \\
&\leqslant \frac{(3|t-s|)^{1/p}}{2} \mathrm{size}_{p,z}(Y) + (3|t-s|)^{1/p} \mathrm{GW}_p^Z(X, Y) + (1-t)^{2/p} \mathrm{GW}_p^Z(X, Y) \quad (27)
\end{aligned}$$

where (26) follows by Lemma 43 and the Lipschitz bound (24), and (27) follows by the lower bound (18) on Z-GW distance in terms of size functions. Choose $\delta > 0$ such that

$$\delta < \min\left\{\frac{1}{3}\left(\frac{2\epsilon}{3\mathrm{size}_{p,z}(Y)}\right)^p, \frac{\epsilon}{3(1-t)^{2/p}}\right\}.$$

The assumption $\max\{\mathrm{GW}_p^Z(X,Y), |t-s|\} < \delta$ then implies

$$\begin{aligned}
&\frac{(3|t-s|)^{1/p}}{2}\mathrm{size}_{p,z}(Y) + (3|t-s|)^{1/p}\mathrm{GW}_p^Z(X,Y) + (1-t)^{2/p}\mathrm{GW}_p^Z(X,Y) \\
&< \frac{2\epsilon}{3\mathrm{size}_{p,z}(Y)} \cdot \frac{1}{2}\mathrm{size}_{p,z}(Y) + \frac{2\epsilon}{3\mathrm{size}_{p,z}(Y)} \cdot \frac{\epsilon}{3(1-t)^{2/p}} + (1-t)^{2/p} \cdot \frac{\epsilon}{3(1-t)^{2/p}} \\
&= \frac{\epsilon}{3} + \frac{2\epsilon}{3\mathrm{size}_{p,z}(Y)(1-t)^{2/p}} \cdot \frac{\epsilon}{3} + \frac{\epsilon}{3} < \epsilon,
\end{aligned} \tag{28}$$

where we have used the assumption (25) to simplify the middle term in (28). This proves continuity of Φ at any Z-network Y with $\mathrm{size}_{p,z}(Y) > 0$ and $t < 1$.

It remains to consider the cases $\mathrm{size}_{p,z}(Y) = 0$ or $t = 1$. In either case, Y_t is weakly isomorphic to X_1, and continuity amounts to controlling $\mathrm{GW}_p^Z(X_s, X_1)$. Applying Lemma 43 and the size bound (18) gives

$$\begin{aligned}
\mathrm{GW}_p^Z(X_s, Y_t) = \mathrm{GW}_p^Z(X_s, X_1) &\leqslant \frac{(3(1-s))^{1/p}}{2}\mathrm{size}_{p,z}(X) \\
&\leqslant \frac{(3(1-s))^{1/p}}{2}(\mathrm{size}_{p,z}(Y) + 2\mathrm{GW}_p^Z(X,Y)).
\end{aligned} \tag{29}$$

If $t = 1$, then (29) can be made arbitrarily small by taking s sufficiently close to 1 and X sufficiently close to Y. If $\mathrm{size}_{p,z}(Y) = 0$, then we get

$$\frac{(3(1-s))^{1/p}}{2}(\mathrm{size}_{p,z}(Y) + 2\mathrm{GW}_p^Z(X,Y)) \leqslant 3^{1/p}\mathrm{GW}_p^Z(X,Y),$$

which gives control of $\mathrm{GW}_p^Z(X_s, Y_t)$ in terms of $\mathrm{GW}_p^Z(X,Y)$, and this completes the proof. ■

In analogy with the discussion in the previous subsection, the proof idea for this result fails in the $p = \infty$ case, which remains open.

Question 44. How does contractibility of $\mathcal{M}_{\sim}^{Z,\infty}$ depend on that of Z?

4.2.5 Geodesicity

A natural question is whether the path-connectedness result Proposition 40 can be pushed further to a statement about the existence of geodesics. Although this question is not fully solved, we have the following result passing the geodesicity of Z to $\mathcal{M}_{\sim}^{Z,p}$.

Theorem 45. *If Z is geodesic, then so is $\mathcal{M}_{\sim}^{Z,p}$.*

Proof Suppose that (Z, d_Z) is geodesic. Let $X_i = (X_i, \omega_i, \mu_i) \in \mathcal{M}^{Z,p}$ for $i \in \{0,1\}$ and let π be an optimal coupling realizing $\mathrm{GW}_p^Z(X_0, X_1)$. For $t \in [0,1]$, let $X_t \in \mathcal{M}^{Z,p}$ be the Z-network

$$X_t = (X_0 \times X_1, \omega_t, \pi),$$

with $\omega_t : (X_0 \times X_1) \times (X_0 \times X_1) \to Z$ defined as follows. For $(x_0, x_1), (x_0', x_1') \in X_0 \times X_1$, choose

$$\omega_\bullet((x_0, x_1), (x_0', x_1')) : [0,1] \to Z : t \mapsto \omega_t((x_0, x_1), (x_0', x_1'))$$

to be a geodesic joining $\omega_0(x_0, x_0')$ to $\omega_1(x_1, x_1')$ in Z; in particular, $\omega_i((x_0, x_1), (x_0', x_1')) = \omega_i(x_i, x_i')$ for $i \in \{0,1\}$. It is enough to prove that $\mathrm{GW}_p^Z(X_s, X_t) \leqslant |s-t| \mathrm{GW}_p^Z(X_0, X_1)$ for any $s, t \in [0,1]$.

Let $X = X_0 \times X_1$ and construct a coupling of π and π on $X \times X$ by $\Delta_* \pi$ where $\Delta : X \to X \times X$ is the standard diagonal map, $\Delta(x) = (x, x)$. Then, for any $s, t \in [0,1]$, and for $p \in [1, \infty)$,

$$\begin{aligned}
&2^p \mathrm{GW}_p^Z(X_s, X_t)^p \\
&\leqslant \int_{X \times X} \int_{X \times X} d_Z(\omega_s(x, x'), \omega_t(y, y'))^p \Delta_* \pi(dx \times dy) \Delta_* \pi(dx' \times dy') \\
&= \int_X \int_X d_Z(\omega_s(x, x'), \omega_t(x, x'))^p \pi(dx) \pi(dx') && (30) \\
&= |s-t|^p \int_X \int_X d_Z(\omega_0(x, x'), \omega_1(x, x'))^p \pi(dx) \pi(dx') && (31) \\
&= |s-t|^p \int_{X_0 \times X_1} \int_{X_0 \times X_1} d_Z(\omega_0(x_0, x_0'), \omega_1(x_1, x_1'))^p \pi(dx_0 \times dx_1) \pi(dx_0' \times dx_1') && (32) \\
&= |s-t|^p 2^p \mathrm{GW}_p^Z(X_0, X_1)^p. && (33)
\end{aligned}$$

The argument uses the change of variable formula in (30), that ω_t is a geodesic in (31), the definition of ω_0 and ω_1 in (32) and that π is an optimal coupling in (33). The $p = \infty$ case follows from the observation that the proof can be rewritten in terms of L^p norms, in which case it applies directly. ∎

It is currently an open question if we have the geodesic property of $\mathcal{M}_\sim^{Z,p}$ when Z is not geodesic. Although not a counterexample, the following example partially explains how the Gromov-Wasserstein distance for a discrete Z behaves.

Example 2. *Fix $p = 1$. Let $Z = \{0,1\}$ with a discrete metric and consider Z-networks $X_0 = ([0,1], \omega_0, \mathscr{L})$ and $X_1 = ([0,1], \omega_1, \mathscr{L})$ where ω_0 and ω_1 are constant functions returning the values 0 and 1, respectively. The path $X_t = ([0,1], \omega_t, \mathscr{L})$ where*

$$\omega_t(u, v) = \begin{cases} 0, & v \leqslant t \\ 1, & \textit{otherwise} \end{cases}$$

defines a geodesic between X_0 and X_1 because, for $0 \leqslant s \leqslant t \leqslant 1$, by Lemma 35, we have $\mathrm{GW}_1^Z(X_s, X_t) \leqslant \frac{1}{2} D_1(\omega_s, \omega_t)$, and

$$D_1(\omega_s, \omega_t) = \int_0^1 \int_0^1 d_Z(\omega_0(u,v), \omega_1(u,v)) du dv = \int_s^t \int_0^1 du dv = t - s$$

Now, since ω_0, ω_1 are constant, X_0 and X_1 are weakly isomorphic to $(\{0\}, \bar{\omega}_0, \delta_0)$ and $(\{1\}, \bar{\omega}_1, \delta_1)$, respectively, where $\bar{\omega}_0, \bar{\omega}_1$ here are again the constant function with values $0, 1$, respectively. As the product measure is the only coupling between Diracs, we can explicitly calculate $\mathrm{GW}_1^Z(X_0, X_1) = \frac{1}{2}$. Combining together with the previous result, we have $\mathrm{GW}_1^Z(X_s, X_t) \leqslant (t-s)\mathrm{GW}_1^Z(X_0, X_1)$, which is sufficient to show that X_t is a geodesic.

As we can see from this example, although $Z = \{0, 1\}$ is discrete, we can find a geodesic connecting between X_0 and X_1. Moreover, it is crucial that $p = 1$ because of the following.

Proposition 46. *Let $Z = \{0, 1\}$ endowed with the discrete metric. Then, $\mathcal{M}_\sim^{Z,p}$ is not geodesic for $p > 1$.*

Proof Consider Z-networks $X = (\{0\}, \omega_0, \delta_0), Y = (\{0\}, \omega_1, \delta_0)$ where ω_0, ω_1 are the functions with $\omega_0(0, 0) = 0$ and $\omega_1(0, 0) = 1$. Suppose that $\mathcal{M}_\sim^{Z,p}$ is geodesic. Then, there is a *midpoint* between them (Burago et al., 2001, Lemma 2.4.8). That is, there exists a Z-network $M = (M, \omega_M, \mu_M)$ such that

$$\mathrm{GW}_p^Z(X, M) = \mathrm{GW}_p^Z(Y, M) = \frac{1}{2}\mathrm{GW}_p^Z(X, Y).$$

We will now calculate each term in the equation. Since the only coupling between δ_0 and μ_M is the product measure $\delta_0 \otimes \mu_M$, it is automatically the optimal coupling, and we can calculate $\mathrm{GW}_p^Z(X, M)$ as

$$\mathrm{GW}_p^Z(X, M) = \frac{1}{2}\left(\int_M \int_M d_Z(0, \omega_M(m, m'))^p \mu_M(dm)\mu_M(dm')\right)^{1/p}$$

Since $d_Z(0, \omega_M(m, m'))$ is nonzero if and only if $\omega_M(m, m') = 1$, we have

$$\mathrm{GW}_p^Z(X, M) = \frac{1}{2}\mu_M \otimes \mu_M(\omega_M^{-1}(\{1\}))^{1/p}$$

Similarly, we have $\mathrm{GW}_p^Z(Y, M) = \frac{1}{2}\mu_M \otimes \mu_M(\omega_M^{-1}(\{0\}))^{1/p}$ and $\mathrm{GW}_p^Z(X, Y) = \frac{1}{2}$. Therefore, we obtain

$$\frac{1}{2}\mu_M \otimes \mu_M(\omega_M^{-1}(\{1\}))^{1/p} = \frac{1}{2}\mu_M \otimes \mu_M(\omega_M^{-1}(\{0\}))^{1/p} = \frac{1}{4}$$

This equation implies $1 = \mu_M \otimes \mu_M(M \times M) = \mu_M \otimes \mu_M(\omega_M^{-1}(\{0\})) + \mu_M \otimes \mu_M(\omega_M^{-1}(\{1\})) = \frac{1}{2^{p-1}}$, but this is contradiction since $p > 1$. ■

These two results suggest that the geodesicity of GW_p^Z differs between $p = 1$ and $p > 1$ cases. In fact, the Wasserstein distance has such a property: the Wasserstein p-space on any Polish metric space Z is geodesic if $p = 1$, but geodesicity depends on the underlying metric if $p > 1$ (Mémoli and Wan, 2023, Theorem 3.16, Remark 3.18). Considering this result, we conclude this subsection by posing the following questions:

Question 47. For $p = 1$, is $\mathcal{M}_\sim^{Z,p}$ geodesic for any Polish metric space Z?

Question 48. For $p > 1$, does $\mathcal{M}_\sim^{Z,p}$ being geodesic imply that Z is geodesic?

5. Lower Bounds and Computational Aspects

As discussed by Mémoli (2011a) and Chowdhury and Mémoli (2019), the exact computation of the Gromov-Wasserstein distance is equivalent to solving non-convex quadratic programming, which is NP-hard. However, we can still obtain computationally tractable lower bounds on the distance through **invariants** of Z-networks. An invariant of a Z-network is defined in terms of some map $\phi : \mathcal{M}_{\sim}^{Z,p} \to (I, d_I)$, where (I, d_I) is a relatively simple pseudometric space, with the property that $\mathrm{GW}_p^Z(X,Y) = 0$ implies $\phi(X) = \phi(Y)$. The associated invariant of X is then $\phi(X)$. Example target spaces (I, d_I) include the real line, the Wasserstein space, or $\mathcal{M}_{\sim}^{Z,p}$ with simpler pseudometrics defined by optimal transport problems. Through such invariants, we will be able to transform the problem of estimating the GW distance into the problem of calculating the distance in (I, d_I).

In this section, we will generalize previous computational results on invariants of the standard Gromov-Wasserstein distance to our framework. We also provide a new result about approximating the Z-GW distance, for arbitrary Z, by the $\mathbb{R}^n$-GW distance.

5.1 Hierarchy of Lower Bounds

Following the previous works (Chowdhury and Mémoli 2019, Section 3; Mémoli 2007, Section 6; Mémoli 2011a, Section 6), we consider the lower bounds and invariants of the Z-network GW distance. We first define the generalizations of invariants given by Chowdhury and Mémoli (2019, Section 3).

Definition 49 (Z-Network Invariants)**.** *Let $X = (X, \omega_X, \mu_X)$ and $Y = (Y, \omega_Y, \mu_Y)$ be (Z,p)-networks. We define the following invariants:*

1. *The **size** of X given a base point $z_0 \in Z$ is*

$$\mathrm{size}_{p,z_0}(X) = \|d_Z(\omega_X(\cdot,\cdot), z_0)\|_{L^p(\mu_X \otimes \mu_X)}$$

2. *The **outgoing joint eccentricity function** $\mathrm{ecc}^{\mathrm{out}}_{p,X,Y} : X \times Y \to \mathbb{R}_+$ is defined by*

$$\mathrm{ecc}^{\mathrm{out}}_{p,X,Y}(x,y) = \inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \|d_Z(\omega_X(x,\cdot), \omega_Y(y,\cdot))\|_{L^p(\pi)}$$

3. *The **incoming joint eccentricity function** $\mathrm{ecc}^{\mathrm{in}}_{p,X,Y} : X \times Y \to \mathbb{R}_+$ is defined by*

$$\mathrm{ecc}^{\mathrm{in}}_{p,X,Y}(x,y) = \inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \|d_Z(\omega_X(\cdot,x), \omega_Y(\cdot,y))\|_{L^p(\pi)}$$

4. *The **outgoing eccentricity** of X given a base point z_0 is*

$$\mathrm{ecc}^{\mathrm{out}}_{p,z_0,X}(x) = \|d_Z(\omega_X(x,\cdot), z_0)\|_{L^p(\mu_X)}$$

5. *The **incoming eccentricity** of X given a base point z_0 is*

$$\mathrm{ecc}^{\mathrm{in}}_{p,z_0,X}(x) = \|d_Z(\omega_X(\cdot,x), z_0)\|_{L^p(\mu_X)}$$

We note that many of the invariants are defined for a base point $z_0 \in Z$. For the case $Z = \mathbb{R}$ considered by Chowdhury and Mémoli (2019), the base point was implicitly chosen to be 0. However, in the general case, there is no canonical choice of the base point. One of the major results by Chowdhury and Mémoli (2019) was the hierarchy of lower bounds for the GW distance for $\mathbb{R}$-networks. We generalize this result to the Z-network case along with the definitions of the lower bounds.

Theorem 50 (Hierarchy of lower bounds)**.** *Let $X = (X, \omega_X, \mu_X), Y = (Y, \omega_Y, \mu_Y)$ be Z-networks. We fix $z_0 \in Z$ and define a cost function $C : X \times Y \to \mathbb{R}_+$ by $C(x, y) = W_p(\omega_X(x, \cdot)_*\mu_X, \omega_Y(y, \cdot)_*\mu_Y)$. Then we have the following, for $p \in [1, \infty]$ and $z_0 \in Z$:*

$$
\begin{aligned}
\mathrm{GW}_p^Z(X,Y) &\geqslant \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \|\mathrm{ecc}^{\mathrm{out}}_{p,X,Y}\|_{L^p(\pi)} && \text{(TLB)}\\
&= \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \|C\|_{L^p(\pi)} && (Z\text{-TLB})\\
&\geqslant \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \|\mathrm{ecc}^{\mathrm{out}}_{p,z_0,X} - \mathrm{ecc}^{\mathrm{out}}_{p,z_0,Y}\|_{L^p(\pi)} && \text{(FLB)}\\
&= \frac{1}{2} \mathrm{W}_p((\mathrm{ecc}^{\mathrm{out}}_{p,z_0,X})_*\mu_X, (\mathrm{ecc}^{\mathrm{out}}_{p,z_0,Y})_*\mu_Y) && (Z\text{-FLB})\\
&\geqslant \frac{1}{2} |\mathrm{size}_{p,z_0}(X) - \mathrm{size}_{p,z_0}(Y)| && \text{(SzLB)}\\
\mathrm{GW}_p^Z(X,Y) &\geqslant \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)} \|d_Z(\omega_X,\omega_Y)\|_{L^p(\pi)} && \text{(SLB)}\\
&= \frac{1}{2} \mathrm{W}_p((\omega_X)_*(\mu_X\otimes\mu_X), (\omega_Y)_*(\mu_Y\otimes\mu_Y)) && (Z\text{-SLB})
\end{aligned}
$$

and similar inequalities for the incoming eccentricity functions.

We note that the terminology FLB, SLB and TLB stands for "first", "second" and "third" lower bound, respectively—this is in reference to the prior conventions (Mémoli, 2011a; Chowdhury and Mémoli, 2019). To prove this theorem, we follow the same strategy as Chowdhury and Mémoli (2019, Theorem 24). Therefore we first introduce the following lemma, whose proof is provided in Section A.4.

Lemma 51 (Chowdhury and Mémoli 2019, Lemma 28)**.** *Let X, Y, Z be Polish, and let $f : X \to Z$ and $g : Y \to Z$ be measurable. Let $T : X \times Y \to Z \times Z$ be the map $(x, y) \mapsto (f(x), g(y))$. Then we have:*

$$T_*\mathcal{C}(\mu_X, \mu_Y) = \mathcal{C}(f_*\mu_X, g_*\mu_Y)$$

Consequently,

$$\mathrm{W}_p(f_*\mu_X, g_*\mu_Y) = \inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \|d_Z(f, g)\|_{L^p(\pi)}$$

Proof of Theorem 50. The inequality (TLB) is obtained through

$$\begin{aligned}
\mathrm{GW}_p^Z(X,Y) &\geqslant \frac{1}{2}\inf_{\pi,\nu\in\mathcal{C}(\mu_X,\mu_Y)} \|d_Z(\omega_X,\omega_Y)\|_{L^p(\pi\otimes\nu)} \\
&= \frac{1}{2}\inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)}\inf_{\nu\in\mathcal{C}(\mu_X,\mu_Y)} \|\|d_Z(\omega_X(x,\cdot),\omega_Y(y,\cdot))\|_{L^p(\nu)}\|_{L^p(\pi)} \\
&\geqslant \frac{1}{2}\inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)} \left\|\inf_{\nu\in\mathcal{C}(\mu_X,\mu_Y)} \|d_Z(\omega_X(x,\cdot),\omega_Y(y,\cdot))\|_{L^p(\nu)}\right\|_{L^p(\pi)} \\
&= \frac{1}{2}\inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \|\mathrm{ecc}^{\mathrm{out}}_{p,X,Y}\|_{L^p(\pi)}.
\end{aligned}$$

The equality (TLB) = (Z-TLB) is obtained by applying Lemma 51, by setting $f = \omega_X(x,\cdot), g = \omega_Y(y,\cdot)$. The inequality (TLB) $\geqslant$ (FLB) is obtained by applying the reverse triangle inequality

$$d_Z(\omega_X(x,\cdot),\omega_Y(x,\cdot)) \geqslant |d_Z(\omega_X(x,\cdot),z_0) - d_Z(\omega_Y(y,\cdot),z_0)|$$

and the reverse Minkowski inequality

$$\|d_Z(\omega_X(x,\cdot),z_0) - d_Z(\omega_Y(y,\cdot),z_0)\|_{L^p(\pi)} \geqslant \left|\|d_Z(\omega_X(x,\cdot),z_0)\|_{L^p(\pi)} - \|d_Z(\omega_Y(y,\cdot),z_0)\|_{L^p(\pi)}\right|.$$

The equality (FLB) = (Z-FLB) is again obtained via Lemma 51 by setting $f = \mathrm{ecc}^{\mathrm{out}}_{p,z_0,X}, g = \mathrm{ecc}^{\mathrm{out}}_{p,z_0,Y}$. The inequality (FLB) $\geqslant$ (SzLB) is obtained by another application of the reverse Minkowski inequality,

$$\|\mathrm{ecc}^{\mathrm{out}}_{p,z_0,X} - \mathrm{ecc}^{\mathrm{out}}_{p,z_0,Y}\|_{L^p(\pi)} \geqslant \left|\|\mathrm{ecc}^{\mathrm{out}}_{p,z_0,X}\|_{L^p(\pi)} - \|\mathrm{ecc}^{\mathrm{out}}_{p,z_0,Y}\|_{L^p(\pi)}\right|,$$

and the fact that

$$\|\mathrm{ecc}^{\mathrm{out}}_{p,z_0,X}\|_{L^p(\pi)} = \|\mathrm{ecc}^{\mathrm{out}}_{p,z_0,X}\|_{L^p(\mu)} = \mathrm{size}_{p,z_0}(X).$$

Finally, the inequality (SLB) is obtained by considering an optimal coupling $\pi \in \mathcal{C}(\mu_X,\mu_Y)$ and applying the following chain of inequalities for $\pi\otimes\pi \in \mathcal{C}(\mu_X\otimes\mu_X, \mu_Y\otimes\mu_Y)$:

$$\mathrm{GW}_p^Z(X,Y) = \frac{1}{2}\|d_Z(\omega_X,\omega_Y)\|_{L^p(\pi\otimes\pi)} \geqslant \frac{1}{2}\inf_{\sigma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)} \|d_Z(\omega_X,\omega_Y)\|_{L^p(\sigma)}$$

and the equality (SLB) = (Z-SLB) is obtained via Lemma 51 by setting $f = \omega_X, g = \omega_Y$. ■

5.2 Approximation of Z-Network GW Distances via $\mathbb{R}^n$-Networks

In the previous subsection, we note that (FLB) was obtained by the reverse triangle inequality $|d_Z(\omega_X,z_0) - d_Z(\omega_Y,z_0)| \leqslant d_Z(\omega_X,\omega_Y)$. This inequality can be seen as relating the Z-networks (X,ω_X,μ_X) and (Y,ω_Y,μ_Y) to the $\mathbb{R}$-networks $(X,d(\omega_X,z_0),\mu_X)$ and $(Y,d(\omega_X,z_0),\mu_Y)$, respectively. We will now generalize this idea to $\mathbb{R}^n$-networks.

Theorem 52. *Suppose Z is a bounded metric space. Let $X = (X, \omega_X, \mu_X)$, $Y = (Y, \omega_Y, \mu_Y)$ be Z-networks. We fix $Q = (q_1, \dots, q_n) \in Z^n$ and define $\mathbb{R}^n$-networks by*

$$X_Q = (X, \omega_{X_Q}, \mu_X) \quad \text{and} \quad Y_Q = (Y, \omega_{Y_Q}, \mu_Y),$$

where

$$\omega_{X_Q} = (d_Z(\omega_X, q_i))_{i=1}^n \quad \text{and} \quad \omega_{Y_Q} = (d_Z(\omega_Y, q_i))_{i=1}^n.$$

Then we obtain

$$n^{-1/r}\mathrm{GW}_p^{\mathbb{R}^n}(X_Q, Y_Q) \leqslant \mathrm{GW}_p^Z(X, Y) \leqslant \mathrm{GW}_p^{\mathbb{R}^n}(X_Q, Y_Q) + \mathrm{H}(Z, Q),$$

where $\mathrm{H}(Z, Q)$ is the Hausdorff distance between Z and Q, the distance on $\mathbb{R}^n$ is the ℓ^r distance with $r \in [1, \infty]$, and we set $n^{-1/\infty} = 1$.

Remark 53. *Before proving the theorem, we provide a few remarks on the statement and its interpretation.*

Generalization to norms other than ℓ^r norms. *Since all norms on a finite-dimensional vector space are equivalent, Theorem 52 can be generalized to the case where we equip $\mathbb{R}^n$ with a norm $\|\cdot\|$ other than the ℓ^r norm. The resulting inequality is as follows:*

$$A \cdot \mathrm{GW}_p^{\mathbb{R}^n}(X_Q, Y_Q) \leqslant \mathrm{GW}_p^Z(X, Y) \leqslant B \cdot \mathrm{GW}_p^{\mathbb{R}^n}(X_Q, Y_Q) + \mathrm{H}(Z, Q),$$

Here, A and B are constants such that $A\|\cdot\| \leqslant \|\cdot\|_{\ell^\infty} \leqslant B\|\cdot\|$.

Necessity of compactness for practical approximation. *Although Theorem 52 is true assuming only the boundedness of Z, it is necessary to impose a stronger condition, such as compactness, to make the Hausdorff distance between Z and Q arbitrarily small as $n \to \infty$.*

The relation between the theorem and the Fréchet-Kuratowski theorem. *In Theorem 52, setting $r = \infty$ gives no constant factor to the GW distance, and thus the difference between the $\mathbb{R}^n$-GW distance and the Z-GW distance admits the following simple Hausdorff distance bound:*

$$\left|\mathrm{GW}_p^{\mathbb{R}^n}(X_Q, Y_Q) - \mathrm{GW}_p^Z(X, Y)\right| \leqslant \mathrm{H}(Z, Q). \tag{34}$$

The inequality shows that the approximation of $\mathrm{GW}_p^Z(X, Y)$ by $\mathrm{GW}_p^Z(X_Q, Y_Q)$ is as accurate as the approximation of Z by Q. In other words, it establishes consistency *and* stability *of the approximation. This result is observed specifically in the case $r = \infty$, which can be understood through the Fréchet-Kuratowski theorem (Ostrovskii, 2013, Proposition 1.17). This theorem states that any separable metric space isometrically embeds into $\ell^\infty(\mathbb{R})$, the space of bounded sequences on $\mathbb{R}$. The embedding $Z \hookrightarrow \ell^\infty(\mathbb{R})$ for bounded Z is provided by the mapping $z \mapsto \{d_Z(z, z_i)\}_{i=1}^\infty$ where $\{z_i\}_{i=1}^\infty$ is any dense subset of Z. Consequently, $\omega_{X_Q} = \{d_Z(\omega_X, q_i)\}_{i=1}^n$ serves as an approximation of ω_X, identified as $\{d_Z(\omega_X, z_i)\}_{i=1}^\infty$, which parallels the Hausdorff approximation of Z by Q. On the other hand, for $r < \infty$, the mapping $z \mapsto \{d_Z(z, z_i)\}_{i=1}^\infty$ is not necessarily an isometric embedding. As a result, the approximation of ω_X by ω_{X_Q} no longer holds.*

Error bounds for the GW distance approximation. *Continuing in the $r = \infty$ case, by (34), an error bound for the Hausdorff distance is automatically an error bound for* *the approximation of the GW distance. In the compact case, previous works (Carlsson and Mémoli, 2010, Theorem 34), (Chazal et al., 2014, Theorem 2) show bounds on the convergence rate to 0 of the Hausdorff distance* $\mathrm{H}(Z,Q)$ *as* $n \to \infty$ *when* $Q = \{q_1, \cdots, q_n\}$ *is a random sample from* Z *(with respect to a measure satisfying certain technical conditions).*

Proof of Theorem 52. We consider the case $r = \infty$, since other cases follow by using the equivalence of the norm $n^{-1/r}\|x\|_{\ell^r} \leqslant \|x\|_{\ell^\infty} \leqslant \|x\|_{\ell^r}$. We first prove the lower bound $\mathrm{GW}_p^{\mathbb{R}^n}(X_Q, Y_Q) \leqslant \mathrm{GW}_p^Z(X,Y)$. Since $|d_Z(\omega_X, q_i) - d_Z(\omega_Y, q_i)| \leqslant d_Z(\omega_X, \omega_Y)$ for any i, we obtain

$$\|\omega_{X_Q} - \omega_{Y_Q}\|_{\ell^\infty} = \max_i |d_Z(\omega_X, q_i) - d_Z(\omega_Y, q_i)| \leqslant d_Z(\omega_X, \omega_Y).$$

Therefore, we obtain the lower bound by integrating both sides against a coupling $\pi \in \mathcal{C}(\mu_X, \mu_Y)$ and taking the infimum in terms of π. For the upper bound, notice that, for any $z, z' \in Z$, we have

$$d_Z(z,z') \leqslant \max_i |d_Z(z,q_i) - d_Z(z',q_i)| + 2\sup_z \min_i d_Z(z,q_i)$$

To verify this, we assume that $d_Z(z,q_i) \geqslant d_Z(z',q_i)$ without loss of generality because of symmetry. Then, we have

$$d_Z(z,z') \leqslant |d_Z(z,q_i) - d_Z(z',q_i)| + 2d_Z(z',q_i) \tag{35}$$

because the right hand side is equal to $d_Z(z,q_i) - d_Z(z',q_i) + 2d_Z(z',q_i) = d_Z(z,q_i) + d_Z(z',q_i)$. By the triangle inequality, we obtain the inequality. Thus, taking the maximum in terms of i, we have

$$d_Z(z,z') \leqslant \max_i |d_Z(z,q_i) - d_Z(z',q_i)| + 2d_Z(z',q_i)$$

Since the above is true for any i, we obtain

$$d_Z(z,z') \leqslant \max_i |d_Z(z,q_i) - d_Z(z',q_i)| + 2\min_i d_Z(z',q_i).$$

Taking the supremum in terms of z',

$$d_Z(z,z') \leqslant \max_i |d_Z(z,q_i) - d_Z(z',q_i)| + 2\sup_{z'} \min_i d_Z(z',q_i),$$

which is equivalent to (35). We briefly point out that the boundedness assumption on Z is used here to ensure that $\sup_{z'} \min_i d_Z(z',q_i)$ is finite. In addition, the second term $\sup_{z'} \min_i d_Z(z',q_i) = \sup_z d_Z(z,Q)$ can be bounded by the Hausdorff distance $\mathrm{H}(Z,Q) = \max\{\sup_z d_Z(z,Q), \sup_i d_Z(q_i,Z)\}$. Now, substituting $z = \omega_X, z' = \omega_Y$, we obtain

$$d_Z(\omega_X,\omega_Y) \leqslant \max_i |d_Z(\omega_X,q_i) - d_Z(\omega_Y,q_i)| + 2\mathrm{H}(Z,Q).$$

For $\pi \in \mathcal{C}(\mu_X, \mu_Y)$, this yields

$$\begin{aligned} \|d_Z(\omega_X,\omega_Y)\|_{L^p(\pi\otimes\pi)} &\leqslant \left\| \max_i |d_Z(\omega_X,q_i) - d_Z(\omega_Y,q_i)| + 2\mathrm{H}(Z,Q) \right\|_{L^p(\pi\otimes\pi)} \\ &\leqslant \left\| \max_i |d_Z(\omega_X,q_i) - d_Z(\omega_Y,q_i)| \right\|_{L^p(\pi\otimes\pi)} + 2\mathrm{H}(Z,Q), \end{aligned}$$

where the last inequality follows from Minkowski's inequality and the fact that $2\mathrm{H}(Z,Q)$ is constant. Dividing both sides by 2 and taking the infimum in terms of π, we obtain the proposition. ■

5.3 Numerical Algorithm

Although the focus of the paper is on theoretical aspects of the Z-GW distance, let us briefly detour to sketch a numerical scheme which adapts existing algorithms.

We start by introducing some notation. Namely, for a 4-dimensional array $(L_{ijkl})_{ijkl}$ and a matrix $(T_{ij})_{ij}$, we define the product $L \otimes T$ via

$$L \otimes T = \left(\sum_{kl} L_{ijkl} T_{kl} \right)_{ij}.$$

Next, we consider two finite Z-networks $(\{1, \cdots, m\}, \omega_X, \mu_X)$ and $(\{1, \cdots, n\}, \omega_Y, \mu_Y)$. In other words, ω_X, ω_Y are Z-valued $m \times m$ and $n \times n$ matrices, respectively, and μ_X, μ_Y are m-dimensional and n-dimensional probability vectors, respectively. To numerically estimate an optimal coupling T—that is, an $\mathbb{R}$-valued $m \times n$ matrix—we perform the following iterations:

$$T^{(l+1)} = \min_{T \in \mathcal{C}(\mu_X, \mu_Y)} \langle d_Z(\omega_X, \omega_Y)^p \otimes T^{(l)}, T \rangle - \epsilon H(T),$$

$$\text{where } d_Z(\omega_X, \omega_Y)_{ijkl} = d_Z\big(\omega_X(i,k), \omega_Y(j,l)\big), \quad H(T) = -\sum_{i,j} T_{ij}\big(\log T_{ij} - 1\big), \quad \epsilon > 0.$$

Here, for matrices A, B of the same size, $\langle A, B \rangle = \sum_{ij} A_{ij} B_{ij}$. The main difference between applying this algorithm to estimate the Z-GW distance as compared to the classical GW distance is the calculation of the product $d_Z(\omega_X, \omega_Y)^p \otimes T^{(l)}$. If $Z = \mathbb{R}$, $d_Z(x, y) = |x - y|$, and $p = 2$, the numerical routine of Peyré et al. (2016) reduces the time complexity of the product calculation to $O(n^2 m + m^2 n)$. In general, the naive calculation of the product costs $O(n^2 m^2)$ in time. We also need $O(n^2 m^2)$ space to store the 4-dimensional array $d_Z(\omega_X, \omega_Y)$. To alleviate this issue, we can use the approximation by $\mathbb{R}^n$-networks proven in Theorem 52. In particular, setting $p = r = 2$ recovers the situation in Peyré et al. (2016) and allows an $O(N(n^2 m + m^2 n))$ time algorithm where N is the size of Q in Theorem 52. This is still a cubic-time algorithm, but we believe that a linear-time algorithm is possible by using the Sampled GW algorithm, cf. Kerdoncuff et al. (2021). We defer addressing these issues and developing a comprehensive numerical framework to future work.

6. Discussion

The Z-Gromov-Wasserstein framework defines a very general setting for reasoning about the mathematical properties of the rich and varied GW-like distances which have appeared in the recent literature. These properties have frequently been derived independently in each instance; in contrast, the work here provides a high-level unified perspective, showing that the metric properties of a Z-GW distance are closely related to those of the space

Z. This initial paper on the topic is meant to lay out its foundations, and we envision several directions for future research, both theoretical and applications-oriented in nature. We conclude the paper by outlining some important future goals:

Geodesic structure and curvature bounds. The geodesic structure of the Z-GW space $\mathcal{M}_{\sim}^{Z,p}$, and how it relates to that of Z, has so far only been partially characterized in Theorem 45. Answers to the related open questions, Question 47 and Question 48, would provide a more complete characterization. Moreover, it would be interesting to give conditions under which the geodesics of Z-GW distance are always of the form described explicitly in the proof of Theorem 45. For example, a result of this form is demonstrated by Sturm for the standard GW distance by Sturm (2023), and this is generalized to certain GW variants by Vayer et al. (2020a); Chowdhury et al. (2023); Zhang et al. (2024a,b). Such a characterization would be a first step in establishing bounds on the Alexandrov curvature of the Z-GW space. Understanding the relationship between the curvature of the space Z and that of the Z-GW space would be very interesting from a theoretical perspective, and could have practical implications to algorithm design. Indeed, a main motivation for the curvature bounds established by Sturm (2023) in the standard GW setting was to enable the application of general tools from Alexandrov geometry to study gradient flows of certain functionals on the space of metric measure spaces. On the practical side, curvature estimates can be used to inform algorithms for computing Fréchet means of point clouds in a metric space—see, e.g., Turner et al. (2014); Chowdhury and Needham (2020); Zhang et al. (2024a) for the case of lower bounds, or Feragen et al. (2011); Bacák (2014) for the case of upper bounds. Additionally, the realization of the Z-GW space as a quotient of an L^p-space, described in Theorem 34, suggests an alternative approach to the theory and numerical modeling of gradient flows in the Z-GW space. We plan to explore these ideas in future work.

Topological Questions. We showed in Theorem 42 that, when $p < \infty$, the Z-GW space is contractible, perhaps surprisingly, independently of the topology of Z. However, the dependence of the topology of $\mathcal{M}_{\sim}^{Z,\infty}$ on that of Z has not been resolved—see Question 41 and Question 44. We show in Example 1 and the ensuing discussion that if Z is not path-connected, then neither is $\mathcal{M}_{\sim}^{Z,\infty}$, so the answer to these questions is potentially subtle. We remark here that the special case of $p = \infty$ is conceptually important—the distance GW_{∞}^{Z} is essentially a notion of Gromov-Hausdorff distance for Z-valued networks, and Gromov-Hausdorff distance for $\mathbb{R}$-valued networks is already a rich area of study, with ties to areas such as topological data analysis (Chowdhury and Mémoli, 2018b, 2023). Other interesting topological questions remain open. For example, Gromov's celebrated *precompactness theorem* (see Burago et al. 2022, Chapter 7) says that a family of metric spaces with certain uniform bounds on geometrical properties of its elements has compact closure in the Gromov-Hausdorff topology, and this is generalized to the Gromov-Wasserstein setting by Mémoli (2011a, Theorem 5.3). A natural question is whether there are interesting precompact families for Z-networks in the Z-GW topology.

Structure of Optimal Couplings. There has recently been significant interest in characterizing the structure of optimal couplings in the GW framework; in particular, several recent articles address the problem of finding general classes of measure networks under which the GW distance is realized by a measure-preserving map (e.g., Mémoli and Needham 2022a; Delon et al. 2022; Sturm 2023; Dumont et al. 2024; Clark et al. 2024). The Z-GW framework gives a more general setting for addressing these questions. The added flexibility in the model might allow for more tractable versions of the problem to be attacked.

Computational Pipeline. The focus of this paper is on the basic theory of Z-GW distances, but the results in Section 5 point toward more practical considerations. The lower bounds in Theorem 50 are polynomial-time computable in the standard GW setting; the computational efficiency for some of these lower bounds relies on the fact that Wasserstein distances between distributions on the real line can be computed via an explicit formula. From an applications perspective, it would be useful to characterize other spaces Z where the Wasserstein distance is efficiently and explicitly computable (e.g., the circle as done by Rabin et al. 2011), and to utilize this for efficient Z-GW distance lower bound computation. Another interesting line of research in this direction is to determine classes of Z-networks for which the lower bounds are *injective* in the sense that vanishing of the lower bound implies that the input Z-networks are weakly isomorphic—this question is quite subtle in the standard GW setting (Mémoli and Needham, 2022b). Finally, an important consequence of Theorem 52 is that, for arbitrary Z, Z-GW distances can be estimated via solvers for the $\mathbb{R}^n$-GW distance, as we outlined in Section 5.3. A serious implementation of this approximation scheme will be the subject of a follow-up paper.

Applications. The surplus of examples of Z-GW distances provided in Section 3 suggests the wide applicability of this framework. Once the computational pipeline described in the previous paragraph is in place, we plan to apply it to various problems involving analysis of complex and non-standard data. We are particularly interested in exploring the more novel settings described in Section 3.2, such as probabilistic metric spaces, shape graphs and connection graphs.

Acknowledgments

We would like to thank Samir Chowdhury for several enlightening discussions about the potential for Z-GW distances. T.N. was partially supported by NSF grants DMS–2107808, DMS–2324962 and CIF–2526630. F.M. acknowledges funding from the NSF under grants IIS–1901360, DMS–2301359 and CCF–2310412, and the BSF under grant 2020124. M.B. was partially supported by NSF grants DMS–2324962 and IIS–2426549 and the BSF under grant 2022076.

Appendix A. Proofs of Technical Results

In this section, we collect proofs of various technical results whose statements appeared in the main body of the paper.

A.1 Proof of Proposition 8

1. The proof of this part of the proposition can be found in Korevaar and Schoen (1993, Section 1.1).

2. The main idea of the proof is due to Majer (2024). By the theorem of Banach (Banach, 1987, Chapter XI, §8, Theorem 10), Y can be isometrically embedded onto a subset

of a separable Banach space $C([0,1])$ of real-valued continuous functions on $[0,1]$ with the sup-norm. By Pettis measurability theorem (Hytönen et al., 2016, Theorem 1.1.6), all measurable functions with values in Y are strongly measurable by separability of Y, so $L^p(X,\mu_X;Y)$ is a subset (or a closed subspace) of the Lebesgue-Bochner space $L^p(X,\mu_X;C([0,1]))$. Since X is separable, its Borel σ-algebra is countably generated, so $L^p(X,\mu_X;C([0,1]))$ is separable (Hytönen et al., 2016, Theorem 1.2.29). Therefore, $L^p(X,\mu_X;Y)$ is separable.

3. Let $\epsilon > 0$ be given. We first consider the case when f is bounded. By Lusin's theorem, there is a compact set $K \subset X$ such that $\mathscr{L}^d(X\backslash K) < \epsilon$ and f is continuous on K. Since f is continuous on a compact set K, it is uniformly continuous on K. Therefore, there exists δ such that for any $x, y \in K$, $\|x-y\|_{\mathbb{R}^d} < \delta$ implies $d_Y(f(x), f(y)) < \epsilon$. Now, let $h > 0$ be small enough so that the maximal distance of each hypercube in the grid is smaller than $\delta > 0$. Then, on each cube, take an arbitrary value of f and define a piecewise constant function g that is constant on each cube. Then, for any $x \in K$, f, g satisfies $d_Y(f(x), g(x)) < \epsilon$ so that $\sup_{x\in K} d_Y(f(x), g(x)) \leqslant \epsilon$. Now, since f is bounded, there exists $M > 0$ such that $d(f(x), f(x')) \leqslant M$ for any $x, x' \in X$. Therefore,

$$\begin{aligned} D_p(f,g)^p &= \int_X d_Y(f(x),g(x))^p \mathscr{L}^d(dx) \\ &= \int_K d_Y(f(x),g(x))^p \mathscr{L}^d(dx) + \int_{X\backslash K} d_Y(f(x),g(x))^p \mathscr{L}^d(dx) \\ &\leqslant \epsilon^p \mathscr{L}^d(K) + M^p \mathscr{L}^d(X\backslash K) \leqslant \epsilon^p + M^p \epsilon \end{aligned}$$

Since ϵ is arbitrary, we have the proposition for the bounded f. We will now consider unbounded f. Fix $y_0 \in Z$ and define a function f_n by

$$f_n(x) = \begin{cases} f(x) & \text{if } d_Y(f(x), z_0) \leqslant n \\ z_0 & \text{otherwise} \end{cases}$$

Then, $f_n(x) \to f(x)$ as $n \to \infty$ for any $x \in [0,1]^d$. By definition, we have $d_Y(f_n(x), z_0) \leqslant d_Y(f(x), z_0)$ so that $d_Y(f_n(x), f(x))^p \leqslant 2^p d_Y(f(x), z_0)^p$ by triangle inequality, and the L^p assumption on f allows us to dominate the function $d_Y(f_n(x), f(x))^p$ by an integrable function $2^p d_Y(f(x), z_0)$. Therefore, by Lebesgue's dominated convergence theorem, we have $D_p(f_n, f) \to 0$ as $n \to \infty$. Thus, taking a bounded function f_n such that $D_p(f_n, f) < \epsilon/2$ and a piecewise constant function g such that $D_p(f_n, g) < \epsilon/2$, we have $D_p(f,g) < \epsilon$.

■

A.2 Proof of Lemma 27

We start from the case when c is bounded. Suppose $0 \leqslant c \leqslant M$ for some $M \geqslant 0$. We note that $X \times X$ and $Y \times Y$ are again Polish and applying Lusin's theorem to a and b, we observe that there are compact sets $K_X \subset X \times X$ and $K_Y \subset Y \times Y$ such that $\mu \otimes \mu(X \times X \backslash K_X) < \delta$ and $\nu \otimes \nu(Y \times Y \backslash K_Y) < \delta$. Now, consider the homeomorphism $s : X \times X \times Y \times Y \ni (x, x', y, y') \mapsto (x, y, x', y') \in X \times Y \times X \times Y$ and the set $K = s(K_X \times K_Y)$.

By using s, we see that $c(a,b)$, seen as a function on $X \times Y \times X \times Y$, is continuous on K. Using that c satisfies $c \leqslant M$,

$$\int_{X\times Y}\int_{X\times Y} c(a,b)d(\gamma_n\otimes\gamma_n) \leqslant \int_{(X\times Y)^2} 1_K\cdot c(a,b)d(\gamma_n\otimes\gamma_n) + \int_{(X\times Y)^2} 1_{K^c}\cdot Md(\gamma_n\otimes\gamma_n)$$
$$= \int_{(X\times Y)^2} 1_K\cdot c(a,b)d(\gamma_n\otimes\gamma_n) + M\gamma_n\otimes\gamma_n(K^c)$$

By definition of K and s, we have

$$\begin{aligned}\gamma_n\otimes\gamma_n(K^c) &= \gamma_n\otimes\gamma_n(s((K_X\times K_Y)^c))\\ &= \gamma_n\otimes\gamma_n(s([K_X^c\times(Y\times Y)]\cup[(X\times X)\times K_Y^c]))\\ &= \gamma_n\otimes\gamma_n(s(K_X^c\times(Y\times Y))\cup s((X\times X)\times K_Y^c))\\ &\leqslant \gamma_n\otimes\gamma_n(s(K_X^c\times(Y\times Y))) + \gamma_n\otimes\gamma_n(s((X\times X)\times K_Y^c))\end{aligned}$$

We note that $(x,y,x',y') \in s(K_X^c\times(Y\times Y))$ if and only if $(x,x')\in K_X^c$, so we have $1_{s(K_X^c\times(Y\times Y))} = 1_{K_X^c}(x,x')$ for any $y,y'\in Y$. Therefore,

$$\begin{aligned}\gamma_n\otimes\gamma_n(s(K_X^c\times(Y\times Y))) &= \int_{X\times Y}\int_{X\times Y} 1_{K_X^c}(x,x')d(\gamma_n\otimes\gamma_n)\\ &= \int_X\int_X 1_{K_X^c}(x,x')d(\mu\otimes\mu)\\ &= \mu\otimes\mu(K_X^c) < \delta\end{aligned}$$

We can argue similarly for K_Y^c. Thus,

$$\int_{X\times Y}\int_{X\times Y} c(a,b)d(\gamma_n\otimes\gamma_n) \leqslant \int_{(X\times Y)^2} 1_K\cdot c(a,b)d(\gamma_n\otimes\gamma_n) + 2M\delta$$

We note that $1_K\cdot c$ is an upper semicontinuous function on $X\times Y$ because it is a nonnegative product of two upper semicontinuous functions. 1_K is upper semicontinuous because K is closed. Now recall that the weak convergence $m_k\to m$ of probability measures is equivalent to $\int fm \geqslant \limsup\int fm_k$ for any upper semicontinuous function f bounded from above. As $1_K c$ satisfies this constraint, using the fact that $\gamma_n\otimes\gamma_n\to\gamma\otimes\gamma$ weakly, we have

$$\begin{aligned}\limsup_n \int_{X\times Y}\int_{X\times Y} c(a,b)d(\gamma_n\otimes\gamma_n) &\leqslant \limsup_n \int_{X\times Y}\int_{X\times Y} 1_K c(a,b)d(\gamma_n\otimes\gamma_n) + 2M\delta\\ &\leqslant \int_{X\times Y}\int_{X\times Y} 1_K c(a,b)d(\gamma\otimes\gamma) + 2M\delta\\ &\leqslant \int_{X\times Y}\int_{X\times Y} c(a,b)d(\gamma\otimes\gamma) + 2M\delta\end{aligned}$$

Since δ is arbitrary, we have that

$$\limsup_n \int_{X\times Y}\int_{X\times Y} c(a,b)d(\gamma_n\otimes\gamma_n) \leqslant \int_{X\times Y}\int_{X\times Y} c(a,b)d(\gamma\otimes\gamma)$$

implying that the integral functional is upper semicontinuous. We can apply the same argument to $M - c$ to obtain that it is also lower semicontinuous. Overall, we have that the functional is continuous with respect to the weak convergence when c is bounded. Now, we will consider the case when c is unbounded. We can approximate the integral functional by the supremum of continuous functionals by replacing c by monotone sequence $c_n = \min(c, n)$, so the functional is lower semicontinuous. For the upper semicontinuity, the assumption that $c(a, b) \leqslant f(a) + g(b)$ with $\int (f \circ a) d(\mu \otimes \mu), \int (g \circ b) d(\mu \otimes \mu) < +\infty$ allows us to replace c by $f(a) + g(b) - c(a, b) \geqslant 0$ and argue the same. Therefore, the continuity is proven for unbounded cases. ■

A.3 Proof of Lemma 43

Observe that there is a unique coupling π between μ_X and $\delta_\star$, characterized by

$$\pi(dx \times d\star) = \mu_X(dx).$$

The integrals $I_{\pi\pi}$, $I_{X\pi}$ and $I_{\pi Y}$ in (16) can then be computed explicitly by applying the definitions:

$$\begin{aligned}
I_{\pi\pi} &= \iint\limits_{(X \amalg \{\star\})^4} d_Z(\omega_X \amalg \omega_Y(u, u'), \omega_X \amalg \omega_Y(v, v'))^p \pi(du \times dv) \pi(du' \times dv') \\
&= \iint\limits_{(X \times \{\star\})^2} d_Z(\omega_X \amalg \omega_Y(x, x'), \omega_X \amalg \omega_Y(\star, \star))^p \pi(dx \times d\star) \pi(dx' \times d\star) \\
&= \iint\limits_{X \times X} d_Z(\omega_X(x, x'), z)^p \mu_X(dx) \mu_X(dx') = \mathrm{size}_{p,z}(X)^p,
\end{aligned}$$

$$\begin{aligned}
I_{X\pi} &= \iint\limits_{(X \amalg \{\star\})^4} d_Z(\omega_X \amalg \omega_Y(u, u'), \omega_X \amalg \omega_Y(v, v'))^p \Delta_*^X \mu_X(du \times dv) \pi(du' \times dv') \\
&= \int_{X \times \{\star\}} \int_X d_Z(\omega_X \amalg \omega_Y(x, x'), \omega_X \amalg \omega_Y(x, \star))^p \mu_X(dx) \pi(dx' \times d\star) \\
&= \iint\limits_{X \times X} d_Z(\omega_X(x, x'), z)^p \mu_X(dx) \mu_X(dx') = \mathrm{size}_{p,z}(X)^p,
\end{aligned}$$

and, similarly,

$$\begin{aligned}
I_{\pi Y} &= \iint\limits_{(X \amalg \{\star\})^4} d_Z(\omega_X \amalg \omega_Y(u, u'), \omega_X \amalg \omega_Y(v, v'))^p \pi(du \times dv) \Delta_*^Y \mu_Y(du' \times dv') \\
&= \int_{X \times \{\star\}} \int_{\{\star\}} d_Z(\omega_X \amalg \omega_Y(x, \star), \omega_X \amalg \omega_Y(\star, \star))^p \pi(dx \times d\star) \delta_\star(d\star) \\
&= \int_{X \times \{\star\}} \int_{\{\star\}} d_Z(z, z)^p \pi(dx \times d\star) = 0.
\end{aligned}$$

The estimate (16) therefore becomes

$$2^p \mathrm{GW}_p^Z(X_s, X_t)^p \leqslant |t-s| \cdot 3 \cdot \mathrm{size}_{p,z}(X)^p,$$

and this can be rewritten as (19). ■

A.4 Proof of Lemma 51

To prove the lemma, we use the following result regarding **analytic sets**, i.e., continuous images of Polish spaces. This result can be found in (Varadarajan, 1963, Lemma 2.2) or (Chowdhury and Mémoli, 2019, Lemma 27)

Lemma 54. *Let X, Y be analytic subsets of Polish spaces equipped with the relative Borel σ-fields. Let $f : X \to Y$ be a surjective, Borel-measurable map. Then for any $\nu \in \mathrm{Prob}(Y)$, there exists $\mu \in \mathrm{Prob}(X)$ such that $\nu = f_*\mu$. Here,* $\mathrm{Prob}(X)$ *is the set of Borel probability measures on X.*

Proof of Lemma 51. The $\subset$ direction is standard (Chowdhury and Mémoli, 2019). For the opposite direction, we briefly point out that the proof by Chowdhury and Mémoli (2019) contains a gap; to fix it, take a coupling $\pi \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)$, and our goal is to find a measure $\sigma \in \mathcal{C}(\mu_X, \mu_Y)$ such that $T_*\sigma = \pi$. The strategy by Chowdhury and Mémoli (2019) was to apply Lemma 54 directly to the mapping T so that we have a probability measure σ such that $\pi = T_*\sigma$, but the issue here is that Lemma 54 only ensures that σ is a probability measure and not necessarily a coupling. Indeed, if f, g are constant maps, $\mathcal{C}(f_*\mu_X, g_*\mu_Y)$ only consists of a Dirac measure, and importantly, any probability measure σ (not necessarily a coupling) on $X \times Y$ satisfies $\pi = T_*\sigma$ for the element $\pi \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)$. To impose an additional restriction that σ is a coupling, we consider the augmented T mapping $\bar{T} : X \times Y \to X \times Z \times Z \times Y$,

$$\bar{T}(x, y) = (\pi_X(x, y), T(x, y), \pi_Y(x, y)) = (x, f(x), g(y), y)$$

where $\pi_X : X \times Y \to X$ is the projection onto the first component and π_Y similarly. Intuitively, we added the projection functions π_X, π_Y so that the condition $\mu_X \otimes \pi \otimes \mu_Y = \bar{T}_*\sigma = (\pi_X, T, \pi_Y)_*\sigma$ we will obtain from Lemma 54 induces $(\pi_X)_*\sigma = \mu_X, (\pi_Y)_*\sigma = \mu_Y$, i.e., the coupling condition.

Since X, Y, Z are Polish, $X \times Y$ and $X \times Z \times Z \times Y$ are again Polish, and $\bar{T}$ can be made surjective by restricting the range to $\bar{T}(X \times Y) = X \times T(X \times Y) \times Y$. The measurability of f, g implies $\bar{T}$ is measurable, so we apply Lemma 54 to the measure $\mu_X \otimes \pi|_{T(X\times Y)} \otimes \mu_Y$ where $\pi|_{T(X\times Y)}(A) = \pi(A \cap T(X \times Y))$ and the mapping $\bar{T}$ to obtain a probability measure σ on $X \times Y$ such that $\mu_X \otimes \pi|_{T(X\times Y)} \otimes \mu_Y = \bar{T}_*\sigma$. σ belongs to $\mathcal{C}(\mu_X, \mu_Y)$ because, by definition, for any measurable $A \subset X$, we have

$$\bar{T}_*\sigma(A \times T(X \times Y) \times Y) = \sigma(A \times Y) = \mu_X(A)\pi|_{T(X\times Y)}(T(X \times Y))\mu_Y(Y) = \mu_X(A)$$

and similarly for Y. Here, the first equality comes from the definition of $\bar{T}^*$ and pushforward, the second is the condition $\mu_X \otimes \pi|_{T(X\times Y)} \otimes \mu_Y = \bar{T}_*\sigma$, and the last one is $\mu_Y(Y) = 1$ and that $\pi|_{T(X\times Y)}(T(X \times Y)) = \pi(T(X \times Y)) = \pi(f(X) \times g(Y)) = 1$ since $\pi \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)$.

Now, we have $\pi|_{T(X\times Y)} = T_*\sigma$ because, for any $B \subset T(X \times Y)$, we have $\bar{T}_*\sigma(X \times B \times Y) = T_*\sigma(B) = \pi|_{T(X\times Y)}(B)$. Finally, notice that $\pi(B \cap T(X \times Y)) = \pi(B)$ for any measurable $B \subset Z \times Z$, so $T_*\sigma = \pi|_{T(X\times Y)} = \pi$. ■

References

D. Alvarez-Melis and T. Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2018.

S. Arya, A. Auddy, R. Edmonds, S. Lim, F. Mémoli, and D. Packer. The Gromov–Wasserstein distance between spheres. *Foundations of Computational Mathematics*, pages 1–61, 2024.

M. Bacák. Computing medians and means in Hadamard spaces. *SIAM journal on optimization*, 24(3):1542–1566, 2014.

H. Bahouri. *Fourier analysis and nonlinear partial differential equations*. Springer, 2011.

A. B. Bal, X. Guo, T. Needham, and A. Srivastava. Statistical shape analysis of shape graphs with applications to retinal blood-vessel networks. *arXiv preprint arXiv:2211.15514*, 2022.

A. B. Bal, X. Guo, T. Needham, and A. Srivastava. Statistical analysis of complex shape graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

S. Banach. *Theory of Linear Operations*. Elsevier, 1987. ISBN 9780444701130. Reprint of the 1932 original.

M. Bauer, N. Charon, E. Klassen, S. Kurtek, T. Needham, and T. Pierron. Elastic metrics on spaces of Euclidean curves: Theory and Algorithms. *Journal of Nonlinear Science*, 34 (3):1–37, 2024.

T. Bhamre, T. Zhang, and A. Singer. Orthogonal matrix retrieval in cryo-electron microscopy. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 1048–1052. IEEE, 2015.

C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.

C. Bothorel, J. D. Cruz, M. Magnani, and B. Micenkova. Clustering attributed graphs: models, measures and methods. *Network Science*, 3(3):408–444, 2015.

D. Burago, I. Burago, and S. Ivanov. *A Course in Metric Geometry*. Crm Proceedings & Lecture Notes. American Mathematical Society, 2001. ISBN 9780821821299. URL `https://books.google.co.jp/books?id=dRmIAwAAQBAJ`.

D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, 2022.

G. Carlsson and F. Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11(47):1425–1470, 2010. URL `http://jmlr.org/papers/v11/carlsson10a.html`.

L. Chapel, M. Z. Alaya, and G. Gasso. Partial Gromov-Wasserstein with applications on positive-unlabeled learning. *Advances in Neural Information Processing Systems*, 2020.

F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot. Gromov-Hausdorff stable signatures for shapes using persistence. In *Computer Graphics Forum*, volume 28, pages 1393–1403. Wiley Online Library, 2009.

F. Chazal, M. Glisse, C. Labruère, and B. Michel. Convergence rates for persistence diagram estimation in topological data analysis. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 163–171, Beijing, China, 22–24 June 2014. PMLR. URL `https://proceedings.mlr.press/v32/chazal14.html`.

J. Cheeger and T. H. Colding. On the structure of spaces with Ricci curvature bounded below. I. *Journal of Differential Geometry*, 46(3):406–480, 1997.

S. Chen, S. Lim, F. Mémoli, Z. Wan, and Y. Wang. Weisfeiler-Lehman meets Gromov-Wasserstein. In *International Conference on Machine Learning*, pages 3371–3416. PMLR, 2022.

S. Chen, S. Lim, F. Mémoli, Z. Wan, and Y. Wang. The Weisfeiler-Lehman distance: Reinterpretation and connection with GNNs. In *Topological, Algebraic and Geometric Learning Workshops 2023*, pages 404–425. PMLR, 2023.

S. Chowdhury and F. Mémoli. Explicit geodesics in Gromov-Hausdorff space. *Electronic Research Announcements*, 25:48–59, 2018a.

S. Chowdhury and F. Mémoli. A functorial Dowker theorem and persistent homology of asymmetric networks. *Journal of Applied and Computational Topology*, 2:115–175, 2018b.

S. Chowdhury and F. Mémoli. The Gromov–Wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787, November 2019. ISSN 2049-8772. doi: 10.1093/imaiai/iaz026. URL `https://doi.org/10.1093/imaiai/iaz026`.

S. Chowdhury and F. Mémoli. Distances and isomorphism between networks: Stability and convergence of network invariants. *Journal of Applied and Computational Topology*, 7(2): 243–361, 2023.

S. Chowdhury and T. Needham. Gromov-Wasserstein averaging in a Riemannian framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 842–843, 2020.

S. Chowdhury and T. Needham. Generalized spectral clustering via Gromov-Wasserstein learning. In *International Conference on Artificial Intelligence and Statistics*, pages 712–720. PMLR, 2021.

S. Chowdhury, D. Miller, and T. Needham. Quantized Gromov-Wasserstein. In *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part III 21*, pages 811–827. Springer, 2021.

S. Chowdhury, T. Needham, E. Semrad, B. Wang, and Y. Zhou. Hypergraph co-optimal transport: Metric and categorical properties. *Journal of Applied and Computational Topology*, pages 1–60, 2023.

R. A. Clark, T. Needham, and T. Weighill. Generalized dimension reduction using semi-relaxed Gromov-Wasserstein distance. *arXiv preprint arXiv:2405.15959*, 2024.

J. Delon, A. Desolneux, and A. Salmona. Gromov–Wasserstein distances between Gaussian distributions. *Journal of Applied Probability*, 59(4):1178–1198, 2022.

P. Demetci, R. Santorella, B. Sandstede, W. S. Noble, and R. Singh. SCOT: single-cell multi-omics alignment with optimal transport. *Journal of Computational Biology*, 29(1): 3–18, 2022.

T. Dumont, T. Lacombe, and F.-X. Vialard. On the existence of Monge maps for the Gromov–Wasserstein problem. *Foundations of Computational Mathematics*, pages 1–48, 2024.

A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze. Means in spaces of tree-like shapes. In *2011 International Conference on Computer Vision*, pages 736–746. IEEE, 2011.

R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

M. Gromov. *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser, December 2006. ISBN 9780817645823. URL `https://books.google.com/books/about/Metric_Structures_for_Riemannian_and_Non.html?hl=&id=KR1engEACAAJ`.

X. Guo, A. Basu Bal, T. Needham, and A. Srivastava. Statistical shape analysis of brain arterial networks (BAN). *The Annals of Applied Statistics*, 16(2):1130–1150, 2022.

T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces: Volume I: Martingales and Littlewood-Paley Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. Springer, 2016. ISBN 978-3-319-48519-5. doi: 10.1007/978-3-319-48520-1.

B. J. Jain and K. Obermayer. Structure spaces. *Journal of Machine Learning Research*, 10 (11), 2009.

S. Janson. Graphons, cut norm and distance, couplings and rearrangements. *arXiv preprint arXiv:1009.2376*, 2010.

A. Kasue and H. Kumura. Spectral convergence of Riemannian manifolds. *Tohoku Mathematical Journal, Second Series*, 46(2):147–179, 1994.

K. Kawano, S. Koide, H. Shiokawa, and T. Amagasa. Multi-dimensional fused Gromov-Wasserstein discrepancy for edge-attributed graphs. *IEICE TRANSACTIONS on Information and Systems*, 107(5):683–693, 2024.

T. Kerdoncuff, R. Emonet, and M. Sebban. Sampled Gromov Wasserstein. *Machine Learning*, 110(2-3):2151–2186, 2021. doi: 10.1007/s10994-021-06035-1. URL `https://doi.org/10.1007/s10994-021-06035-1`.

W. Kim. *The persistent topology of dynamic data*. PhD thesis, The Ohio State University, 2020. `http://rave.ohiolink.edu/etdc/view?acc_num=osu1587503336988272`.

W. Kim and F. Mémoli. Spatiotemporal persistent homology for dynamic metric spaces. *Discrete & Computational Geometry*, 66:831–875, 2021.

W. Kim, F. Mémoli, and Z. Smith. Analysis of dynamic graphs and dynamic metric spaces via zigzag persistence. In *Topological Data Analysis: The Abel Symposium 2018*, pages 371–389. Springer, 2020.

N. J. Korevaar and R. M. Schoen. Sobolev spaces and harmonic maps for metric space targets. *Communications in Analysis and Geometry*, 1:561–659, 1993. URL `https://api.semanticscholar.org/CorpusID:55639601`.

I. Kramosil and J. Michálek. Fuzzy metrics and statistical metric spaces. *Kybernetika*, 11 (5):336–344, 1975.

M. Li, J. Yu, H. Xu, and C. Meng. Efficient approximation of Gromov-Wasserstein distance using importance sparsification. *Journal of Computational and Graphical Statistics*, 32 (4):1512–1523, 2023.

S. Liang, M. P. Segundo, S. N. Aakur, S. Sarkar, and A. Srivastava. Shape-graph matching network (SGM-net): Registration for statistical shape analysis. *arXiv preprint arXiv:2308.06869*, 2023.

S. Lim. *Geometry, Topology, and Spectral Methods in Data Analysis: from Injective Metric Spaces, through Gromov-type Distances, to Generalized MDS*. PhD thesis, The Ohio State University, 2021. `http://rave.ohiolink.edu/etdc/view?acc_num=osu1618959416483507`.

L. Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

P. Majer. For Polish X, Y, $L^p(X, Y)$ is separable. MathOverflow, 2024. URL `https://mathoverflow.net/q/469072`. version: 2024-04-14.

F. Mémoli. On the use of Gromov-Hausdorff Distances for Shape Comparison. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. ISBN 978-3-905673-51-7. doi: 10.2312/SPBG/SPBG07/081-090.

F. Mémoli. Spectral Gromov-Wasserstein distances for shape matching. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 256–263. IEEE, 2009.

F. Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, August 2011a. doi: 10.1007/s10208-011-9093-5. URL https://doi.org/10.1007/s10208-011-9093-5.

F. Mémoli. A spectral notion of Gromov–Wasserstein distance and related methods. *Applied and Computational Harmonic Analysis*, 30(3):363–401, 2011b.

F. Mémoli. Some properties of Gromov–Hausdorff distances. *Discrete & Computational Geometry*, 48:416–440, 2012.

F. Mémoli and T. Needham. Comparison results for Gromov-Wasserstein and Gromov-Monge distances. *arXiv preprint arXiv:2212.14123*, 2022a.

F. Mémoli and T. Needham. Distance distributions and inverse problems for metric measure spaces. *Studies in Applied Mathematics*, 149(4):943–1001, 2022b.

F. Mémoli and G. Sapiro. Comparing point clouds. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 32–40, 2004.

F. Mémoli and G. Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics*, 5:313–347, 2005.

F. Mémoli and Z. Wan. Characterization of Gromov-type geodesics. *Differential Geometry and its Applications*, 88:102006, 2023. ISSN 0926-2245. doi: https://doi.org/10.1016/j.difgeo.2023.102006. URL https://www.sciencedirect.com/science/article/pii/S0926224523000323.

F. Mémoli, A. Munk, Z. Wan, and C. Weitkamp. The ultrametric Gromov–Wasserstein distance. *Discrete & Computational Geometry*, 70(4):1378–1450, 2023.

K. Menger. Statistical metrics. *Proceedings of the National Academy of Sciences of the United States of America*, 28(12):535–537, 1942.

M. I. Ostrovskii. *Metric Embeddings*. De Gruyter, Berlin, Boston, 2013. ISBN 9783110264012. doi: 10.1515/9783110264012. URL https://doi.org/10.1515/9783110264012.

G. Peyré, M. Cuturi, and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International conference on machine learning*, pages 2664–2672. PMLR, 2016.

I. Pinelis. Does complete and separable Wasserstein space imply a complete base space? MathOverflow, 2024. URL https://mathoverflow.net/q/470226. version: 2024-04-29.

J. Rabin, J. Delon, and Y. Gousseau. Transportation distances on the circle. *Journal of Mathematical Imaging and Vision*, 41(1):147–167, 2011.

S. Robertson, D. Kohli, G. Mishne, and A. Cloninger. On a generalization of Wasserstein distance and the Beckmann problem to connection graphs. *arXiv preprint arXiv:2312.10295*, 2023.

S. Rosenberg. *The Laplacian on a Riemannian manifold: an introduction to analysis on manifolds.* Number 31 in London Mathematical Society Student Texts. Cambridge University Press, 1997.

F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs and Modeling.* Birkhäuser, Cham, Switzerland, 2015. ISBN 9783319208282. doi: 10.1007/978-3-319-20828-2.

M. Scetbon, G. Peyré, and M. Cuturi. Linear-time Gromov-Wasserstein distances using low rank couplings and costs. In *International Conference on Machine Learning*, pages 19347–19365. PMLR, 2022.

R. L. Schilling and F. Kühn. *Counterexamples in Measure and Integration.* Cambridge University Press, 2021.

B. Schweizer and A. Sklar. Statistical metric spaces. *Pacific J. Math*, 10(1):313–334, 1960.

T. Shioya. *Metric Measure Geometry.* European Mathematical Society, 2016. ISBN 9783037191583. URL https://books.google.com/books/about/Metric_Measure_Geometry.html?hl=&id=8M5GjwEACAAJ.

A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.

A. Srivastava and E. P. Klassen. *Functional and shape data analysis*, volume 1. Springer, 2016.

A. Srivastava, X. Guo, and H. Laga. Advances in geometrical analysis of topologically-varying shapes. In *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, pages 1–4. IEEE, 2020.

S. M. Srivastava. *A course on Borel sets*, volume 180. Springer Science & Business Media, 2008.

K. Sturm. *The Space of Spaces: Curvature Bounds and Gradient Flows on the Space of Metric Measure Spaces.* Memoirs of the American Mathematical Society. American Mathematical Society, 2023. ISBN 9781470466961. URL https://books.google.com/books?id=CEPsEAAAQBAJ.

Y. Sukurdeep, M. Bauer, and N. Charon. A new variational model for shape graph registration with partial matching constraints. *SIAM Journal on Imaging Sciences*, 15(1): 261–292, 2022.

D. J. Sumpter. *Collective animal behavior.* Princeton University Press, 2010.

K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52:44–70, 2014.

V. S. Varadarajan. Groups of automorphisms of Borel spaces. *Transactions of the American Mathematical Society*, 109(2):191–220, 1963. ISSN 0002-9947. doi: 10.1090/s0002-9947-1963-0159923-5. URL http://dx.doi.org/10.1090/S0002-9947-1963-0159923-5.

T. Vayer, N. Courty, R. Tavenard, and R. Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284. PMLR, 2019.

T. Vayer, L. Chapel, R. Flamary, R. Tavenard, and N. Courty. Fused Gromov-Wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020a.

T. Vayer, I. Redko, R. Flamary, and N. Courty. Co-optimal transport. *Advances in Neural Information Processing Systems*, 33:17559–17570, 2020b.

S. Vedula, V. Maiorca, L. Basile, F. Locatello, and A. Bronstein. Scalable unsupervised alignment of general metric and non-metric structures. *arXiv preprint arXiv:2406.13507*, 2024.

C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003. ISBN 9780821833124. URL https://books.google.com/books?id=idyFAwAAQBAJ.

A. Wald. On a statistical generalization of metric spaces. *Proceedings of the National Academy of Sciences*, 29(6):196–197, 1943.

H. Xu. Gromov-Wasserstein factorization models for graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 6478–6485, 2020.

H. Xu, D. Luo, and L. Carin. Scalable Gromov-Wasserstein learning for graph partitioning and matching. *Advances in Neural Information Processing Systems*, 32, 2019.

J. Yang, M. Labeau, and F. d'Alché Buc. Exploiting edge features in graphs with fused network Gromov-Wasserstein distance. *arXiv preprint arXiv:2309.16604*, 2023.

S. Y. Zhang, F. Lan, Y. Zhou, A. Barbensi, M. P. Stumpf, B. Wang, and T. Needham. Geometry of the space of partitioned networks: A unified theoretical and computational framework. *arXiv preprint arXiv:2409.06302*, 2024a.

S. Y. Zhang, M. P. Stumpf, T. Needham, and A. Barbensi. Topological optimal transport for geometric cycle matching. *arXiv preprint arXiv:2403.19097*, 2024b.