THE GROMOV-WASSERSTEIN DISTANCE BETWEEN NETWORKS AND STABLE NETWORK INVARIANTS

SAMIR CHOWDHURY AND FACUNDO MÉMOLI

Department of Mathematics, The Ohio State University.

ABSTRACT. We define a metric—the Network Gromov-Wasserstein distance—on weighted, directed networks that is sensitive to the presence of outliers. In addition to proving its theoretical properties, we supply easily computable network invariants that approximate this distance by means of lower bounds.

CONTENTS

1. INTRODUCTION

1.1. **Motivation and related literature.** Advances in data mining are beginning to lead to the acquisition of large networks that are directed, weighted, and possibly even signed [KSSF16]. In light of the ready availability of such data, a natural problem is to devise methods for comparing network datasets. These methods in turn lead to a wide range of applications. An example is the *network retrieval task*: given a database of networks and a query network, return an ordered list of the networks in the database that are most similar to the query. Additionally, because there may be redundant data in the networks that are not relevant to the query, one may wish to impose a notion of significance to certain substructures of the query network. The task then is to retrieve networks which are similar to the query network both globally and also at the scale of relevant substructures.

E-mail address: chowdhury.57@osu.edu, memoli@math.osu.edu.

Date: August 14, 2018.

While there has been some work in devising directed, weighted analogues of conventional network analysis tools such as edge overlap and clustering coefficients, we are more interested in pairwise comparison of individual networks. The intuitive idea behind this comparison is to search for the best possible alignment of edges (according to weights) while simultaneously aligning nodes with similar significance.

Techniques based on optimal transport provide an elegant solution to this problem by endowing a network with a probability measure. The user adjusts the measure to signify important network substructures and to smooth out the effect of outliers. This approach was adopted in [Hen16] to compare various real-world network datasets modeled as *metric measure (mm) spaces*—metric spaces equipped with a probability measure. This work was based in turn on the formulation of the *Gromov-Wasserstein (GW) distance* between mm spaces presented in [Mém07, Mém11].

An alternative definition of the GW distance due to Sturm (the *transportation* formulation) appeared in [Stu06], although this formulation is less amenable to practical computations than the one in [Mém07] (the *distortion* formulation). Both the transportation and distortion formulations were studied carefully in [Mém07, Mém11, Stu12]. It was further observed by Sturm in [Stu12] that the definition of the (distortion) GW distance can be extended to *gauged measure spaces* of the form (X, d_X, μ_X). Here X is a Polish space, d_X is a symmetric L^2 function on $X \times X$ (that does not necessarily satisfy the triangle inequality), and μ_X is a Borel probability measure on X. These results are particularly important in the context of the current paper.

Exact computation of GW distances amounts to solving a nonconvex quadratic program. Towards this end, the computational techniques presented in [Mém07, Mém11] included both readily-computable lower bounds and an alternate minimization scheme for reaching a local minimum of the GW objection function. This alternate minimization scheme involved solving successive linear optimization problems, and was used for the computations in [Hen16].

From now on, we reserve the term *network* for network datasets that cannot necessarily be represented as metric spaces, unless qualified otherwise. Already in [Hen16], it was observed that numerical computation of GW distances between networks worked well for network comparison even when the underlying datasets failed to be metric. This observation was further developed in [PCS16], where the focus from the outset was to define generalized discrepancies between matrices that are not necessarily metric.

On the computational front, the authors of [PCS16] directly attacked the nonconvex optimization problem by considering an *entropy-regularized* form of the GW distance (ERGW) following [SPKS16], and using a projected gradient descent algorithm based on results in [BCC+15, SPKS16]. This approach was also used (for a generalized GW distance) on graph-structured datasets in [VCF+18]. It was pointed out in [VCF+18] that the gradient descent approach for the ERGW problem occasionally requires a large amount of regularization to obtain convergence, and that this could possibly lead to over-regularized solutions. A different approach, developed in [Mém07, Mém11], considers the use of lower bounds on the GW distance as opposed to solving the full GW optimization problem. This is a practical approach for many use cases, in which it may be sufficient to simply obtain lower bounds for the GW distance.

In the current paper, we use the GW distance formulation to define and develop a metric structure on the "space of networks". Additionally, by following the approaches used in [Mém07, Mém11], we are able to produce quantitatively stable network invariants that produce polynomial-time lower bounds on this Network GW distance. We defer experiments on network datasets to a future update of this paper.

1.2. **Organization of the paper.** In the following section, we define some notation and terms that will be used throughout the paper. §2 contains details about couplings and the Network Gromov-Wasserstein and Gromov-Prokhorov distances. In §3 we present network invariants along with quantitative stability results that yield lower bounds on the GW distance.

1.3. **Notation and basic terminology.** The indicator function of a set S is denoted $\mathbb{1}_S$. We denote measure spaces via the triple $(X, \mathcal{F}, \mu)$, where X is a set, $\mathcal{F}$ is a σ-field on X, and μ is the measure on $\mathcal{F}$. Given a measure space $(X, \mathcal{F}, \mu)$, we write $L^0 = L^0(\mu)$ to denote the collection of $\mathcal{F}$-measurable functions $f : X \to \mathbb{R}$. For all $p \in (0, \infty)$ and all $f \in L^0$, we define $\|f\|_p := (\int |f|^p \, d\mu)^{1/p}$. For $p = \infty$, $\|f\|_\infty := \inf\{M \in [0, \infty] : \mu(|f| > M) = 0\}$. Then for any $p \in (0, \infty]$, $L^p = L^p(\mu) := \{f \in L^0 : \|f\|_p < \infty\}$.

Given a measurable real-valued function $f : X \to \mathbb{R}$ and $t \in \mathbb{R}$, we will occasionally write $\{f \le t\}$ to denote the set $\{x \in X : f(x) \le t\}$.

Lebesgue measure on the reals will be denoted by λ. We write λ_I to denote the Lebesgue measure on the unit interval $I = [0, 1]$.

Suppose we have a measure space $(X, \mathcal{F}, \mu)$, a measurable space $(Y, \mathcal{G})$, and a measurable function $f : X \to Y$. The *pushforward* or *image measure* of f is defined to be the measure $f_*\mu$ on $\mathcal{G}$ given by writing $f_*\mu(A) := \mu(f^{-1}[A])$ for all $A \in \mathcal{G}$.

A particular case where we deal with pushforward measures is the following: given a product space $\mathcal{X} = X_1 \times X_2 \times \ldots \times X_n$ and a measure μ on $\mathcal{X}$, the canonical projection maps $\pi_i : \mathcal{X} \to X_i$, for $i = 1, \ldots, n$, define pushforward measures that we denote $(\pi_i)_*\mu$. If each X_i is itself a measure space with measure μ_i, then we say that μ has *marginals* μ_i, for $i = 1, \ldots, n$, if $(\pi_i)_*\mu = \mu_i$ for each i. We also consider projection maps of the form $(\pi_i, \pi_j, \pi_k) : \mathcal{X} \to X_i \times X_j \times X_k$ for $i, j, k \in \{1, \ldots, n\}$, and denote the corresponding pushforward by $(\pi_i, \pi_j, \pi_k)_*\mu$. Notice that we can take further projections of the form $(\pi_i, \pi_j)^{ijk} : X_i \times X_j \times X_k \to X_i \times X_j$, and the images of these projections are precisely those given by projections of the form $(\pi_i, \pi_j) : \mathcal{X} \to X_i \times X_j$.

Remark 1. Let $\mathcal{X} = X_1 \times X_2 \times \ldots \times X_n$ be a product space with a measure μ as above, and suppose each X_i is equipped with a measure μ_i such that μ has marginals μ_i. Let $i, j, k \in \{1, \ldots, n\}$. Then the measure $(\pi_i, \pi_j, \pi_k)_*\mu$ on $X_i \times X_j \times X_k$ has marginals $(\pi_i, \pi_k)_*\mu$ and $(\pi_j)_*\mu$ on $X_i \times X_k$ and X_j, respectively. To see this, let the projections $X_i \times X_j \times X_k \to X_i \times X_k$ and $X_i \times X_j \times X_k \to X_j$ be denoted $(\pi_i, \pi_k)^{ijk}$ and $(\pi_j)^{ijk}$, respectively. Let $B \subseteq X_j$ be measurable. Then

$$\begin{aligned}(\pi_j)^{ijk}_*(B) = (\pi_i, \pi_j, \pi_k)_*\mu(X_i \times B \times X_k) &= \mu(X_1 \times \ldots \times X_{j-1} \times B \times X_{j+1} \times \ldots \times X_n) \\ &= (\pi_j)_*(B).\end{aligned}$$

Next let $A \subseteq X_i$ and $C \subseteq X_k$ be measurable. Then

$$\begin{aligned}(\pi_i, \pi_k)^{ijk}_*(A \times C) &= (\pi_i, \pi_j, \pi_k)_*\mu(A \times X_j \times C) \\ &= \mu(X_1 \times \ldots \times X_{i-1} \times A \times X_{i+1} \times \ldots \times X_{k-1} \times C \times X_{k+1} \times \ldots \times X_n) \\ &= (\pi_i, \pi_k)_*(A \times C).\end{aligned}$$

2. THE STRUCTURE OF MEASURE NETWORKS

Let X be a Polish space with Borel σ-field denoted by writing $\mathrm{Borel}(X)$, and let μ_X be a Borel probability measure on $\mathrm{Borel}(X)$. We will write $\mathrm{Prob}(X)$ to denote the collection of Borel probability measures supported on X. For each $1 \le p < \infty$, denote by $L^p(\mu_X)$ the space of μ_X-measurable functions $f : X \to \mathbb{R}$ such that $|f|^p$ is μ_X-integrable. For $p = \infty$, denote by $L^p(\mu_X)$ the space of essentially bounded μ_X-measurable functions, i.e. functions that are bounded except on a set of measure zero. Formally, these spaces are equivalence classes of functions, where functions are equivalent if they agree μ_X-a.e.

We write $\mu_X \otimes \mu_X$ (equivalently $\mu_X^{\otimes 2}$) to denote the product measure on $\mathrm{Borel}(X) \otimes \mathrm{Borel}(X)$ (equivalently $\mathrm{Borel}(X)^{\otimes 2}$). Next let $\omega_X \in L^\infty(\mu_X^{\otimes 2})$. Then ω_X is essentially bounded. Since μ_X is finite, it follows that $|\omega_X|^p$ is integrable for any $1 \le p < \infty$.

By a *measure network*, we mean a triple (X, ω_X, μ_X). The naming convention arises from the case when X is finite; in such a case, we can view the pair (X, ω_X) as a complete directed graph with asymmetric real-valued edge weights. Accordingly, the points of X are called *nodes*, pairs of nodes are called *edges*, and ω_X is called the *edge weight function* of X. The collection of all measure networks will be denoted $\mathcal{N}$.

Remark 2. Sturm has studied symmetric, L^2 versions of measure networks (called *gauged measure spaces*) in [Stu12], and we point to his work as an excellent reference on the geometry of such spaces. Our motivation comes from studying networks, hence the difference in our naming conventions.

The information contained in a network should be preserved when we relabel the nodes in a compatible way; we formalize this idea by the following notion of *strong isomorphism* of measure networks.

Definition 1 (Strong isomorphism)**.** To say $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y)\in\mathcal{N}$ are *strongly isomorphic* means that there exists a Borel measurable bijection $\varphi:\mathrm{supp}(X)\to\mathrm{supp}(Y)$ (with Borel measurable inverse φ^{-1}) such that

- $\omega_X(x,x')=\omega_Y(\varphi(x),\varphi(x'))$ for all $x,x'\in\mathrm{supp}(X)$, and
- $\varphi_*\mu_X=\mu_Y$.

We will denote a strong isomorphism between measure networks by $X\cong^s Y$.

Example 3. Networks with one or two nodes will be very instructive in providing examples and counterexamples, so we introduce them now with some special terminology.

- By $N_1(a)$ we will refer to the network with one node $X=\{p\}$, a weight $\omega_X(p,p)=a$, and the Dirac measure $\delta_p=\mathbb{1}_p$.
- By $N_2(\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right),\alpha,\beta)$ we will mean a two-node network with node set $X=\{p,q\}$, and weights and measures given as follows:

$$\begin{aligned} \omega_X(p,p)&=a & \mu_X(\{p\})&=\alpha\\ \omega_X(p,q)&=b & \mu_X(\{q\})&=\beta\\ \omega_X(q,p)&=c\\ \omega_X(q,q)&=d \end{aligned}$$

- Given a k-by-k matrix $\Sigma\in\mathbb{R}^{k\times k}$ and a $k\times 1$ vector $v\in\mathbb{R}^k_+$ with sum 1, we automatically obtain a network on k nodes that we denote as $N_k(\Sigma,v)$. Notice that $N_k(\Sigma,v)\cong_1 N_\ell(\Sigma',v')$ if and only if $k=\ell$ and there exists a permutation matrix P of size k such that $\Sigma'=P\,\Sigma\,P^T$ and $PA=A'$.

Notation. Even though μ_X takes sets as its argument, we will often omit the curly braces and use $\mu_X(p,q,r)$ to mean $\mu_X(\{p,q,r\})$.

We wish to define a notion of distance on $\mathcal{N}$ that is compatible with isomorphism. A natural analog is the Gromov-Wasserstein distance defined between metric measure spaces [Mém07]. To adapt that definition for our needs, we first recall the definition of a measure coupling.

2.1. **Couplings and the distortion functional.** Let $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y)$ be two measure networks. A *coupling* between these two networks is a probability measure μ on $X\times Y$ with marginals μ_X and μ_Y, respectively. Stated differently, couplings satisfy the following property:

$$\mu(A\times Y)=\mu_X(A)\text{ and }\mu(X\times B)=\mu_Y(B),\text{ for all }A\in\mathrm{Borel}(X)\text{ and for all }B\in\mathrm{Borel}(Y).$$

The collection of all couplings between (X,ω_X,μ_X) and (Y,ω_Y,μ_Y) will be denoted $\mathscr{C}(\mu_X,\mu_Y)$, abbreviated to $\mathscr{C}$ when the context is clear.

In the case where we have a coupling μ between two measures ν,ν' on the same network (X,ω_X), the quantity $\mu(A\times B)$ is interpreted as the amount of mass transported from A to B when interpolating between the two distributions ν and ν'. In this special case, a coupling is also referred to as a *transport plan*.

Here we also recall that the product σ-field on $X\times Y$, denoted $\mathrm{Borel}(X)\otimes\mathrm{Borel}(Y)$, is defined as the σ-field generated by the measurable rectangles $A\times B$, where $A\in\mathrm{Borel}(X)$ and $B\in\mathrm{Borel}(Y)$. Because our spaces are all Polish, we always have $\mathrm{Borel}(X\times Y)=\mathrm{Borel}(X)\otimes\mathrm{Borel}(Y)$ [Bog07a, Lemma 6.4.2].

The product measure $\mu_X\otimes\mu_Y$ is defined on the measurable rectangles by writing

$$\mu_X\otimes\mu_Y(A\times B):=\mu_X(A)\mu_X(B),\text{ for all }A\in\mathrm{Borel}(X)\text{ and for all }B\in\mathrm{Borel}(Y).$$

By a consequence of Fubini's theorem and the π-λ theorem, the property above uniquely defines the product measure $\mu_X\otimes\mu_Y$ among measures on $\mathrm{Borel}(X\times Y)$.

Example 4 (Product coupling)**.** Let (X,ω_X,μ_X), $(Y,\omega_Y,\mu_Y)\in\mathcal{N}$. The set $\mathscr{C}(\mu_X,\mu_Y)$ is always nonempty, because the *product measure* $\mu:=\mu_X\otimes\mu_Y$ is always a coupling between μ_X and μ_Y.

Example 5 (1-point coupling)**.** Let X be a set, and let $Y = \{p\}$ be the set with one point. Then for any probability measure μ_X on X there is a unique coupling $\mu = \mu_X \otimes \delta_p$ between μ_X and δ_p. To see this, first we check that μ as defined above is a coupling. Let $A \in \mathrm{Borel}(X)$. Then $\mu(A \times Y) = \mu_X(A)\delta_p(Y) = \mu_X(A)$, and similarly $\mu(X \times \{p\}) = \mu_X(X)\delta_p(\{p\}) = \delta_p(\{p\})$. Thus $\mu \in \mathscr{C}(X, Y)$. For uniqueness, let ν be another coupling. It suffices to show that ν agrees with μ on the measurable rectangles. Let $A \in \mathrm{Borel}(X)$, and observe that

$$\nu(A \times \{p\}) = (\pi_X)_*\nu(A) = \mu_X(A) = \mu_X(A)\delta_p(\{p\}) = \mu(A \times \{p\}).$$

On the other hand, $\nu(A \times \varnothing) \leq \nu(X \times \varnothing) = (\pi_Y)_*\nu(\varnothing) = 0 = \mu_X(A)\delta_p(\varnothing) = \mu(A \times \varnothing)$.

Thus ν satisfies the property $\nu(A \times B) = \mu_X(A)\delta_p(B)$. Thus by uniqueness of the product measure, $\nu = \mu_X \otimes \delta_p$. Finally, note that we can endow X and Y with weight functions ω_X and ω_Y, thus adapting this example to the case of networks.

Example 6 (Diagonal coupling)**.** Let $(X, \omega_X, \mu_X) \in \mathcal{N}$. The *diagonal coupling* between μ_X and itself is defined by writing

$$\Delta(A \times B) := \int_{X \times X} \mathbb{1}_{A \times B}(x, x')\, d\mu_X(x)\, d\delta_x(x') \qquad \text{for all } A \in \mathrm{Borel}(X), B \in \mathrm{Borel}(Y).$$

To see that this is a coupling, let $A \in \mathrm{Borel}(X)$. Then,

$$\Delta(A \times X) = \int_{X \times X} \mathbb{1}_{A \times X}(x, x')\, d\mu_X(x)\, d\delta_x(x') = \int_X \mathbb{1}_A(x)\, d\mu_X(x) = \mu_X(A),$$

and similarly $\Delta(X \times A) = \mu_X(A)$. Thus $\Delta \in \mathscr{C}(\mu_X, \mu_X)$.

Now we turn to the notion of the *distortion* of a coupling. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ be two measure networks. For convenience, we define the function

$$\Omega_{X,Y} : (X \times Y)^2 \to \mathbb{R} \text{ by writing } (x, y, x', y') \mapsto \omega_X(x, x') - \omega_Y(y, y').$$

Next let $\mu \in \mathscr{C}(\mu_X, \mu_Y)$, and consider the probability space $(X \times Y)^2$ equipped with the product measure $\mu \otimes \mu$. For each $p \in [1, \infty)$ the p-distortion of μ is defined as:

$$\begin{aligned}\mathrm{dis}_p(\mu) &= \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p \, d\mu(x, y)\, d\mu(x', y') \right)^{1/p} \\ &= \|\Omega_{X,Y}\|_{L^p(\mu \otimes \mu)}.\end{aligned}$$

For $p = \infty$, the p-distortion is defined as:

$$\mathrm{dis}_p(\mu) := \sup\{|\omega_X(x, x') - \omega_Y(y, y')| : (x, y), (x', y') \in \mathrm{supp}(\mu)\}.$$

When the context is clear, we will often write $\|f\|_p$ to denote $\|f\|_{L^p(\mu \otimes \mu)}$.

2.2. **Interval representation and continuity of distortion.** We now record some standard results about Polish spaces (see also [Stu12, §1.3]). Recall that for a measure space $(X, \mathcal{F}, \mu)$, an *atom* is an element $A \in \mathcal{F}$ such that $0 < \mu(A) < \infty$ and for every $B \in \mathcal{F}$ such that $B \subseteq A$, we have $\mu(B) = 0$ or $\mu(B) = \mu(A)$. In our network setting, the atoms are singletons. To see this, let $(X, \omega_X, \mu_X) \in \mathcal{N}$. The underlying measurable space consists of the Polish space X and its Borel σ-field. Because the topology on X is just the metric topology for a suitable metric, we can use standard techniques involving intersections of elements of covers to show that any atom is necessarily a singleton. Next, since μ_X is a finite measure, $\mathrm{Borel}(X)$ can have at most countably many atoms. In particular, μ_X can be decomposed as the sum of a countable number of atomic (Dirac) measures and a nonatomic measure [Joh70]:

$$\mu_X = \sum_{i=1}^{\infty} c_i \delta_{x_i} + \mu_X', \qquad x_i \in X,\ c_i \in [0, 1] \text{ for each } i \in \mathbb{N}.$$

In what follows, we follow the presentation in [Stu12]. Since X is Polish, it can be viewed as a *standard Borel space* [Sri08] and therefore as the pushforward of Lebesgue measure on the unit interval I. More specifically, let $C_0 = 0$, write $C_i = \sum_{j=1}^{i} c_i$ for $i \in \mathbb{N} \cup \{\infty\}$, $I' = [C_\infty, 1]$, and $X' = \mathrm{supp}(\mu'_X)$. Now X' is a standard Borel space equipped with a nonatomic measure, so by [Sri08, Theorem 3.4.23], there is a Borel isomorphism $\rho' : I' \to X'$ such that $\mu'_X = \rho'_* \lambda_{I'}$, where $\lambda_{I'}$ denotes Lebesgue measure restricted to I'. Define the representation map $\rho : I \to X$ as follows:

$$\rho([C_{i-1}, C_i)) := \{x_i\} \text{ for all } i \in \mathbb{N}, \qquad \rho|_{[C_\infty,1]} := \rho'.$$

The map ρ' is not necessarily unique, and therefore neither is ρ. Any such map ρ is called a *parametrization* of X. In particular, we have $\mu_X = \rho_* \lambda_I$.

The benefit of this construction is that it allows us to represent the underlying measurable space of a network via the unit interval I. Moreover, by taking the pullback of ω_X via ρ, we obtain a network $(I, \rho^*\omega_X, \lambda_I)$. As we will see in the next section, this permits the strategy of proving results over I and transporting them back to X using ρ.

Remark 7 (A 0-distortion coupling between a space and its interval representation)**.** Let $(X, \omega_X, \mu_X) \in \mathcal{N}$, and let $(I, \rho^*\omega_X, \lambda_I)$ be an interval representation of X for some parametrization ρ. Consider the map $(\rho, \mathrm{id}) : I \to X \times I$ given by $i \mapsto (\rho(i), i)$. Define $\mu := (\rho, \mathrm{id})_* \lambda_I$. Let $A \in \mathrm{Borel}(X)$ and $B \in \mathrm{Borel}(I)$. Then $\mu(A \times I) = \lambda_I(\{j \in I : \rho(j) \in A\}) = \mu_X(A)$. Also, $\mu(X \times B) = \lambda_I(\{j \in B : \rho(j) \in X\}) = \lambda_I(B)$. Thus μ is a coupling between μ_X and λ_I. Moreover, for any $A \in \mathrm{Borel}(X)$ and any $B \in \mathrm{Borel}(I)$, if for each $j \in B$ we have $\rho(j) \notin A$, then we have $\mu(A \times B) = 0$. In particular, $\mu(A \times B) = \mu\left((A \cap \rho(B)) \times B\right)$. Also, given $(x, i) \in X \times I$, we have that $\rho(i) \neq x$ implies $(x, i) \notin \mathrm{supp}(\mu)$.

Let $1 \leq p < \infty$. For convenience, define $\omega_I := \rho^*\omega_X$. An explicit computation of $\mathrm{dis}_p(\mu)$ shows:

$$\begin{aligned}
\mathrm{dis}_p(\mu)^p &= \int_{X \times I} \int_{X \times I} \left|\omega_X(x, x') - \omega_I(i, i')\right|^p \, d\mu(x, i)\, d\mu(x', i') \\
&= \int_I \int_I \left|\omega_X(\rho(i), \rho(i')) - \omega_I(i, i')\right|^p \, d\lambda_I(i)\, d\lambda_I(i') \\
&= 0.
\end{aligned}$$

For $p = \infty$, we have:

$$\begin{aligned}
\mathrm{dis}_\infty(\mu) &= \sup\{|\omega_X(x, x') - \omega_I(i, i')| : (x, i), (x', i') \in \mathrm{supp}(\mu)\} \\
&= \sup\{|\omega_X(\rho(i), \rho(i')) - \omega_I(i, i')| : i, i' \in \mathrm{supp}(\lambda)\} \\
&= 0.
\end{aligned}$$

2.3. **Optimality of couplings in the network setting.** We now collect some results about probability spaces. Let X be a Polish space. A subset $P \subseteq \mathrm{Prob}(X)$ is said to be *tight* if for all $\varepsilon > 0$, there is a compact subset $K_\varepsilon \subseteq X$ such that $\mu_X(X \setminus K_\varepsilon) \leq \varepsilon$ for all $\mu_X \in P$.

A sequence $(\mu_n)_{n \in \mathbb{N}} \in \mathrm{Prob}(X)^{\mathbb{N}}$ is said to *converge narrowly* to $\mu_X \in \mathrm{Prob}(X)$ if

$$\lim_{n \to \infty} \int_X f \, d\mu_n = \int_X f \, d\mu_X \qquad \text{for all } f \in C_b(X),$$

the space of continuous, bounded, real-valued functions on X. Narrow convergence is induced by a distance [AGS08, Remark 5.1.1], hence the convergent sequences in $\mathrm{Prob}(X)$ completely determine a topology on $\mathrm{Prob}(X)$. This topology on $\mathrm{Prob}(X)$ is called the *narrow topology*. In some references [Stu12], narrow convergence (resp. narrow topology) is called *weak convergence* (resp. *weak topology*).

A further consequence of having a metric on $\mathrm{Prob}(X)$ [AGS08, Remark 5.1.1] is that singletons are closed. This simple fact will be used below.

Theorem 8 (Prokhorov, [AGS08] Theorem 5.1.3)**.** *Let X be a Polish space. Then $P \subseteq \mathrm{Prob}(X)$ is tight if and only if it is relatively compact, i.e. its closure is compact in $\mathrm{Prob}(X)$.*

Lemma 9 (Lemma 4.4, [Vil08])**.** *Let X, Y be two Polish spaces, and let $P_X \subseteq \mathrm{Prob}(X)$, $P_Y \subseteq \mathrm{Prob}(Y)$ be tight in their respective spaces. Then the set $\mathscr{C}(P_X, P_Y) \subseteq \mathrm{Prob}(X \times Y)$ of couplings with marginals in P_X and P_Y is tight in $\mathrm{Prob}(X \times Y)$.*

Lemma 10 (Compactness of couplings; Lemma 1.2, [Stu12]). *Let X, Y be two Polish spaces. Let $\mu_X \in \mathrm{Prob}(X)$, $\mu_Y \in \mathrm{Prob}(Y)$. Then $\mathscr{C}(X,Y)$ is compact in $\mathrm{Prob}(X \times Y)$.*

Proof. The singletons $\{\mu_X\}$, $\{\mu_Y\}$ are closed and of course compact in $\mathrm{Prob}(X)$, $\mathrm{Prob}(Y)$. Hence by Prokhorov's theorem, they are tight. Now consider $\mathscr{C}(X,Y) \subseteq \mathrm{Prob}(X \times Y)$. Since this is obtained by intersecting the preimages of the continuous projections onto the marginals μ_X and μ_Y, we know that it is closed. Furthermore, $\mathscr{C}(X,Y)$ is tight by Lemma 9. Then by another application of Prokhorov's theorem, it is compact. □

The following lemma appeared for the L^2 case in [Stu12].

Lemma 11 (Continuity of the distortion functional on intervals). *Let $1 \le p < \infty$, and let (I, σ_X, λ_I), $(I, \sigma_Y, \lambda_I) \in \mathcal{N}$. The distortion functional dis_p is continuous on $\mathscr{C}(\lambda_I, \lambda_I) \subseteq \mathrm{Prob}(I \times I)$. For $p = \infty$, dis_∞ is lower semicontinuous.*

Proof. First suppose $p \in [1, \infty)$. We will construct a sequence of continuous functionals that converges uniformly to dis_p. Since the uniform limit of continuous functions is continuous, this will show that dis_p is continuous.

Continuous functions are dense in $L^p(\lambda_I^{\otimes 2})$ (see e.g. [Bog07b]), so for each $n \in \mathbb{N}$, we pick $\sigma_X^n, \sigma_Y^n \in L^p(\lambda_I^{\otimes 2})$ such that

$$\|\sigma_X - \sigma_X^n\|_{L^p(\lambda_I^{\otimes 2})} \le 1/n, \qquad \|\sigma_Y - \sigma_Y^n\|_{L^p(\lambda_I^{\otimes 2})} \le 1/n.$$

For each $n \in \mathbb{N}$, define the functional $\mathrm{dis}_p^n : \mathscr{C}(\lambda_I, \lambda_I) \to \mathbb{R}_+$ as follows:

$$\mathrm{dis}_p^n(\nu) := \left(\int_{I \times I} \int_{I \times I} \left|\sigma_X^n(i,i') - \sigma_Y^n(j,j')\right|^p d\nu(i,j) d\nu(i',j') \right)^{1/p}.$$

Because $|\sigma_X^n - \sigma_Y^n|^p$ is continuous and hence bounded on the compact cube $I^2 \times I^2$, we know that $|\sigma_X^n - \sigma_Y^n|^p \in C_b(I^2 \times I^2)$.

We claim that dis_p^n is continuous. Since the narrow topology on $\mathrm{Prob}(I \times I)$ is induced by a distance [AGS08, Remark 5.1.1], it suffices to show sequential continuity. Let $\nu \in \mathscr{C}(\lambda_I, \lambda_I)$, and let $(\nu_m)_{m \in \mathbb{N}}$ be a sequence in $\mathscr{C}(\lambda_I, \lambda_I)$ converging narrowly to ν. Then we have

$$\begin{aligned}
\lim_{m \to \infty} \mathrm{dis}_p^n(\nu_m) &= \lim_{m \to \infty} \left(\int_{I^2 \times I^2} |\sigma_X^n - \sigma_Y^n|^p \, d\nu_m \otimes d\nu_m \right)^{1/p} \\
&= \left(\int_{I^2 \times I^2} |\sigma_X^n - \sigma_Y^n|^p \, d\nu \otimes d\nu \right)^{1/p} \\
&= \mathrm{dis}_p^n(\nu).
\end{aligned}$$

Here the second equality follows from the definition of convergence in the narrow topology and the fact that the integrand is bounded and continuous. This shows sequential continuity (hence continuity) of dis_p^n.

Finally, we show that $(\mathrm{dis}_p^n)_{n \in \mathbb{N}}$ converges to dis_p uniformly. Let $\mu \in \mathscr{C}(\lambda_I, \lambda_I)$. Then,

$$\begin{aligned}
\left|\mathrm{dis}_p(\mu) - \mathrm{dis}_p^n(\mu)\right| &= \left| \|\sigma_X - \sigma_Y\|_{L^p(\mu^{\otimes 2})} - \|\sigma_X^n - \sigma_Y^n\|_{L^p(\mu^{\otimes 2})} \right| \\
&\le \|\sigma_X - \sigma_Y - \sigma_X^n + \sigma_Y^n\|_{L^p(\mu^{\otimes 2})} \\
&\le \|\sigma_X - \sigma_X^n\|_{L^p(\mu^{\otimes 2})} + \|\sigma_Y - \sigma_Y^n\|_{L^p(\mu^{\otimes 2})} \\
&= \left(\int_{I \times I} \int_{I \times I} \left|\sigma_X(i,i') - \sigma_X^n(i,i')\right|^p \, d\mu(i,j) \, d\mu(i',j') \right)^{1/p} \dots \\
&\quad + \left(\int_{I \times I} \int_{I \times I} \left|\sigma_Y(j,j') - \sigma_Y^n(j,j')\right|^p \, d\mu(i,j) \, d\mu(i',j') \right)^{1/p}
\end{aligned}$$

$$= \left(\int_I \int_I \left| \sigma_X(i,i') - \sigma_X^n(i,i') \right|^p \, d\lambda_I(i) \, d\lambda_I(i') \right)^{1/p} \dots$$
$$+ \left(\int_I \int_I \left| \sigma_Y(j,j') - \sigma_Y^n(j,j') \right|^p \, d\lambda_I(j) \, d\lambda_I(j') \right)^{1/p}$$
$$= \|\sigma_X - \sigma_X^n\|_{L^p(\lambda_I^{\otimes 2})} + \|\sigma_Y - \sigma_Y^n\|_{L^p(\lambda_I^{\otimes 2})}$$
$$\leq 2/n.$$

But $\mu \in \mathscr{C}(\lambda_I, \lambda_I)$ was arbitrary. This shows that dis_p is the uniform limit of continuous functions, hence is continuous. Here the first and second inequalities followed from Minkowski's inequality.

Now suppose $p = \infty$. Let $\mu \in \mathscr{C}(\lambda_I, \lambda_I)$ be arbitrary. Recall that because we are working over probability spaces, Jensen's inequality can be used to show that for any $1 \leq q \leq r < \infty$, we have $\mathrm{dis}_q(\mu) \leq \mathrm{dis}_r(\mu)$. Moreover, we have $\lim_{q\to\infty} \mathrm{dis}_q(\mu) = \mathrm{dis}_\infty(\mu)$. The supremum of a family of continuous functions is lower semicontinuous. In our case, $\mathrm{dis}_\infty = \sup\{\mathrm{dis}_q : q \in [1,\infty)\}$, and we have shown above that all the functions in this family are continuous. Hence dis_∞ is lower semicontinuous. □

The next lemma is standard.

Lemma 12 (Gluing lemma, Lemma 1.4 in [Stu12], also Lemma 7.6 in [Vil03])**.** *Let $\mu_1, \mu_2, \dots, \mu_k$ be probability measures supported on Polish spaces $X_1, \dots, X_k$. For each $i \in \{1, \dots, k-1\}$, let $\mu_{i,i+1} \in \mathscr{C}(\mu_i, \mu_{i+1})$. Then there exists $\mu \in \mathrm{Prob}(X_1 \times X_2 \times \dots \times X_k)$ with marginals $\mu_{i,i+1}$ on $X_i \times X_{i+1}$ for each $i \in \{1, \dots, k-1\}$.*

2.4. **The Network Gromov-Wasserstein distance.** For each $p \in [1,\infty]$, we define:

$$d_{\mathcal{N},p}(X,Y) := \frac{1}{2} \inf_{\mu \in \mathscr{C}(\mu_X,\mu_Y)} \mathrm{dis}_p(\mu) \qquad \text{for each } (X,\omega_X,\mu_X), (Y,\omega_Y,\mu_Y) \in \mathcal{N}.$$

As we will see below, $d_{\mathcal{N},p}$ is a legitimate pseudometric on $\mathcal{N}$. The structure of $d_{\mathcal{N},p}$ is analogous to a formulation of the *Gromov-Wasserstein distance* between metric measure spaces [Mém11, Stu12].

Remark 13 (Boundedness of $d_{\mathcal{N},p}$)**.** Recall from Example 4 that for any $X, Y \in \mathcal{N}$, $\mathscr{C}(\mu_X, \mu_Y)$ always contains the product coupling, and is thus nonempty. A consequence is that $d_{\mathcal{N},p}(X,Y)$ is bounded for any $p \in [1,\infty]$. Indeed, by taking the product coupling $\mu := \mu_X \otimes \mu_Y$ we have

$$d_{\mathcal{N},p}(X,Y) \leq \frac{1}{2} \mathrm{dis}_p(\mu).$$

Suppose first that $p \in [1,\infty)$. Applying Minkowski's inequality, we obtain:

$$\begin{aligned}
\mathrm{dis}_p(\mu) &= \|\omega_X - \omega_Y\|_{L^p(\mu\otimes\mu)} \\
&\leq \|\omega_X\|_{L^p(\mu\otimes\mu)} + \|\omega_Y\|_{L^p(\mu\otimes\mu)} \\
&= \left(\int_{X\times Y} \int_{X\times Y} \left| \omega_X(x,x') \right|^p \, d\mu(x,y) \, d\mu(x',y') \right)^{1/p} \\
&\qquad + \left(\int_{X\times Y} \int_{X\times Y} \left| \omega_Y(y,y') \right|^p \, d\mu(x,y) \, d\mu(x',y') \right)^{1/p} \\
&= \left(\int_X \int_X \left| \omega_X(x,x') \right|^p \, d\mu_X(x) \, d\mu_X(x') \right)^{1/p} + \left(\int_Y \int_Y \left| \omega_Y(y,y') \right|^p \, d\mu_Y(y) \, d\mu_Y(y') \right)^{1/p} \\
&= \|\omega_X\|_{L^p(\mu_X\otimes\mu_X)} + \|\omega_Y\|_{L^p(\mu_Y\otimes\mu_Y)} < \infty.
\end{aligned}$$

The case $p = \infty$ case is analogous, except that integrals are replaced by taking essential suprema as needed.

In some simple cases, we obtain explicit formulas for computing $d_{\mathcal{N},p}$.

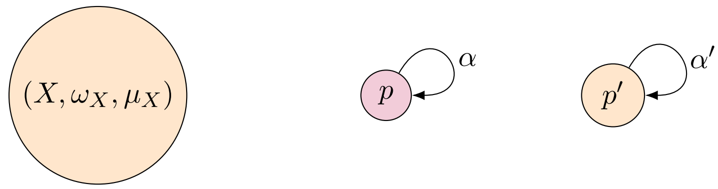


FIGURE 1. The $d_{\mathcal{N},p}$ distance between the two one-node networks is simply $\frac{1}{2}|\alpha - \alpha'|$. In Example 14 we give an explicit formula for computing $d_{\mathcal{N},p}$ between an arbitrary network and a one-node network.

Example 14 (Easy examples of $d_{\mathcal{N},p}$)**.** Let $a, b \in \mathbb{R}$ and consider the networks $N_1(a)$ and $N_1(b)$. The unique coupling between the two networks is the product measure $\mu = \delta_x \otimes \delta_y$, where we understand x, y to be the nodes of the two networks. Then for any $p \in [1, \infty]$, we obtain:

$$\mathrm{dis}_p(\mu) = |\omega_{N_1(a)}(x, x) - \omega_{N_1(b)}(y, y)| = |a - b|.$$

Thus $d_{\mathcal{N},p}(N_1(a), N_1(b)) = \frac{1}{2}|a - b|$.

Let $(X, \omega_X, \mu_X) \in \mathcal{N}$ be any network and let $N_1(a) = (\{y\}, a)$ be a network with one node. Once again, there is a unique coupling $\mu = \mu_X \otimes \delta_y$ between the two networks. For any $p \in [1, \infty)$, we obtain:

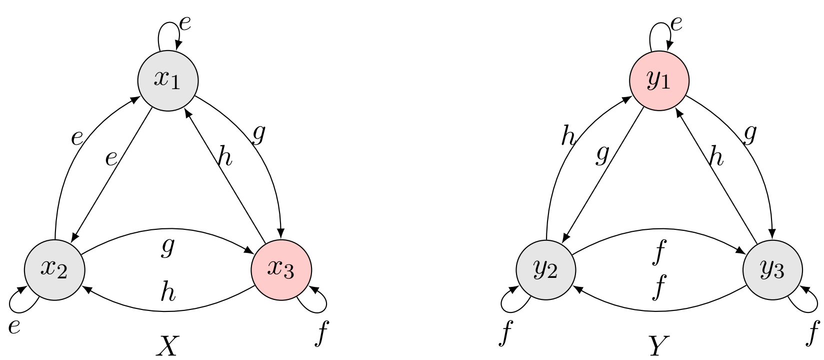
$$d_{\mathcal{N},p}(X, N_1(a)) = \frac{1}{2}\,\mathrm{dis}_p(\mu) = \frac{1}{2}\left(\int_X \int_X \left|\omega_X(x, x') - a\right|^p d\mu_X(x) d\mu_X(x')\right)^{1/p}.$$

For $p = \infty$, we have $d_{\mathcal{N},p}(X, \mathbb{N}_1(a)) = \sup\{\frac{1}{2}|\omega_X(x, x') - a| : x, x' \in \mathrm{supp}(\mu_X)\}$.

Remark 15. $d_{\mathcal{N},p}$ is not necessarily a metric modulo strong isomorphism. Let $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3\}$. Consider a coupling μ given as:

$$\mu = \begin{matrix} \\ x_1 \\ x_2 \\ x_3 \end{matrix} \begin{matrix} \begin{matrix} y_1 & y_2 & y_3 \end{matrix} \\ \begin{pmatrix} 1/3 & 0 & 0 \\ 1/6 & 0 & 0 \\ 0 & 1/6 & 1/3 \end{pmatrix} \end{matrix}.$$

Next equip X and Y with edge weights $\{e, f, g, h\}$ as in the following figure:



Comparing the edge weights, it is clear that X and Y are not strongly isomorphic. However, $d_{\mathcal{N},p}(X, Y) = 0$ for all $p \in [1, \infty]$. To see this, define:

$$G = \{(x_1, y_1), (x_2, y_1), (x_3, y_2), (x_3, y_3)\}$$

Then G contains all the points with positive μ-measure. Given any two points $(x, y), (x', y') \in G$, we observe that $|\omega_X(x, x') - \omega_Y(y, y')| = 0$. Thus for any $p \in [1, \infty]$, $\mathrm{dis}_p(\mu) = 0$, and so $d_{\mathcal{N},p}(X, Y) = 0$.

The definition of $d_{\mathcal{N},p}$ is sensible in the sense that it captures the notion of a distance:

Theorem 16. *For each $p \in [1,\infty]$, $d_{\mathcal{N},p}$ is a pseudometric on $\mathcal{N}$.*

Proof of Theorem 16. Let $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y),(Z,\omega_Y,\mu_Y) \in \mathcal{N}$. It is clear that $d_{\mathcal{N},p}(X,Y) \geq 0$. To show $d_{\mathcal{N},p}(X,X) = 0$, consider the diagonal coupling Δ (see Example 6). For $p \in [1,\infty)$, we have:

$$\begin{aligned}
\operatorname{dis}_p(\Delta) &= \left(\int_{X\times X}\int_{X\times X} \left|\omega_X(x,x') - \omega_X(z,z')\right|^p d\Delta(x,z) d\Delta(x',z')\right)^{1/p}\\
&= \left(\int_X \int_X \left|\omega_X(x,x') - \omega_X(x,x')\right|^p d\mu_X(x) d\mu_X(x')\right)^{1/p}\\
&= 0.
\end{aligned}$$

For $p=\infty$, we have:

$$\begin{aligned}
\operatorname{dis}_p(\Delta) &= \sup\{\left|\omega_X(x,x') - \omega_X(z,z')\right| : (x,z),(x',z') \in \operatorname{supp}(\Delta)\}\\
&= \sup\{\left|\omega_X(x,x') - \omega_X(x,x')\right| : x,x' \in \operatorname{supp}(\mu_X)\}\\
&= 0.
\end{aligned}$$

Thus $d_{\mathcal{N},p}(X,X) = 0$ for any $p \in [1,\infty]$. For symmetry, notice that for any $\mu \in \mathscr{C}(\mu_X,\mu_Y)$, we can define $\widetilde{\mu} \in \mathscr{C}(\mu_Y,\mu_X)$ by $\widetilde{\mu}(y,x) = \mu(x,y)$. Then $\operatorname{dis}_p(\mu) = \operatorname{dis}_p(\widetilde{\mu})$, and this will show $d_{\mathcal{N},p}(X,Y) = d_{\mathcal{N},p}(Y,X)$.

Finally, we need to check the triangle inequality. Let $\varepsilon > 0$, and let $\mu_{12} \in \mathscr{C}(\mu_X,\mu_Y)$ and $\mu_{23} \in \mathscr{C}(\mu_Y,\mu_Z)$ be couplings such that $2d_{\mathcal{N},p}(X,Y) \geq \operatorname{dis}_p(\mu_{12}) - \varepsilon$ and $2d_{\mathcal{N},p}(Y,Z) \geq \operatorname{dis}_p(\mu_{23}) - \varepsilon$. Invoking Lemma 12, we obtain a probability measure $\mu \in \operatorname{Prob}(X\times Y\times Z)$ with marginals μ_{12},μ_{23}, and a marginal μ_{13} that is a coupling between μ_X and μ_Z. This coupling is not necessarily optimal. For $p \in [1,\infty)$ we have:

$$\begin{aligned}
2d_{\mathcal{N},p}(X,Z) &\leq \operatorname{dis}_p(\mu_{13})\\
&= \left(\int_{X\times Z}\int_{X\times Z} \left|\omega_X(x,x') - \omega_Z(z,z')\right|^p \, d\mu_{13}(x,z)\, d\mu_{13}(x',z')\right)^{1/p}\\
&= \left(\int_{X\times Y\times Z}\int_{X\times Y\times Z} \left|\omega_X(x,x') - \omega_Z(z,z')\right|^p \, d\mu(x,y,z)\, d\mu(x',y',z')\right)^{1/p}\\
&= \|\omega_X - \omega_Y + \omega_Y - \omega_Z\|_{L^p(\mu\otimes\mu)}\\
&\leq \|\omega_X - \omega_Y\|_{L^p(\mu\otimes\mu)} + \|\omega_Y - \omega_Z\|_{L^p(\mu\otimes\mu)}\\
&= \left(\int_{X\times Y}\int_{X\times Y} \left|\omega_X(x,x') - \omega_Y(y,y')\right|^p \, d\mu_{12}(x,y)\, d\mu_{12}(x',y')\right)^{1/p} \ldots\\
&\quad + \left(\int_{Y\times Z}\int_{Y\times Z} \left|\omega_Y(y,y') - \omega_Z(z,z')\right|^p \, d\mu_{23}(y,z)\, d\mu_{23}(y',z')\right)^{1/p}\\
&\leq 2d_{\mathcal{N},p}(X,Y) + 2d_{\mathcal{N},p}(Y,Z) + 2\varepsilon.
\end{aligned}$$

The second inequality above follows from Minkowski's inequality. Letting $\varepsilon \to 0$ now proves the triangle inequality in the case $p \in [1,\infty)$.

For $p = \infty$ we have:

$$\begin{aligned}
2d_{\mathcal{N},p}(X,Z) &\leq \operatorname{dis}_p(\mu_{13})\\
&= \sup\{\left|\omega_X(x,x') - \omega_Z(z,z')\right| : (x,z),(x',z') \in \operatorname{supp}(\mu_{13})\}\\
&= \sup\{\left|\omega_X(x,x') - \omega_Y(y,y') + \omega_Y(y,y') - \omega_Z(z,z')\right| : (x,y,z),(x',y',z') \in \operatorname{supp}(\mu)\}\\
&\leq \sup\{\left|\omega_X(x,x') - \omega_Y(y,y')\right| + \left|\omega_Y(y,y') - \omega_Z(z,z')\right|
\end{aligned}$$

$$: (x,y),(x',y') \in \operatorname{supp}(\mu_{12}),(y,z),(y',z') \in \operatorname{supp}(\mu_{23})\}$$
$$\leq \operatorname{dis}_p(\mu_{12}) + \operatorname{dis}_p(\mu_{23})$$
$$\leq 2d_{\mathcal{N},p}(X,Y) + 2d_{\mathcal{N},p}(Y,Z) + 2\varepsilon.$$

Letting $\varepsilon \to 0$ now proves the triangle inequality in the case $p = \infty$. This concludes our proof. □

By the next result, this infimum is actually attained. Hence we may write:

$$d_{\mathcal{N},p}(X,Y) := \frac{1}{2} \min_{\mu \in \mathscr{C}(X,Y)} \operatorname{dis}_p(\mu).$$

Definition 2 (Optimal couplings)**.** Let $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y) \in \mathcal{N}$, and let $p \in [1,\infty)$. A coupling $\mu \in \mathscr{C}(\mu_X,\mu_Y)$ is *optimal* if $\operatorname{dis}_p(\mu) = 2d_{\mathcal{N},p}(X,Y)$.

Theorem 17. *Let (X,ω_X,μ_X) and (Y,ω_Y,μ_Y) be two measure networks, and let $p \in [1,\infty]$. Then there exists an optimal coupling, i.e. a minimizer for $\operatorname{dis}_p(\cdot)$ in $\mathscr{C}(\mu_X,\mu_Y)$.*

Proof. First suppose $p \in [1,\infty)$. By the construction in Section 2.2, we pass into interval representations of X and Y. As noted in Section 2.2, the choice of parametrization is not necessarily unique, but this does not affect the argument. Let (I,σ_X,λ_I), (I,σ_Y,λ_I) denote these representations. By Lemma 11, the dis_p functional is continuous on the space of couplings between these two networks. By Lemma 10, this space of couplings is compact. Thus dis_p achieves its infimum.

Let $\mu \in \mathscr{C}(\lambda_I,\lambda_I)$ denote this minimizer of dis_p. By Remark 7, we can also take couplings $\mu_X \in \mathscr{C}(\mu_X,\lambda_I)$ and $\mu_Y \in \mathscr{C}(\lambda_I,\mu_Y)$ which have zero distortion. By Lemma 12, we can glue together μ_X,μ, and μ_Y to obtain a coupling $\nu \in \mathscr{C}(\mu_X,\mu_Y)$. By the proof of the triangle inequality in Theorem 16, we have:

$$\operatorname{dis}_p(\nu) \leq \operatorname{dis}_p(\mu_X) + \operatorname{dis}_p(\mu) + \operatorname{dis}_p(\mu_Y) = \operatorname{dis}_p(\mu) = 2d_{\mathcal{N},p}((I,\sigma_X,\lambda_I),(I,\sigma_Y,\lambda_I)).$$

Also by the triangle inequality, we have $d_{\mathcal{N},p}((I,\sigma_X,\lambda_I),(I,\sigma_Y,\lambda_I)) \leq d_{\mathcal{N},p}(X,Y)$. It follows that $\nu \in \mathscr{C}(\mu_X,\mu_Y)$ is optimal.

The case $p = \infty$ is analogous, because lower semicontinuity (Lemma 11) combined with compactness (Lemma 10) is sufficient to guarantee that $\operatorname{dis}_\infty$ achieves its infimum on $\mathscr{C}(\lambda_I,\lambda_I)$. □

It remains to discuss the precise pseudometric structure of $d_{\mathcal{N},p}$. The following definition is a relaxation of strong isomorphism.

Definition 3 (Weak isomorphism)**.** $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y) \in \mathcal{N}$ are *weakly isomorphic*, denoted $X \cong^w Y$, if there exists a Borel probability space (Z,μ_Z) with measurable maps $f : Z \to X$ and $g : Z \to Y$ such that

- $f_*\mu_Z = \mu_X$, $g_*\mu_Z = \mu_Y$, and
- $\|f^*\omega_X - g^*\omega_Y\|_\infty = 0$.

Here $f^*\omega_X : Z \times Z \to \mathbb{R}$ is the pullback weight function given by the map $(z,z') \mapsto \omega_X(f(z),f(z'))$. The map $g^*\omega_Y$ is defined analogously. For the definition to make sense, we need to check that $f^*\omega_X$ is measurable. Let $(a,b) \in \operatorname{Borel}(\mathbb{R})$. Then $B := \{\omega_X \in (a,b)\}$ is measurable because ω_X is measurable. Because f is measurable, we know that $(f,f) : Z \times Z \to X \times X$ is measurable. Thus $A := (f,f)^{-1}(B)$ is measurable. Now we write:

$$\begin{aligned} A &= \{(z,z') \in Z^2 : \big((f(z),f(z')\big) \in B\} \\ &= \{(z,z') \in Z^2 : \omega_X(f(z),f(z')) \in (a,b)\} \\ &= (f^*\omega_X)^{-1}(a,b). \end{aligned}$$

Thus $f^*\omega_X$ is measurable. Similarly, we verify that $g^*\omega_Y$ is measurable.

Theorem 18 (Pseudometric structure of $d_{\mathcal{N},p}$)**.** *Let $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y) \in \mathcal{N}$, and let $p \in [1,\infty]$. Then $d_{\mathcal{N},p}(X,Y) = 0$ if and only if $X \cong^w Y$.*

Proof of Theorem 18. Fix $p \in [1, \infty)$. For the backwards direction, suppose there exist Z and measurable maps $f : Z \to X$ and $g : Z \to Y$ such that the appropriate conditions are satisfied. We first claim that $d_{\mathcal{N},p}((X, \omega_X, \mu_X), (Z, f^*\omega_X, \mu_Z)) = 0$.

To see the claim, define $\mu \in \operatorname{Prob}(X \times Z)$ by $\mu := (f, \mathrm{id})_*\mu_Z$. Then,

$$\int_{X\times Z}\int_{X\times Z} \left|\omega_X(x,x') - f^*\omega_X(z,z')\right|^p \, d\mu(x,z)\, d\mu(x',z')$$
$$\int_{X\times Z}\int_{X\times Z} \left|\omega_X(x,x') - \omega_X(f(z),f(z'))\right|^p \, d\mu(x,z)\, d\mu(x',z')$$
$$\int_{Z}\int_{Z} \left|\omega_X(f(z),f(z')) - \omega_X(f(z),f(z'))\right|^p \, d\mu_Z(z)\, d\mu_Z(z') = 0.$$

This verifies the claim. Similarly we have $d_{\mathcal{N},p}((Y, \omega_Y, \mu_X), (Z, g^*\omega_Y, \mu_Z)) = 0$. Using the diagonal coupling along with the assumption, we have $d_{\mathcal{N},p}((Z, f^*\omega_X, \mu_Z), (Z, g^*\omega_Y, \mu_Z)) = 0$. By triangle inequality, we then have $d_{\mathcal{N},p}(X, Y) = 0$.

For the forwards direction, let $\mu \in \mathscr{C}(\mu_X, \mu_Y)$ be an optimal coupling with $\operatorname{dis}_p(\mu) = 0$ (Theorem 17). Define $Z := X \times Y$, $\mu_Z := \mu$. Then the projection maps $\pi_X : Z \to X$ and $\pi_Y : Z \to Y$ are measurable. We also have $(\pi_X)_*\mu = \mu_X$ and $(\pi_Y)_*\mu = \mu_Y$. Since $\operatorname{dis}_p(\mu) = 0$, we also have $\|(\pi_X)^*\omega_X - (\pi_Y)^*\omega_Y\|_\infty = \|\omega_X - \omega_Y\|_\infty = 0$.

The $p = \infty$ case is proved analogously. This concludes the proof. □

Remark 19. Theorem 18 is in the same spirit as related results for gauged measure spaces [Stu12] and for networks without measure equipped with a Gromov-Hausdorff-type network distance [CM17]. The "tripod structure" $X \leftarrow Z \rightarrow Y$ described above is much more difficult to obtain in the setting of [CM17]. This highlights an advantage of the measure-theoretic setting of the current paper.

In the next section we follow a brief diversion to study a *Gromov-Prokhorov* distance between measure networks. While it is not the main focus of the current paper, it turns out to be useful for the notion of *interleaving stability* that we define in §3.

2.5. **The Network Gromov-Prokhorov distance.** Let $\alpha \in [0, \infty)$. For any $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$, we write $\mathscr{C} := \mathscr{C}(\mu_X, \mu_Y)$ and define:

$$d^{\mathcal{GP}}_{\mathcal{N},\alpha}(X,Y) := \frac{1}{2}\inf_{\mu\in\mathscr{C}} \inf\{\varepsilon > 0 : \mu\otimes\mu\left(\{x,y,x',y' \in (X\times Y)^2 : \left|\omega_X(x,x') - \omega_Y(y,y')\right| \geq \varepsilon\}\right) \leq \alpha\varepsilon\}.$$

Theorem 20. *For each $\alpha \in [0, \infty)$, $d^{\mathcal{GP}}_{\mathcal{N},\alpha}$ is a pseudometric on $\mathcal{N}$.*

Proof. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y), (Z, \omega_Z, \mu_Z) \in \mathcal{N}$. The proofs that $d^{\mathcal{GP}}_{\mathcal{N},\alpha}(X,Y) \geq 0$, $d^{\mathcal{GP}}_{\mathcal{N},\alpha}(X,X) = 0$, and that $d^{\mathcal{GP}}_{\mathcal{N},\alpha}(X,Y) = d^{\mathcal{GP}}_{\mathcal{N},\alpha}(Y,X)$ are analogous to those used in Theorem 16. Hence we only check the triangle inequality. Let $\varepsilon_{XY} > 2d^{\mathcal{GP}}_{\mathcal{N},\alpha}(X,Y)$, $\varepsilon_{YZ} > 2d^{\mathcal{GP}}_{\mathcal{N},\alpha}(Y,Z)$, and let μ_{XY}, μ_{YZ} be couplings such that

$$\mu_{XY}^{\otimes 2}\left(\{(x,y,x',y') : \left|\omega_X(x,x') - \omega_Y(y,y')\right| \geq \varepsilon_{XY}\}\right) \leq \alpha\varepsilon_{XY},$$
$$\mu_{YZ}^{\otimes 2}\left(\{(y,z,y',z') : \left|\omega_Y(y,y') - \omega_Z(z,z')\right| \geq \varepsilon_{YZ}\}\right) \leq \alpha\varepsilon_{YZ}.$$

For convenience, define:

$$A := \{((x,y,z),(x',y',z')) \in (X\times Y\times Z)^2 : \left|\omega_X(x,x') - \omega_Y(y,y')\right| \geq \varepsilon_{XY}\}$$
$$B := \{((x,y,z),(x',y',z')) \in (X\times Y\times Z)^2 : \left|\omega_Y(y,y') - \omega_Z(z,z')\right| \geq \varepsilon_{YZ}\}$$
$$C := \{((x,y,z),(x',y',z')) \in (X\times Y\times Z)^2 : \left|\omega_X(x,x') - \omega_Z(z,z')\right| \geq \varepsilon_{XY} + \varepsilon_{YZ}\}.$$

Next let μ denote the probability measure obtained from gluing μ_{XY} and μ_{YZ} (cf. Lemma 12). This has marginals μ_{XY}, μ_{YZ}, and a marginal $\mu_{XZ} \in \mathscr{C}(\mu_X, \mu_Z)$. We need to show:

$$\mu_{XZ}^{\otimes 2}((\pi_X, \pi_Z)(C)) \leq \alpha(\varepsilon_{XY} + \varepsilon_{YZ}).$$

To show this, it suffices to show $C \subseteq A \cup B$, because then we have $\mu^{\otimes 2}(C) \leq \mu^{\otimes 2}(A) + \mu^{\otimes 2}(B)$ and consequently

$$\begin{aligned}\mu_{XZ}^{\otimes 2}((\pi_X, \pi_Z)(C)) = \mu^{\otimes 2}(C) \leq \mu^{\otimes 2}(A) + \mu^{\otimes 2}(B) &= \mu_{XY}^{\otimes 2}((\pi_X, \pi_Y)(A)) + \mu_{YZ}^{\otimes 2}((\pi_Y, \pi_Z)(B)) \\ &\leq \alpha(\varepsilon_{XY} + \varepsilon_{YZ}).\end{aligned}$$

Let $((x,y,z),(x',y',z')) \in (X \times Y \times Z)^2 \setminus (A \cup B)$. Then we have

$$\left|\omega_X(x,x') - \omega_Y(y,y')\right| < \varepsilon_{XY} \text{ and } \left|\omega_Y(y,y') - \omega_Z(z,z')\right| < \varepsilon_{YZ}.$$

By the triangle inequality, we then have:

$$\left|\omega_X(x,x') - \omega_Z(z,z')\right| \leq \left|\omega_X(x,x') - \omega_Y(y,y')\right| + \left|\omega_Y(y,y') - \omega_Z(z,z')\right| < \varepsilon_{XY} + \varepsilon_{YZ}.$$

Thus $((x,y,z),(x',y',z')) \in (X \times Y \times Z)^2 \setminus C$. This shows $C \subseteq A \cup B$.

The preceding work shows that $2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X,Z) \leq \varepsilon_{XY} + \varepsilon_{YZ}$. Since $\varepsilon_{XY} > 2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X,Y)$ and $\varepsilon_{YZ} > 2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(Y,Z)$ were arbitrary, it follows that $d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X,Z) \leq d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X,Y) + d_{\mathcal{N},\alpha}^{\mathcal{GP}}(Y,Z)$. □

Lemma 21 (Relation between Gromov-Prokhorov and Gromov-Wasserstein)**.** *Let* $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$. *We always have:*

$$d_{\mathcal{N},0}^{\mathcal{GP}}(X,Y) = d_{\mathcal{N},\infty}(X,Y).$$

Proof. Write $\mathscr{C} := \mathscr{C}(X,Y)$. When $\alpha = 0$ in the $d_{\mathcal{N},\alpha}^{\mathcal{GP}}$ formulation, we have:

$$\begin{aligned}2d_{\mathcal{N},0}^{\mathcal{GP}}(X,Y) &= \inf_{\mu \in \mathscr{C}} \inf\{\varepsilon > 0 : \mu^{\otimes 2}\left(\{x,y,x',y' \in (X \times Y)^2 : \left|\omega_X(x,x') - \omega_Y(y,y')\right| \geq \varepsilon\}\right) = 0\} \\ &= \inf_{\mu \in \mathscr{C}} \sup\{\varepsilon > 0 : \mu^{\otimes 2}\left(\{x,y,x',y' \in (X \times Y)^2 : \left|\omega_X(x,x') - \omega_Y(y,y')\right| < \varepsilon\}\right) = 1\} \\ &= \inf_{\mu \in \mathscr{C}} \sup\{\left|\omega_X(x,x') - \omega_Y(y,y')\right| : (x,y),(x',y') \in \operatorname{supp}(\mu)\} \\ &= 2d_{\mathcal{N},\infty}(X,Y).\end{aligned}$$

□

3. Invariants and lower bounds

Let (V, d_V) denote a pseudometric space. By a *(pseudo)metric-valued network invariant*, we mean a function $\iota : \mathcal{N} \to V$ such that $X \cong Y$ implies $d_V(\iota(X), \iota(Y)) = 0$. We are also interested in $\mathbb{R}$*-parametrized* network invariants, which are functions $\iota : \mathcal{N} \times \mathbb{R} \to V$ such that $X \cong Y$ implies $d_V(\iota(X), \iota(Y)) = 0$. This is a bona fide generalization of the non-parametrized setting, because any map $\iota : \mathcal{N} \to V$ can be viewed as being parametrized by a constant object $\{0\}$.

There are two notions of stability that we are interested in.

Definition 4 (Lipschitz stability)**.** Let $p \in [1, \infty]$. A *Lipschitz-stable network invariant* is an invariant $\iota_p : \mathcal{N} \to V$ for which there exists a Lipschitz constant $L(\iota_p) > 0$ such that

$$d_V(\iota_p(X), \iota_p(Y)) \leq L(\iota_p) d_{\mathcal{N},p}(X,Y) \text{ for all } X, Y \in \mathcal{N}.$$

Definition 5 (Interleaving stability)**.** Let $p \in [1, \infty]$. An *interleaving-stable network invariant* is an $\mathbb{R}$-parametrized invariant $\iota_p : \mathcal{N} \times \mathbb{R} \to V$ for which there exists an interleaving constant $\alpha \in \mathbb{R}$ and a symmetric interleaving function $\varepsilon : \mathcal{N} \times \mathcal{N} \to \mathbb{R}$ such that

$$\iota_p(X,t) \leq \iota_p(Y, t + \varepsilon_{XY}) + \alpha\varepsilon_{XY} \leq \iota_p(X, t + 2\varepsilon_{XY}) + 2\alpha\varepsilon_{XY} \text{ for all } t \in \mathbb{R} \text{ and all } X, Y \in \mathcal{N}.$$

Here $\varepsilon_{XY} := \varepsilon(X,Y)$.

Occasionally, some of our proofs can be compacted by using the following auxiliary definition. Consider the family of set functions $\mathfrak{a} := \{A : \mathcal{N} \to \mathbf{Set} : A(X) \subseteq X \times X \text{ for each } X \in \mathcal{N}\}$.

Definition 6 ((A, p)-indicated invariants)**.** For each $A \in \mathfrak{a}$ and $p \in [1, \infty]$, we define an (A, p)-*indicated invariant* $\iota_{A,p} : \mathcal{N} \to \mathbb{R}$ by writing

$$\iota_{A,p}(X, \omega_X, \mu_X) := \left\|\omega_X \mathbb{1}_{A(X)}\right\|_{L^p(\mu_X \otimes \mu_X)}$$

for each $(X, \omega_X, \mu_X) \in \mathcal{N}$.

3.1. **Global invariants.** The first class of invariants that we consider are those which incorporate data from the entire network at once.

Example 22 (The size_p invariant)**.** Let $p \in [1, \infty]$. The pth size invariant is the map $\text{size}_p : \mathcal{N} \to \mathbb{R}_+$ given by writing, for each $(X, \omega_X, \mu_X) \in \mathcal{N}$,

$$\begin{aligned} \text{size}_p(X) &:= \left(\int_X \int_X \left|\omega_X(x, x')\right|^p \, d\mu_X(x) \, d\mu_X(x') \right)^{1/p} && \text{for } p \in [1, \infty), \\ &:= \sup\{\left|\omega_X(x, x')\right| : x, x' \in \text{supp}(\mu_X)\} && \text{for } p = \infty. \end{aligned}$$

Then size_p is an (A, p)-indicated invariant where A is given by writing $A(X) = X \times X$ for each $(X, \omega_X, \mu_X) \in \mathcal{N}$. The naming convention for this invariant follows [Stu12].

Example 23 (A map that sums the diagonal)**.** The pth trace invariant is the map $\text{tr}_p : \mathcal{N} \to \mathbb{R}_+$ given by taking $A(X) = \text{diag}(X \times X)$ for each $(X, \omega_X, \mu_X) \in \mathcal{N}$. As an example, for $p \in [1, \infty)$ we have:

$$\begin{aligned} \text{tr}_p(X, \omega_X, \mu_X) &= \left(\int_X |w_X(x, x)|^p \, d\mu_X(x) \right)^{1/p} \\ &= \left(\int_X \int_X \left|\omega_X(x, x') \mathbb{1}_{\text{diag}(X \times X)}(x, x')\right|^p \, d\mu_X(x) \, d\mu_X(x') \right)^{1/p}. \end{aligned}$$

Next we present some $\mathbb{R}$-parametrized network invariants.

Example 24 (A map that ignores/emphasizes large edge weights)**.** Let $t \in \mathbb{R}$. For each $p \in [1, \infty]$, the pth t-sublevel set map for the weight function, denoted $\text{sub}^w_{p,t} : \mathcal{N} \to \mathbb{R}_+$, is an (A, p)-indicated invariant obtained by writing $A(X) = \{\omega_X \leq t\}$ for each $(X, \omega_X, \mu_X) \in \mathcal{N}$. This map de-emphasizes large edge weights in a measure network. This map is explicitly given as:

$$\begin{aligned} \text{sub}^w_{p,t}(X, \omega_X, \mu_X) &= \left(\int_{\{\omega_X \leq t\}} \left|\omega_X(x, x')\right|^p \, d(\mu_X \otimes \mu_X)(x, x') \right)^{1/p} && \text{for } p \in [1, \infty), \\ \text{sub}^w_{p,t}(X, \omega_X, \mu_X) &= \sup\{\left|\omega_X(x, x')\right| : x, x' \in \text{supp}(\mu_X), \ \omega_X(x, x') \leq t\} && \text{for } p = \infty. \end{aligned}$$

Analogously, one can consider integrating over the set $\{\omega_X \geq t\}$. In this case, $A(X) = \{\omega_X \geq t\}$ and the larger edge weights are emphasized. The corresponding superlevel set invariant is denoted $\text{sup}^w_{p,t}$.

The invariants we have introduced so far are all examples of *global* invariants. In particular, each of these invariants compresses all the information in a network into a single real number. The next result shows that this compression occurs in a quantitatively stable manner.

Theorem 25 (Lipschitz stability of size and tr invariants)**.** *The* size *and* tr *invariants are quantitatively stable for each* $p \in [1, \infty]$*, with Lipschitz constant* $L = 2$.

Proof. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$, and let $p \in [1, \infty]$. For convenience, define $\varphi_{XY} : (X \times Y)^2 \to \mathbb{R}$ by writing $\varphi_{XY}(x, y, x', y') := \omega_X(x, x')$, and define $\psi_{XY} : (X \times Y)^2 \to \mathbb{R}$ by writing $\psi_{XY}(x, y, x', y') := \omega_Y(y, y')$. Also let $A \subseteq (X \times Y)^2$ be some subset to be determined later.

Let $\eta > d_{\mathcal{N},p}(X,Y)$, and let $\mu \in \mathscr{C}(\mu_X,\mu_Y)$ be a coupling such that $\operatorname{dis}_p(\mu) < 2\eta$. Then by applying Minkowski's inequality, we obtain:

$$\left| \|\varphi_{XY}\mathbb{1}_A\|_{L^p(\mu\otimes\mu)} - \|\psi_{XY}\mathbb{1}_A\|_{L^p(\mu\otimes\mu)} \right| \le \|\varphi_{XY}\mathbb{1}_A - \psi_{XY}\mathbb{1}_A\|_{L^p(\mu\otimes\mu)}. \tag{1}$$

To make the proof more readable, we will now restrict to the case $p \in [1,\infty)$ and write out the integrals fully. However, at each step, we only use properties of couplings and norms, so the same technique works for the $p = \infty$ case.

The right hand side of the preceding inequality is equal to:

$$\left(\int_{(X\times Y)^2} \left| (\omega_X(x,x') - \omega_Y(y,y'))\, \mathbb{1}_A \right|^p \, d\mu(x,y)\, d\mu(x',y') \right)^{1/p}$$
$$\le \left(\int_{(X\times Y)^2} \left| \omega_X(x,x') - \omega_Y(y,y') \right|^p \, d\mu(x,y)\, d\mu(x',y') \right)^{1/p} < 2\eta.$$

Suppose now that $A = (X\times Y)^2$. Then we have:

$$\begin{aligned} \|\varphi_{XY}\mathbb{1}_A\|_{L^p(\mu\otimes\mu)} &= \left(\int_{X\times Y}\int_{X\times Y} \left|\omega_X(x,x')\right|^p \, d\mu(x,y)\, d\mu(x',y') \right)^{1/p} \\ &= \left(\int_X \int_X \left|\omega_X(x,x')\right|^p \, d\mu_X(x)\, d\mu_X(x') \right)^{1/p} = \operatorname{size}_p(X). \end{aligned}$$

Similarly, $\|\psi_{XY}\mathbb{1}_A\|_{L^p(\mu\otimes\mu)} = \operatorname{size}_p(Y)$. Hence $|\operatorname{size}_p(X) - \operatorname{size}_p(Y)| < 2\eta$.

Next suppose $A = \operatorname{diag}((X\times Y)\times(X\times Y))$. Then we have:

$$\begin{aligned} \|\varphi_{XY}\mathbb{1}_A\|_{L^p(\mu\otimes\mu)} &= \left(\int_{X\times Y}\int_{X\times Y} \left|\omega_X(x,x')\mathbb{1}_A(x,y,x',y')\right|^p \, d\mu(x,y)\, d\mu(x',y') \right)^{1/p} \\ &= \left(\int_X \left|\omega_X(x,x)\right|^p \, d\mu_X(x) \right)^{1/p} = \operatorname{tr}_p(X). \end{aligned}$$

Similarly, $\|\psi_{XY}\mathbb{1}_A\|_{L^p(\mu\otimes\mu)} = \operatorname{tr}_p(Y)$. Thus $|\operatorname{tr}_p(X) - \operatorname{tr}_p(Y)| < 2\eta$. Since $\eta > d_{\mathcal{N},p}(X,Y)$ was arbitrary, it follows that:

$$\begin{aligned} |\operatorname{size}_p(X) - \operatorname{size}_p(Y)| &\le 2d_{\mathcal{N},p}(X,Y) \\ |\operatorname{tr}_p(X) - \operatorname{tr}_p(Y)| &\le 2d_{\mathcal{N},p}(X,Y). \end{aligned}$$

We have already remarked that the proof for $p = \infty$ is analogous. Thus we conclude that the preceding inequalities hold for all $p \in [1,\infty]$. □

Theorem 26 (Interleaving stability of the sublevel/superlevel set weight invariants)**.** *Let $p \in [1,\infty]$. The sub^w_p invariant is interleaving-stable with interleaving constant $\alpha = 1$ and interleaving function $d_{\mathcal{N},\infty}$. The sup^w_p invariant is interleaving-stable with interleaving constant $\alpha = -1$ and interleaving function $-d_{\mathcal{N},\infty}$.*

Proof. Let $t_0 \in \mathbb{R}$, and let $(X,\omega_X,\mu_X), (Y,\omega_Y,\mu_Y) \in \mathcal{N}$. Via Lemma 21, write $\varepsilon := d^{\mathcal{GP}}_{\mathcal{N},0}(X,Y) = d_{\mathcal{N},\infty}(X,Y)$. Using Theorem 17, let μ be an optimal coupling between μ_X and μ_Y for which $d_{\mathcal{N},\infty}(X,Y)$ is achieved.

For each $t \in \mathbb{R}$, write $A(X,t) := \{(x,x') \in X\times X : \omega_X(x,x') \le t\} = \{\omega_X \le t\}$. Similarly write $A(Y,t) := \{\omega_Y \le t\}$ for each $t \in \mathbb{R}$.

Let $B := \{(x,y,x',y') \in (X\times Y)^2 : |\omega_X(x,x') - \omega_Y(y,y')| \ge \varepsilon\}$. Also let G denote the complement of B, i.e. $G := \{(x,y,x',y') \in (X\times Y)^2 : |\omega_X(x,x') - \omega_Y(y,y')| < \varepsilon\}$. In particular, notice that for any $(x,y,x',y') \in G$, we have $\omega_X(x,x') < \varepsilon + \omega_Y(y,y')$.

By the definition of ε, we have $\mu^{\otimes 2}(B) = 0$, and hence $\mu^{\otimes 2}(G) = 1$.

In what follows, we will focus on the case $p \in [1,\infty)$ and write out the integrals explicitly. An analogous proof holds for $p=\infty$. We have:

$$\begin{aligned}
\mathrm{sub}^w_{p,t_0}(X) &= \left(\int_{A(X,t_0)} \left|\omega_X(x,x')\right|^p d\mu_X^{\otimes 2}(x,x')\right)^{1/p}\\
&= \left(\int_{A(X,t_0)\times Y^2} \left|\omega_X(x,x')\right|^p d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p}\\
&= \left(\int_{G\cap(A(X,t_0)\times Y^2)} \left|\omega_X(x,x')\right|^p d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p}\\
&= \left(\int_{(X\times Y)^2} \mathbb{1}^p_{G\cap(A(X,t_0)\times Y^2)} \left|\omega_X(x,x')-\omega_Y(y,y')+\omega_Y(y,y')\right|^p\, d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p}.
\end{aligned}$$

For convenience, write $H := G\cap \big(A(X,t_0)\times Y^2\big)$. Using Minkowski's inequality, we have:

$$\begin{aligned}
&\leq \left(\int_{(X\times Y)^2} \mathbb{1}^p_{H} \left|\omega_X(x,x')-\omega_Y(y,y')\right|^p\, d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p}\dots\\
&\qquad + \left(\int_{(X\times Y)^2} \mathbb{1}^p_{H} \left|\omega_Y(y,y')\right|^p\, d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p}.
\end{aligned}$$

For any $(x,y,x',y')\in H$, we have $|\omega_X(x,x')-\omega_Y(y,y')|<\varepsilon$. Also we have $\omega_Y(y,y')<\varepsilon+\omega_X(x,x')\leq \varepsilon+t_0$. From the latter, we know $G\cap\big(A(X,t_0)\times Y^2\big)\subseteq X^2\times A(Y,t_0+\varepsilon)$. So we continue the previous expression as below:

$$\leq \left(\int_{(X\times Y)^2} \mathbb{1}^p_{H} |\varepsilon|^p\, d\mu^{\otimes 2}\right)^{1/p} + \left(\int_{X^2\times A(Y,t_0+\varepsilon)} \left|\omega_Y(y,y')\right|^p d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p} \tag{2}$$

$$\begin{aligned}
&\leq \varepsilon + \left(\int_{A(Y,t_0+\varepsilon)} \left|\omega_Y(y,y')\right|^p\, d\mu_Y^{\otimes 2}(y,y')\right)^{1/p}\\
&= \mathrm{sub}^w_{p,t_0+\varepsilon}(Y)+\varepsilon.
\end{aligned}$$

Analogously, we have

$$\mathrm{sub}^w_{p,t_0+\varepsilon}(Y)\leq \mathrm{sub}^w_{p,t_0+2\varepsilon}(X)+\varepsilon.$$

This yields interleaving for $p\in[1,\infty)$. For $p=\infty$, we use the same arguments about G and B to obtain:

$$\begin{aligned}
\mathrm{sub}^w_{p,t_0}(X) &= \sup\{\left|\omega_X(x,x')\right| : x,x'\in\mathrm{supp}(\mu_X),\ \omega_X(x,x')\leq t_0\}\\
&\leq \sup\{\left|\omega_Y(y,y')\right|+\varepsilon : y,y'\in\mathrm{supp}(\mu_Y),\ \omega_Y(y,y')\leq t_0+\varepsilon\}\\
&\leq \mathrm{sub}^w_{p,t_0+\varepsilon}(Y)+\varepsilon.
\end{aligned}$$

Thus we have interleaving for all $p\in[1,\infty]$.

The case for the $\sup^w_p$ invariant is similar, except in step 2 above. In this case, we note that for any $(x,y,x',y')\in H$, we have $\omega_Y(y,y')>\omega_X(x,x')-\varepsilon\geq t_0-\varepsilon$. Thus we have $H=G\cap\big(A(X,t_0)\times Y^2\big)\subseteq X^2\times A(Y,t_0-\varepsilon)$, and so:

$$\left(\int_{(X\times Y)^2} \mathbb{1}^p_{H} \left|\omega_Y(y,y')\right|^p\, d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p} \leq \left(\int_{X^2\times A(Y,t_0-\varepsilon)} \left|\omega_Y(y,y')\right|^p\, d\mu^{\otimes 2}(x,y,x',y')\right)^{1/p}.$$

It follows that we have:

$$\sup^{w}_{p,t_0}(X) \leq \sup^{w}_{p,t_0-\varepsilon}(X) + \varepsilon \leq \sup^{w}_{p,t_0-2\varepsilon}(Y) + 2\varepsilon.$$

The $p = \infty$ is proved analogously. □

3.2. **Local invariants.** The global invariants defined above have local counterparts that we now define.

Example 27 (A generalized eccentricity function)**.** Let (X, ω_X, μ_X) be a measure network. Then consider the $\operatorname{ecc}^{\text{out}}_{p,X} : X \to \mathbb{R}_+$ map

$$\operatorname{ecc}^{\text{out}}_{p,X}(s) := \left(\int_X |\omega_X(s,x)|^p \, d\mu_X(x) \right)^{1/p} = \|\omega_X(s,\cdot)\|_{L^p(\mu_X)}.$$

The $p = \infty$ version is defined analogously, with the integral replaced by a supremum over the support. We can also replace $\omega_X(s,\cdot)$ above with $\omega_X(\cdot,s)$ to obtain another map $\operatorname{ecc}^{\text{in}}_{p,X}$. In general, the two maps will not agree due to the asymmetry of the network. This invariant is an asymmetric generalization of the *p-eccentricity* function for metric measure spaces [Mém11, Definition 5.3]

Example 28 (A joint eccentricity function)**.** Let (X, ω_X, μ_X) and (Y, ω_Y, μ_Y) be two measure networks, and let $p \in [1, \infty]$. Define the *(outer) joint eccentricity function* $\operatorname{ecc}^{\text{out}}_{p,X,Y} : X \times Y \to \mathbb{R}_+$ of X and Y as follows: for each $(s,t) \in X \times Y$,

$$\operatorname{ecc}^{\text{out}}_{p,X,Y}(s,t) := \inf_{\mu \in \mathscr{C}(\mu_X,\mu_Y)} \|\omega_X(s,\cdot) - \omega_Y(t,\cdot)\|_{L^p(\mu)}.$$

For $p \in [1, \infty)$, this invariant has the following form:

$$\operatorname{ecc}^{\text{out}}_{p,X,Y}(s,t) := \left(\inf_{\mu \in \mathscr{C}(\mu_X,\mu_Y)} \int_{X \times Y} |\omega_X(s,x') - \omega_Y(t,y')|^p \, d\mu(x',y') \right)^{1/p}.$$

One obtains the *inner joint eccentricity function* by using the term $\omega_X(\cdot,s) - \omega_Y(\cdot,t)$ above, and we denote this by $\operatorname{ecc}^{\text{in}}_{p,X,Y}$.

Theorem 29 (Stability of local $\mathbb{R}$-valued invariants)**.** *The eccentricity and joint eccentricity invariants are both Lipschitz stable, with Lipschitz constant* 2*. Formally, for any* $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$*, we have:*

$$\inf_{\mu \in \mathscr{C}(\mu_X,\mu_Y)} \left\| \operatorname{ecc}^{\text{out}}_{p,X} - \operatorname{ecc}^{\text{out}}_{p,Y} \right\|_{L^p(\mu)} \leq 2 d_{\mathcal{N},p}(X,Y), \qquad \text{(eccentricity bound)}$$

$$\inf_{\mu \in \mathscr{C}(\mu_X,\mu_Y)} \left\| \operatorname{ecc}^{\text{out}}_{p,X,Y} \right\|_{L^p(\mu)} \leq 2 d_{\mathcal{N},p}(X,Y). \qquad \text{(joint eccentricity bound)}$$

Moreover, the joint eccentricity invariant provides a stronger bound than the eccentricity bound, i.e.

$$\inf_{\mu \in \mathscr{C}(\mu_X,\mu_Y)} \left\| \operatorname{ecc}^{\text{out}}_{p,X} - \operatorname{ecc}^{\text{out}}_{p,Y} \right\|_{L^p(\mu)} \leq \inf_{\mu \in \mathscr{C}(\mu_X,\mu_Y)} \left\| \operatorname{ecc}^{\text{out}}_{p,X,Y} \right\|_{L^p(\mu)} \leq 2 d_{\mathcal{N},p}(X,Y).$$

Finally, the analogous bounds hold in the case of the inner eccentricity and inner joint eccentricity functions.

Remark 30. The analogous bounds in the setting of metric measure spaces were provided in [Mém07], where the eccentricity and joint eccentricity bounds were called the First and Third Lower Bounds, respectively. The TLB later appeared in [SS13].

Proof of Theorem 29. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$. For each $s \in X$ and $t \in Y$, define $\varphi^{st}_{XY} : (X \times Y)^2 \to \mathbb{R}$ by writing $\varphi^{st}_{XY}(x,y,x',y') := \omega_X(s,x')$, and define $\psi^{st}_{XY} : (X \times Y)^2 \to \mathbb{R}$ by writing $\psi^{st}_{XY}(x,y,x',y') := \omega_Y(t,y')$. For convenience, we write $\mathscr{C} := \mathscr{C}(\mu_X, \mu_Y)$.

First let $p \in [1, \infty)$. Let $\eta > d_{\mathcal{N},p}(X,Y)$, and let $\mu \in \mathscr{C}(\mu_X, \mu_Y)$ be a coupling such that $\operatorname{dis}_p(\mu) < 2\eta$. Then by applying Minkowski's inequality, we obtain

$$\left| \|\varphi^{st}_{XY}\|_{L^p(\mu \otimes \mu)} - \|\psi^{st}_{XY}\|_{L^p(\mu \otimes \mu)} \right| \leq \|\varphi^{st}_{XY} - \psi^{st}_{XY}\|_{L^p(\mu \otimes \mu)}. \qquad (3)$$

In particular, because $x \mapsto x^p$ is increasing on $\mathbb{R}_+$, we also have

$$\left| \|\varphi^{st}_{XY}\|_{L^p(\mu\otimes\mu)} - \|\psi^{st}_{XY}\|_{L^p(\mu\otimes\mu)} \right|^p \leq \|\varphi^{st}_{XY} - \psi^{st}_{XY}\|^p_{L^p(\mu\otimes\mu)}. \tag{4}$$

Next we observe:

$$\begin{aligned}\|\varphi^{st}_{XY}\|_{L^p(\mu\otimes\mu)} &= \left(\int_{X\times Y}\int_{X\times Y} \left|\varphi^{st}_{XY}(x,y,x',y')\right|^p \, d\mu(x,y)\, d\mu(x',y') \right)^{1/p} \\ &= \left(\int_{X\times Y}\int_{X\times Y} \left|\omega_X(s,x')\right|^p \, d\mu(x,y)\, d\mu(x',y') \right)^{1/p} \\ &= \left(\int_{X} \left|\omega_X(s,x')\right|^p \, d\mu_X(x') \right)^{1/p} = \mathrm{ecc}^{\mathrm{out}}_{p,X}(s).\end{aligned}$$

Similarly, $\|\psi^{st}_{XY}\|_{L^p(\mu\otimes\mu)} = \mathrm{ecc}^{\mathrm{out}}_{p,Y}(t)$.

For the right side of Inequality 4, we have:

$$\begin{aligned}\|\varphi^{st}_{XY} - \psi^{st}_{XY}\|^p_{L^p(\mu\otimes\mu)} &= \int_{X\times Y}\int_{X\times Y} \left|\varphi^{st}_{XY}(x,y,x',y') - \psi^{st}_{XY}(x,y,x',y')\right|^p \, d\mu(x,y)\, d\mu(x',y') \\ &= \int_{X\times Y}\int_{X\times Y} \left|\omega_X(s,x') - \omega_Y(t,y')\right|^p \, d\mu(x,y)\, d\mu(x',y') \\ &= \int_{X\times Y} \left|\omega_X(s,x') - \omega_Y(t,y')\right|^p \, d\mu(x',y').\end{aligned}$$

Putting all these observations together with Inequality 4, we have:

$$\left|\mathrm{ecc}^{\mathrm{out}}_{p,X}(s) - \mathrm{ecc}^{\mathrm{out}}_{p,Y}(t)\right|^p \leq \int_{X\times Y} \left|\omega_X(s,x') - \omega_Y(t,y')\right|^p \, d\mu(x',y').$$

The left hand side above is independent of the coupling μ, so we can infimize over $\mathscr{C}(\mu_X,\mu_Y)$:

$$\left|\mathrm{ecc}^{\mathrm{out}}_{p,X}(s) - \mathrm{ecc}^{\mathrm{out}}_{p,Y}(t)\right|^p \leq \inf_{\nu\in\mathscr{C}} \int_{X\times Y} \left|\omega_X(s,x') - \omega_Y(t,y')\right|^p \, d\nu(x',y') = \left(\mathrm{ecc}^{\mathrm{out}}_{p,X,Y}(s,t)\right)^p.$$

Also observe:

$$\begin{aligned}&\left(\int_{X\times Y} \|\varphi^{st}_{XY} - \psi^{st}_{XY}\|^p_{L^p(\mu\otimes\mu)} \, d\mu(s,t) \right)^{1/p} \\ &\qquad = \left(\int_{X\times Y}\int_{X\times Y} \left|\omega_X(s,x') - \omega_Y(t,y')\right|^p \, d\mu(x',y')\, d\mu(s,t) \right)^{1/p} \\ &\qquad = \mathrm{dis}_p(\mu) < 2\eta.\end{aligned}$$

Thus we obtain:

$$\left(\int_{X\times Y} \left|\mathrm{ecc}^{\mathrm{out}}_{p,X}(s) - \mathrm{ecc}^{\mathrm{out}}_{p,Y}(t)\right|^p \, d\mu(s,t) \right)^{1/p} \leq \left(\int_{X\times Y} \left(\mathrm{ecc}^{\mathrm{out}}_{p,X,Y}(s,t)\right)^p \, d\mu(s,t) \right)^{1/p} < 2\eta.$$

Since $\eta > 2d_{\mathcal{N},p}(X,Y)$ was arbitrary, it follows that

$$\begin{aligned}\left(\inf_{\mu\in\mathscr{C}} \int_{X\times Y} \left|\mathrm{ecc}^{\mathrm{out}}_{p,X}(s) - \mathrm{ecc}^{\mathrm{out}}_{p,Y}(t)\right|^p \, d\mu(s,t) \right)^{1/p} &\leq \left(\inf_{\mu\in\mathscr{C}} \int_{X\times Y} \left(\mathrm{ecc}^{\mathrm{out}}_{p,X,Y}(s,t)\right)^p \, d\mu(s,t) \right)^{1/p} \\ &\leq 2d_{\mathcal{N},p}(X,Y).\end{aligned} \tag{5}$$

This proves the $p \in [1,\infty)$ case. The $p = \infty$ case follows by applying Minkowski's inequality to obtain Inequality 3, and working analogously from there. Finally, we remark that the same proof holds for the $\mathrm{ecc}^{\mathrm{in}}_{p,X}$ and $\mathrm{ecc}^{\mathrm{in}}_{p,X,Y}$ functions. □

3.3. **Distribution-valued invariants: local and global pushforwards.** The invariants we have defined above have all been $\mathbb{R}$-valued invariants, where there is a clear choice of metric. Next we define some *distribution-valued invariants* that arise naturally from our setup. Specifically, these invariants map measure networks into distributions over $\mathbb{R}$. There are different choices one can make when comparing distributions: the Wasserstein metric is one natural candidate, but one can also consider the Prokhorov metric or dissimilarities such as the Kullback-Leibler divergence.

Example 31 (Pushforward via ω_X)**.** Recall that given any (X, ω_X, μ_X), the corresponding pushforward of $\mu_X \otimes \mu_X$ via ω_X is given as follows: for any generator of $\operatorname{Borel}(\mathbb{R})$ of the form $(a, b) \subseteq \mathbb{R}$,

$$\begin{aligned}(\omega_X)_*(\mu_X \otimes \mu_X)(a, b) &:= (\mu_X \otimes \mu_X)(\{\omega_X \in (a, b)\}) \\ &= \int_X \int_X \mathbb{1}_{\{\omega_X \in (a,b)\}}(x, x')\, d\mu_X(x)\, d\mu_X(x').\end{aligned}$$

For convenience, we define $\nu_X := (\omega_X)_*(\mu_X^{\otimes 2})$. This distribution is completely determined by its cumulative distribution function, which we denote by F_{ω_X}. This is a function $\mathbb{R} \to [0, 1]$ given by:

$$F_{\omega_X}(t) := (\mu_X \otimes \mu_X)(\{\omega_X \leq t\}) = \int_X \int_X \mathbb{1}_{\{\omega_X \leq t\}}(x, x')\, d\mu_X(x)\, d\mu_X(x').$$

The distribution-valued invariant above is a global invariant. The corresponding local versions are below.

Example 32 (Pushforward via a single coordinate of ω_X)**.** Let (X, ω_X, μ_X) and $x \in X$ be given. Then we can define *local* distribution-valued invariants as follows: for any generator of $\operatorname{Borel}(\mathbb{R})$ of the form $(a, b) \subseteq \mathbb{R}$,

$$\begin{aligned}(\omega_X(x, \cdot))_*\mu_X(a, b) &:= \mu_X\big(\{x' \in X : \omega_X(x, x') \in (a, b)\}\big) \\ (\omega_X(\cdot, x))_*\mu_X(a, b) &:= \mu_X\big(\{x' \in X : \omega_X(x', x) \in (a, b)\}\big).\end{aligned}$$

We adopt the following shorthand:

$$\lambda_X(x) := (\omega_X(x, \cdot))_*\mu_X, \qquad \rho_X(x) := (\omega_X(\cdot, x)_*\mu_X.$$

Here we write λ and ρ to refer to the "left" and "right" arguments, respectively. The corresponding distribution functions are defined as follows: for any $t \in \mathbb{R}$,

$$\begin{aligned}F_{\omega_X(x,\cdot)}(t) &:= \mu_X(\{\omega_X(x, \cdot) \leq t\}) = \int_X \mathbb{1}_{\{\omega_X(x,\cdot) \leq t\}}(x')\, d\mu_X(x') \\ F_{\omega_X(\cdot,x)}(t) &:= \mu_X(\{\omega_X(\cdot, x) \leq t\}) = \int_X \mathbb{1}_{\{\omega_X(\cdot,x) \leq t\}}(x')\, d\mu_X(x').\end{aligned}$$

It is interesting to note that we get such a pair of distributions for each $x \in X$. Thus we can add yet another layer to this construction, via the maps $\mathcal{N} \to \mathscr{P}(\operatorname{Prob}(\mathbb{R}))$ defined by writing

$$\begin{aligned}&(X, \omega_X, \mu_X) \mapsto \{\lambda_X(x) : x \in X\}, \text{ and} \\ &(X, \omega_X, \mu_X) \mapsto \{\rho_X(x) : x \in X\} \text{ for each } (X, \omega_X, \mu_X) \in \mathcal{N}.\end{aligned}$$

Assume for now that we equip $\operatorname{Prob}(\mathbb{R})$ with the Wasserstein metric. Write $\mathbb{X} := \{\lambda_X(x)\}_{x \in X}$, let $d_{\mathbb{X}}$ denote the Wasserstein metric, and let $\mu_{\mathbb{X}} := (\lambda_X)_*\mu_X$. More specifically, for any $A \in \operatorname{Borel}(\mathbb{X})$, we have $\mu_{\mathbb{X}}(A) = \mu_X(\{x \in X : \lambda_X(x) \in A\})$. This yields a metric measure space $(\mathbb{X}, d_{\mathbb{X}}, \mu_{\mathbb{X}})$. So even though we do not start off with a metric space, the operation of passing into distributions over $\mathbb{R}$ forces a metric structure on (X, ω_X, μ_X).

Next let $(Y, \omega_Y, \mu_Y) \in \mathcal{N}$, and suppose $(\mathbb{Y}, d_{\mathbb{Y}}, \mu_{\mathbb{Y}})$ are defined as above. Since $\mathbb{X}, \mathbb{Y} \subseteq \operatorname{Prob}(\mathbb{R})$, we know that $\mu_{\mathbb{X}}, \mu_{\mathbb{Y}}$ are both distributions on $\operatorname{Prob}(\mathbb{R})$. Thus we can compare them via the p-Wasserstein distance as follows, for $p \in [1, \infty)$:

$$d_{W,p}(\mu_{\mathbb{X}}, \mu_{\mathbb{Y}}) = \inf_{\mu \in \mathscr{C}(\mu_{\mathbb{X}}, \mu_{\mathbb{Y}})} \left(\int_{\operatorname{Prob}(\mathbb{R})^2} d_{W,p}(\lambda_X(x), \lambda_Y(y))^p\, d\mu(\lambda_X(x), \lambda_Y(y)) \right)^{1/p}$$

By the change of variables formula, this quantity coincides with one that we show below to be a lower bound for $2d_{\mathcal{N},p}(X,Y)$ (cf. Inequality 13 of Theorem 35).

Example 33 (Pushforward via eccentricity)**.** Let $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y)\in\mathcal{N}$, and let $(a,b)\in\text{Borel}(\mathbb{R})$. Recall the outer and inner eccentricity functions $\text{ecc}^{\text{out}}_{p,X}$ and $\text{ecc}^{\text{in}}_{p,X}$ from Example 27. These functions induce distributions as follows:

$$\left(\text{ecc}^{\text{out}}_{p,X}\right)_* \mu_X(a,b) = \mu_X\left(\left\{x\in X : \text{ecc}^{\text{out}}_{p,X}(x)\in(a,b)\right\}\right),$$
$$\left(\text{ecc}^{\text{in}}_{p,X}\right)_* \mu_X(a,b) = \mu_X\left(\left\{x\in X : \text{ecc}^{\text{in}}_{p,X}(x)\in(a,b)\right\}\right).$$

Next let $\mu\in\mathscr{C}(\mu_X,\mu_Y)$ and recall the joint outer/inner eccentricity functions $\text{ecc}^{\text{out}}_{p,X,Y}$ and $\text{ecc}^{\text{in}}_{p,X,Y}$ from Example 28. These functions induce distributions as below:

$$\left(\text{ecc}^{\text{out}}_{p,X,Y}\right)_* \mu(a,b) = \mu\left(\left\{(x,y)\in X\times Y : \text{ecc}^{\text{out}}_{p,X,Y}(x,y)\in(a,b)\right\}\right),$$
$$\left(\text{ecc}^{\text{in}}_{p,X,Y}\right)_* \mu(a,b) = \mu\left(\left\{(x,y)\in X\times Y : \text{ecc}^{\text{in}}_{p,X,Y}(x,y)\in(a,b)\right\}\right).$$

In general, each local invariant $\iota_{p,X}: X\to\mathbb{R}_+$ yields a distribution on $\mathbb{R}$ by taking the pushforward of μ_X via $\iota_{p,X}$. Distribution valued invariants provide interesting means of compressing the information in a network into a distribution or histogram over $\mathbb{R}_+$. We now prove quantitative stability results for the preceding invariants. The following lemma is a particular statement of the change of variables theorem that we use later.

Lemma 34 (Change of variables)**.** *Let $(X,\mathcal{F}_X,\mu_X)$ and $(Y,\mathcal{F}_Y,\mu_Y)$ be two probability spaces. Let $f: X\to\mathbb{R}$ and $g: Y\to\mathbb{R}$ be two measurable functions. Write $f_*\mu_X$ and $g_*\mu_Y$ to denote the pushforward distributions on $\mathbb{R}$. Let $T: X\times Y\to\mathbb{R}^2$ be the map $(x,y)\mapsto(f(x),g(y))$ and let $h:\mathbb{R}^2\to\mathbb{R}_+$ be measurable. Next let $\mu\in\mathscr{C}(\mu_X,\mu_Y)$. Then $T_*\mu\in\mathscr{C}(f_*\mu_X,g_*\mu_Y)$, and the following inequality holds:*

$$\left(\inf_{\nu\in\mathscr{C}(f_*\mu_X,g_*\mu_Y)}\int_{\mathbb{R}^2} h(a,b)\,d(T_*\mu)(a,b)\right)^{1/p} \le \left(\int_{X\times Y} h(T(x,y))\,d\mu(x,y)\right)^{1/p}.$$

This is essentially the same as [Mém11, Lemma 6.1] but stated for general probability spaces instead of metric measure spaces. The form of the statement in [Mém11, Lemma 6.1] is slightly different, but it can be obtained from the statement presented above by using [Vil03, Remark 2.19].

Proof of Lemma 34. First we check that $T_*\mu\in\mathscr{C}(f_*\mu_X,g_*\mu_Y)$. Let $A\in\text{Borel}(\mathbb{R})$. Then,

$$\begin{aligned} T_*\mu(A\times\mathbb{R}) = \mu\left(\{(x,y)\in X\times Y : T(x,y)\in A\times\mathbb{R}\}\right) &= \mu\left(\{(x,y)\in X\times Y : f(x)\in A\}\right) \\ &= f_*\mu_X(A). \end{aligned}$$

Similarly we check $T_*\mu(\mathbb{R}\times A) = g_*\mu_Y(A)$.

Next we check the inequality. By the change of variables formula, we have:

$$\left(\int_{\mathbb{R}^2} h(a,b)\,d(T_*\mu)(a,b)\right)^{1/p} = \left(\int_{X\times Y} h(T(x,y))\,d\mu(x,y)\right)^{1/p}.$$

We have already verified that $T_*\mu\in\mathscr{C}(f_*\mu_X,g_*\mu_Y)$. The inequality is obtained by infimizing the left hand side over all possible couplings $\nu\in\mathscr{C}(f_*\mu_X,g_*\mu_Y)$. This does not affect the right hand side, which is independent of such couplings. ∎

Theorem 35 (Stability of the ω_X and eccentricity-pushforward distributions)**.** *Let $(X,\omega_X,\mu_X),(Y,\omega_Y,\mu_Y)\in\mathcal{N}$. Then we have the following statements about Lipschitz stability, for $p\in[1,\infty)$:*

$$2d_{\mathcal{N},p}(X,Y) \ge \inf_{\mu\in\mathscr{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)}\left(\int_{X^2\times Y^2}\left|\omega_X(x,x')-\omega_Y(y,y')\right|^p\,d\mu(x,x',y,y')\right)^{1/p} \tag{6}$$
$$\ge \inf_{\nu\in\mathscr{C}(\nu_X,\nu_Y)}\left(\int_{\mathbb{R}^2}|a-b|^p\,d\nu(a,b)\right)^{1/p}. \tag{7}$$

$$2d_{\mathcal{N},p}(X,Y) \geq \inf_{\mu\in\mathscr{C}(\mu_X,\mu_Y)} \left(\int_{X\times Y} \left|\mathrm{ecc}^{\mathrm{out}}_{p,X}(x) - \mathrm{ecc}^{\mathrm{out}}_{p,Y}(y)\right|^p \, d\mu(x,y)\right)^{1/p} \tag{8}$$

$$\geq \inf_{\gamma\in\mathscr{C}((\mathrm{ecc}^{\mathrm{out}}_{p,X})_*\mu_X,(\mathrm{ecc}^{\mathrm{out}}_{p,Y})_*\mu_Y)} \left(\int_{R^2} |a-b|^p \, d\gamma(a,b)\right)^{1/p}. \tag{9}$$

$$2d_{\mathcal{N},p}(X,Y) \geq \inf_{\mu\in\mathscr{C}(\mu_X,\mu_Y)} \left(\int_{X\times Y} \left|\mathrm{ecc}^{\mathrm{in}}_{p,X}(x) - \mathrm{ecc}^{\mathrm{in}}_{p,Y}(y)\right|^p \, d\mu(x,y)\right)^{1/p} \tag{10}$$

$$\geq \inf_{\gamma\in\mathscr{C}((\mathrm{ecc}^{\mathrm{in}}_{p,X})_*\mu_X,(\mathrm{ecc}^{\mathrm{in}}_{p,Y})_*\mu_Y)} \left(\int_{R^2} |a-b|^p \, d\gamma(a,b)\right)^{1/p}. \tag{11}$$

$$2d_{\mathcal{N},p}(X,Y) \geq \inf_{\mu\in\mathscr{C}(\mu_X,\mu_Y)} \left(\int_{X\times Y} \inf_{\gamma\in\mathscr{C}(\mu_X,\mu_Y)} \int_{X\times Y} \left|\omega_X(x,x') - \omega_Y(y,y')\right|^p \, d\gamma(x',y')\, d\mu(x,y)\right)^{1/p} \tag{12}$$

$$\geq \inf_{\mu\in\mathscr{C}(\mu_X,\mu_Y)} \left(\int_{X\times Y} \inf_{\gamma\in\mathscr{C}(\lambda_X(x),\lambda_Y(y))} \int_{\mathbb{R}^2} |a-b|^p \, d\gamma(a,b)\, d\mu(x,y)\right)^{1/p}. \tag{13}$$

$$2d_{\mathcal{N},p}(X,Y) \geq \inf_{\mu\in\mathscr{C}(\mu_X,\mu_Y)} \left(\int_{X\times Y} \inf_{\gamma\in\mathscr{C}(\mu_X,\mu_Y)} \int_{X\times Y} \left|\omega_X(x,x') - \omega_Y(y,y')\right|^p \, d\gamma(x,y)\, d\mu(x',y')\right)^{1/p} \tag{14}$$

$$\geq \inf_{\mu\in\mathscr{C}(\mu_X,\mu_Y)} \left(\int_{X\times Y} \inf_{\gamma\in\mathscr{C}(\rho_X(x'),\rho_Y(y'))} \int_{\mathbb{R}^2} |a-b|^p \, d\gamma(a,b)\, d\mu(x',y')\right)^{1/p}. \tag{15}$$

Here recall that $\nu_X = (\omega_X)_*(\mu_X^{\otimes 2})$, $\nu_Y = (\omega_Y)_*(\mu_Y^{\otimes 2})$, $\lambda_X = (\omega_X(x,\cdot))_*\mu_X$, $\lambda_Y = (\omega_Y(y,\cdot))_*\mu_Y$, $\rho_X = (\omega_X(\cdot,x))_*\mu_X$, and $\rho_Y = (\omega_Y(\cdot,y))_*\mu_Y$. Inequalities 6-7 appeared as the Second Lower Bound and its relaxation in [Mém07]. *Inequalities 8, 10, 12, and 14 are the eccentricity bounds in Theorem 29. Inequalities 9, 11, 13, and 15 are their relaxations. In the symmetric case, these outer/inner pairs of inequalities coincide; they appeared as the First and Third Lower Bounds and their relaxations in* [Mém07].

In Inequality 7, both ν_X and ν_Y are probability distributions on $\mathbb{R}$, and the right hand side is precisely the p-Wasserstein distance between ν_X and ν_Y. Analogous statements hold for Inequalities 13 and 15.

Proof of Theorem 35. Consider the probability spaces $X\times X$ and $Y\times Y$, equipped with the product measures $\mu_X\otimes\mu_X$ and $\mu_Y\otimes\mu_Y$. For convenience, we define the shorthand notation $\nu_X := (\omega_X)_*(\mu_X\otimes\mu_X)$ and $\nu_Y := (\omega_Y)_*(\mu_Y\otimes\mu_Y)$. Let $T : (X\times X)\times(Y\times Y)\to\mathbb{R}^2$ be the map $(x,x',y,y')\mapsto(\omega_X(x,x'),\omega_Y(y,y'))$. Also let $h:\mathbb{R}^2\to\mathbb{R}$ be the map $(a,b)\mapsto|a-b|^p$.

Let $\eta > d_{\mathcal{N},p}(X,Y)$, and let $\mu\in\mathscr{C}(\mu_X,\mu_Y)$ be a coupling such that $\mathrm{dis}_p(\mu) < 2\eta$. Also let τ be a measure on $X\times X\times Y\times Y$ defined by writing $\tau(A,A',B,B') := \mu(A,B)\mu(A',B')$ for $A,A'\in\mathrm{Borel}(X)$ and $B,B'\in\mathrm{Borel}(Y)$. Then $\tau\in\mathscr{C}(\mu_X^{\otimes 2},\mu_Y^{\otimes 2})$.

By Lemma 34, we know that $T_*\tau\in\mathscr{C}(\nu_X,\nu_Y)$. By the change of variables formula and Fubini's theorem,

$$\begin{aligned}\left(\int_{\mathbb{R}^2} |a-b|^p \, d(T_*\tau)(a,b)\right)^{1/p} &= \left(\int_{X^2\times Y^2} \left|\omega_X(x,x')-\omega_Y(y,y')\right|^p \, d\tau(x,x',y,y')\right)^{1/p}\\ &= \left(\int_{X^2\times Y^2} \left|\omega_X(x,x')-\omega_Y(y,y')\right|^p \, d\left(\mu(x,y)\mu(x',y')\right)\right)^{1/p}\\ &= \left(\int_{X\times Y}\int_{X\times Y} \left|\omega_X(x,x')-\omega_Y(y,y')\right|^p \, d\mu(x,y)\, d\mu(x',y')\right)^{1/p} < 2\eta.\end{aligned}$$

We infimize over $\mathscr{C}(\mu_X^{\otimes 2}, \mu_Y^{\otimes 2})$, use the fact that $\eta > d_{\mathcal{N},p}(X,Y)$ was arbitrary, and apply Lemma 34 to obtain:

$$2d_{\mathcal{N},p}(X,Y) \geq \left(\inf_{\mu \in \mathscr{C}(\mu_X^{\otimes 2}, \mu_Y^{\otimes 2})} \int_{X^2 \times Y^2} \left|\omega_X(x,x') - \omega_Y(y,y')\right|^p \, d\mu(x,x',y,y') \right)^{1/p}$$
$$\geq \inf_{\gamma \in \mathscr{C}(\nu_X, \nu_Y)} \left(\int_{\mathbb{R}^2} |a-b|^p \, d\gamma(a,b) \right)^{1/p}.$$

This yields Inequalities 6-7.

Next we consider the distributions induced by the $\mathrm{ecc}^{\mathrm{out}}$ function. For convenience, write $e_X := (\mathrm{ecc}^{\mathrm{out}}_{p,X})_* \mu_X$ and $e_Y := (\mathrm{ecc}^{\mathrm{out}}_{p,Y})_* \mu_Y$. Now let $T : X \times Y \to \mathbb{R}$ be the map $(x,y) \mapsto (\mathrm{ecc}^{\mathrm{out}}_{p,X}(x), \mathrm{ecc}^{\mathrm{out}}_{p,Y}(y))$, and let $h : \mathbb{R}^2 \to \mathbb{R}$ be the map $(a,b) \mapsto |a-b|^p$. By the change of variables formula and Theorem 29, we know

$$\inf_{\mu \in \mathscr{C}(\mu_X, \mu_Y)} \left(\int_{\mathbb{R}^2} |a-b|^p \, d(T_*\mu)(a,b) \right)^{1/p} = \inf_{\mu \in \mathscr{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \left|\mathrm{ecc}^{\mathrm{out}}_{p,X}(x) - \mathrm{ecc}^{\mathrm{out}}_{p,X}(y)\right|^p \, d\mu(x,y) \right)^{1/p}$$
$$\leq 2d_{\mathcal{N},p}(X,Y).$$

By Lemma 34, we know that $T_*\mu \in \mathscr{C}(e_X, e_Y)$ and also the following:

$$\inf_{\gamma \in \mathscr{C}(e_X, e_Y)} \left(\int_{\mathbb{R}^2} |a-b|^p \, d\gamma(a,b) \right)^{1/p} \leq \inf_{\mu \in \mathscr{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \left|\mathrm{ecc}^{\mathrm{out}}_{p,X}(x) - \mathrm{ecc}^{\mathrm{out}}_{p,X}(y)\right|^p \, d\mu(x,y) \right)^{1/p}$$
$$\leq 2d_{\mathcal{N},p}(X,Y).$$

This proves Inequalities 8-9. Inequalities 10-11 are proved analogously.

Finally we consider the distributions obtained as pushforwards of the joint eccentricity function, i.e. Inequalities 12-15. For each $x \in X$ and $y \in Y$ let $T^{xy} : X \times Y \to \mathbb{R}$ be the map $(x',y') \mapsto (\omega_X(x,x'), \omega_Y(y,y'))$, and let $h : \mathbb{R}^2 \to \mathbb{R}$ be the map $(a,b) \mapsto |a-b|^p$. Let $\gamma \in \mathscr{C}(\mu_X, \mu_Y)$. By the change of variables formula, we have

$$\int_{\mathbb{R}^2} |a-b|^p \, d(T^{xy}_*\gamma)(a,b) = \int_{X \times Y} \left|\omega_X(x,x') - \omega_Y(y,y')\right|^p \, d\gamma(x',y'), \text{ and so}$$

$$\int_{X \times Y} \int_{\mathbb{R}^2} |a-b|^p \, d(T^{xy}_*\gamma)(a,b) \, d\mu(x,y) = \int_{X \times Y} \int_{X \times Y} \left|\omega_X(x,x') - \omega_Y(y,y')\right|^p \, d\gamma(x',y') \, d\mu(x,y).$$

By Lemma 34, $T^{xy}_*\mu \in \mathscr{C}(\lambda_X(x), \lambda_Y(y))$. Applying Theorem 29 and Lemma 34, we have:

$$2d_{\mathcal{N},p}(X,Y) \geq \inf_{\mu \in \mathscr{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathscr{C}(\mu_X, \mu_Y)} \int_{X \times Y} \left|\omega_X(x,x') - \omega_Y(y,y')\right|^p \, d\gamma(x',y') \, d\mu(x,y) \right)^{1/p}$$
$$\geq \inf_{\mu \in \mathscr{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathscr{C}(\lambda_X(x), \lambda_Y(y))} \int_{\mathbb{R}^2} |a-b|^p \, d\gamma(a,b) \, d\mu(x,y) \right)^{1/p}.$$

This verifies Inequalities 12-13. Inequalities 14-15 are proved analogously. □

4. Discussion

We have presented the GW distance as a valid pseudometric on the space of all directed, weighted networks. The crux of this approach is that even though the GW distance was originally formulated for metric measure spaces, the structure of the GW distance automatically forces a metric structure on networks. This yields the insight that the metric structure on the "space of spaces" is not inherited from the metric on the ground spaces.

We have also presented quantitatively stable network invariants that yield readily computable lower bounds on the GW distance. Applications of these invariants to network datasets and numerical experiments will be made available in a later release of this paper.

Acknowledgments This project was supported by NSF grants IIS-1422400, DMS-1723003, and TRIPODS-1740761.

REFERENCES

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[BCC+15] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[Bog07a] Vladimir I Bogachev. *Measure theory*, volume 2. Springer Science & Business Media, 2007.

[Bog07b] Vladimir I Bogachev. *Measure theory*, volume 1. Springer Science & Business Media, 2007.

[CM17] Samir Chowdhury and Facundo Mémoli. Distances and isomorphism between networks and the stability of network invariants. *arXiv preprint arXiv:1708.04727*, 2017.

[Hen16] Reigo Hendrikson. Using Gromov-Wasserstein distance to explore sets of networks. Master's thesis, University of Tartu, 2016.

[Joh70] Roy A Johnson. Atomic and nonatomic measures. *Proceedings of the American Mathematical Society*, 25(3):650–655, 1970.

[KSSF16] Srijan Kumar, Francesca Spezzano, VS Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 221–230. IEEE, 2016.

[Mém07] Facundo Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. *The Eurographics Association*, 2007.

[Mém11] Facundo Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, pages 1–71, 2011. 10.1007/s10208-011-9093-5.

[PCS16] Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pages 2664–2672, 2016.

[SPKS16] Justin Solomon, Gabriel Peyré, Vladimir G Kim, and Suvrit Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (TOG)*, 35(4):72, 2016.

[Sri08] Sashi Mohan Srivastava. *A course on Borel sets*, volume 180. Springer Science & Business Media, 2008.

[SS13] Bernhard Schmitzer and Christoph Schnörr. Modelling convex shape priors and matching based on the Gromov-Wasserstein distance. *Journal of mathematical imaging and vision*, 46(1):143–159, 2013.

[Stu06] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.

[Stu12] Karl-Theodor Sturm. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *arXiv preprint arXiv:1208.0434*, 2012.

[VCF+18] Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Optimal transport for structured data. *arXiv preprint arXiv:1805.09114*, 2018.

[Vil03] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.