Discrete & Computational Geometry
https://doi.org/10.1007/s00454-023-00583-0

The Ultrametric Gromov–Wasserstein Distance

Facundo Mémoli[1] · Axel Munk[2] · Zhengchao Wan[3] · Christoph Weitkamp[2]

Received: 2 July 2021 / Revised: 28 March 2023 / Accepted: 1 April 2023

Abstract
We investigate compact ultrametric measure spaces which form a subset $\mathcal{U}^{\mathrm{w}}$ of the collection of all metric measure spaces $\mathcal{M}^{\mathrm{w}}$. In analogy with the notion of the ultrametric Gromov–Hausdorff distance on the collection of ultrametric spaces $\mathcal{U}$, we define ultrametric versions of two metrics on $\mathcal{U}^{\mathrm{w}}$, namely of Sturm's Gromov–Wasserstein distance of order p and of the Gromov–Wasserstein distance of order p. We study the basic topological and geometric properties of these distances as well as their relation and derive for $p = \infty$ a polynomial time algorithm for their calculation. Further, several lower bounds for both distances are derived and some of our results are generalized to the case of finite ultra-dissimilarity spaces. Finally, we study the relation between the Gromov–Wasserstein distance and its ultrametric version (as well as the relation between the corresponding lower bounds) in simulations and apply our findings for phylogenetic tree shape comparisons.

Editor in Charge: Kenneth Clarkson

F.M. and Z.W. acknowledge funding from the National Science Foundation under grants CCF 1740761, DMS 1723003, and RI 1901360. A.M. and C.W. gratefully acknowledge support by the DFG Research Training Group 2088, CRC 1456 project A04 and Cluster of Excellence MBExC 2067.

Facundo Mémoli
memoli@math.osu.edu

Axel Munk
munk@math.uni-goettingen.de

Zhengchao Wan
zcwan@ucsd.edu

Christoph Weitkamp
cweitka@mathematik.uni-goettingen.de

1 Department of Mathematics, The Ohio State University, Columbus, OH, USA

2 Institute for Mathematical Stochastics, University of Göttingen, Göttingen, Germany

3 Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA, USA

Published online: 07 October 2023

Keywords Ultrametric space · Gromov–Hausdorff distance · Gromov–Wasserstein distance · Optimal transport

Mathematics Subject Classification Primary: 51F99 · Second: 49Q22 · 53C23 · 53Z50

1 Introduction

Over the last decade the acquisition of ever more complex data, structures and shapes has increased dramatically. Consequently, the need to develop meaningful methods for comparing general objects has become more and more apparent. In numerous applications, e.g. in molecular biology [16, 40, 49], computer vision [42, 56] and electrical engineering [50, 70], it is important to distinguish between different objects in a pose invariant manner: two instances of a given object in *different* spatial orientations are deemed to be equal. Furthermore, also the comparisons of graphs, trees, ultrametric spaces and networks, where mainly the underlying connectivity structure matters, have grown in importance [20, 26]. One possibility to compare two general objects in a pose invariant manner is to model them as metric spaces (X, d_X) and (Y, d_Y) and regard them as elements of the collection of isometry classes of compact metric spaces denoted by $\mathcal{M}$ (i.e. two compact metric spaces (X, d_X) and (Y, d_Y) are in the same class if and only if they are isometric to each other which we denote by $X \cong Y$). It is possible to compare (X, d_X) and (Y, d_Y) via the *Gromov–Hausdorff distance* [29, 38], which is a metric on $\mathcal{M}$. It is defined as

$$d_{\mathrm{GH}}(X, Y) := \inf_{Z,\phi,\psi} d_{\mathrm{H}}^{(Z,d_Z)}(\phi(X), \psi(Y)), \tag{1}$$

where $\phi\colon X \to Z$ and $\psi\colon Y \to Z$ are isometric embeddings into a common metric space (Z, d_Z) and $d_{\mathrm{H}}^{(Z,d_Z)}$ denotes the *Hausdorff distance in* Z. The Hausdorff distance is a metric on the collection $\mathcal{S}(Z)$ of all compact subsets of a metric space (Z, d_Z) and, for $A, B \in \mathcal{S}(Z)$, is defined as follows:

$$d_{\mathrm{H}}^{(Z,d_Z)}(A, B) := \max\bigl(\sup_{a\in A}\inf_{b\in B} d_Z(a, b),\ \sup_{b\in B}\inf_{a\in A} d_Z(a, b)\bigr).$$

While the Gromov–Hausdorff distance has been applied successfully to various shape and data analysis tasks (see e.g. [11–15, 18, 19, 62]), it turns out that it is generally convenient to equip the modelled objects with additional structure rendering them as *metric measure spaces* [59, 60]. A metric measure space $\mathcal{X} = (X, d_X, \mu_X)$ is a triple, where (X, d_X) denotes a metric space and μ_X is a Borel probability measure on X with full support. This additional probability measure can be thought of as signalling the importance of different regions in the modelled object. Moreover, two metric measure spaces $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$ are considered as isomorphic (denoted by $\mathcal{X} \cong_{\mathrm{w}} \mathcal{Y}$) iff there exists an isometry $\varphi\colon (X, d_X) \to (Y, d_Y)$ such that $\varphi_\# \mu_X = \mu_Y$. Here, $\varphi_\#$ denotes the pushforward map induced by φ. From now on, $\mathcal{M}^{\mathrm{w}}$ denotes the collection of all (isomorphism classes of) compact metric measure spaces.

The metric measure space structure allows us to regard objects as probability measures instead of compact sets. Hence, it is possible to substitute the Hausdorff component in (1) by a relaxed notion of proximity, namely the *Wasserstein distance*. This distance is fundamental to a variety of mathematical developments and is also known as Kantorovich distance [44], Kantorovich–Rubinstein distance [45], Mallows distance [57] or as the Earth Mover's distance [78]. Given a compact metric space (Z, d_Z), let $\mathcal{P}(Z)$ denote the space of probability measures on Z and let $\alpha, \beta \in \mathcal{P}(Z)$. Then, the Wasserstein distance of order p, for $1 \leq p < \infty$, between α and β is defined as

$$d_{\mathrm{W},p}^{(Z,d_Z)}(\alpha, \beta) := \left(\inf_{\mu \in \mathcal{C}(\alpha,\beta)} \int_{Z \times Z} (d_Z(x, y))^p \, \mu(dx \times dy) \right)^{1/p}, \tag{2}$$

and for $p = \infty$ as

$$d_{\mathrm{W},\infty}^{(Z,d_Z)}(\alpha, \beta) := \inf_{\mu \in \mathcal{C}(\alpha,\beta)} \sup_{(x,y) \in \mathrm{supp}(\mu)} d_Z(x, y), \tag{3}$$

where $\mathrm{supp}(\mu)$ stands for the support of μ and $\mathcal{C}(\alpha, \beta)$ denotes the set of all couplings of α and β, i.e., the set of all probability measures μ on the product space $Z \times Z$ such that

$$\mu(A \times Z) = \alpha(A) \quad \text{and} \quad \mu(Z \times B) = \beta(B)$$

for all Borel measurable sets A and B of Z. It is worth noting that the Wasserstein distance between probability measures on the real line admits a closed form solution (see [90] and also Remark 2.10). We note that (2) and (3) can be unified into a more compact expression via L^p-norms:

$$d_{\mathrm{W},p}^{(Z,d_Z)}(\alpha, \beta) := \inf_{\mu \in \mathcal{C}(\alpha,\beta)} \|d_Z\|_{L^p(\mu)}, \quad 1 \leq p \leq \infty.$$

To simplify the presentation of our results in this paper, we will adopt this notation throughout what follows. To facilitate readers to understand our notation, we provide expanded version of important formulas in Section A of the Supplementary Material.

Sturm [83] has shown that replacing the Hausdorff distance in (1) with the Wasserstein distance yields a meaningful metric on $\mathcal{M}^{\mathrm{w}}$. Let $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$ be two metric measure spaces. Then, *Sturm's Gromov–Wasserstein distance* of order p, $1 \leq p \leq \infty$, is defined as

$$d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) := \inf_{Z,\phi,\psi} d_{\mathrm{W},p}^{(Z,d_Z)}(\phi_\# \mu_X, \psi_\# \mu_Y), \tag{4}$$

for isometric embeddings $\phi \colon X \to Z$, $\psi \colon Y \to Z$ into a metric space (Z, d_Z).

Based on similar ideas but starting from a different representation of the Gromov–Hausdorff distance, Mémoli [59, 60] derived a computationally more tractable and topologically equivalent metric on $\mathcal{M}^{\mathrm{w}}$, namely the *Gromov–Wasserstein* distance: For $1 \le p \le \infty$, the *p-distortion*[1] of a coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ is defined as

$$\mathrm{dis}_p(\mu) := \|d_X - d_Y\|_{L^p(\mu\otimes\mu)}.$$

The *Gromov–Wasserstein distance of order* p, $1 \le p \le \infty$, is defined as

$$d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) := \frac{1}{2} \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \mathrm{dis}_p(\mu). \tag{5}$$

It is known that in general $d_{\mathrm{GW},p} \le d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and that the inequality can be strict [60]. Although both $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \le p \le \infty$, are in general NP-hard to compute [60], it is possible to efficiently approximate $d_{\mathrm{GW},p}$ via conditional gradient descent [60, 72]. This has led to numerous applications and extensions of this distance [4, 17, 22, 79, 86].

In many cases, since the direct computation of either of these distances can be onerous, the determination of the degree of similarity between two datasets is performed via firstly computing *invariant features* out of each dataset (e.g. global distance distributions [68]) and secondly by suitably comparing these features. This point of view has motivated the exploration of inverse problems arising from the study of such features [10, 60, 61, 84].

Clearly, $\mathcal{M}^{\mathrm{w}}$ contains various, extremely general spaces. However, in many applications one has prior knowledge about the metric measure spaces under consideration and it is often reasonable to restrict oneself to work on specific sub-collections $\mathcal{O}^{\mathrm{w}} \subseteq \mathcal{M}^{\mathrm{w}}$. For instance, it could be known that the metrics of the spaces considered are induced by the shortest path metric on some underlying trees and hence it is unnecessary to consider the calculation of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \le p \le \infty$, for all of $\mathcal{M}^{\mathrm{w}}$. The potential advantages of focusing on a specific sub-collection $\mathcal{O}^{\mathrm{w}}$ are twofold. On the one hand, it might be possible to use the features of $\mathcal{O}^{\mathrm{w}}$ to gain computational benefits. On the other hand, it might be possible to refine the definition $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \le p \le \infty$, to obtain more informative comparisons on $\mathcal{O}^{\mathrm{w}}$. Naturally, it is of interest to identify and study these subclasses and the corresponding refinements. This approach has been pursued to study (variants of) the Gromov–Hausdorff distance on compact *ultrametric spaces* by Zarichnyi [93] and Qiu [73], and on compact *p-metric spaces* by Mémoli and Wan [64]. Here, the metric space (X, d_X) is called a p-metric space ($1 \le p < \infty$), if for all $x, x', x'' \in X$ it holds $d_X(x, x'') \le (d_X(x, x')^p + d_X(x', x'')^p)^{1/p}$. Further, the metric space (X, u_X) is called an ultrametric space, if u_X fulfills the *strong triangle inequality*, i.e., it holds for all $x, x', x'' \in X$ that

$$u_X(x', x'') \le \max(u_X(x, x'), u_X(x', x'')). \tag{6}$$

[1] The term "p-distortion" is not used in [59, 60]. However, the quantity $\mathrm{dis}_p(\mu)$ is introduced as $J_p(\mu)$ in both references.

In particular, Mémoli et al. [63] derived a polynomial time algorithm for computing the *ultrametric Gromov–Hausdorff* distance u_{GH} between two finite ultrametric spaces (X, u_X) and (Y, u_Y) (see Sect. 2.2) defined as

$$u_{\mathrm{GH}}(X, Y) := \inf_{Z,\phi,\psi} d_{\mathrm{H}}^{(Z,u_Z)}(\phi(X), \psi(Y)), \tag{7}$$

where $\phi\colon X \to Z$ and $\psi\colon Y \to Z$ are isometric embeddings into a common *ultrametric* space (Z, u_Z) and $d_{\mathrm{H}}^{(Z,u_Z)}$ denotes the Hausdorff distance on Z.

A further motivation to study (surrogates of) the distances $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$ restricted on a subset $\mathcal{O}^{\mathrm{w}}$ comes from the idea of *slicing* which originated as a method to efficiently estimate the Wasserstein distance $d_{\mathrm{W},p}^{\mathbb{R}^d}(\alpha, \beta)$ between probability measures α and β supported in a high dimensional Euclidean space $\mathbb{R}^d$ [78]. The original idea is that given any line ℓ in $\mathbb{R}^d$ one first obtains α_ℓ and β_ℓ, the respective pushforwards of α and β under the orthogonal projection map $\pi_\ell\colon \mathbb{R}^d \to \ell$, and then one invokes the explicit formula for the Wasserstein distance for probability measures on $\mathbb{R}$ (see Remark 2.10) to obtain a lower bound to $d_{\mathrm{W},p}^{\mathbb{R}^d}(\alpha, \beta)$ without incurring the possibly high computational cost associated to solving an optimal transportation problem. This lower bound is improved via repeated (often random) selections of the line ℓ [8, 48, 78].

Recently, Le et al. [54] pointed out that, thanks to the fact that the 1-Wasserstein distance also admits an explicit formula when the underlying metric space is a tree [25, 31, 58], one can also devise *tree slicing* estimates of the distance between two given probability measures by suitably projecting them onto tree-like structures. Most likely, the same strategy is successful for suitable projections on random ultrametric spaces, as on these there is also an explicit formula for the Wasserstein distance [46]. The same line of work has also recently explored in the Gromov–Wasserstein scenario [53, 89] and could be extended based on efficiently computable restrictions (or surrogates) of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$. Inspired by the results of Mémoli and Wan [64] and Mémoli [63] on the ultrametric Gromov–Hausdorff distance as well as the results of Kloeckner [46], who derived an explicit representation of the Wasserstein distance on ultrametric spaces, we study the collection of compact *ultrametric measure spaces* $\mathcal{U}^{\mathrm{w}} \subseteq \mathcal{M}^{\mathrm{w}}$, where $\mathcal{X} = (X, u_X, \mu_X) \in \mathcal{U}^{\mathrm{w}}$, whenever the underlying metric space (X, u_X) is a compact ultrametric space.

In terms of applications, ultrametric spaces (and thus also ultrametric *measure* spaces) arise naturally in statistics as metric encodings of dendrograms [18, 43] which is a graph theoretical representation of ultrametric spaces, in the context of phylogenetic trees [82], in theoretical computer science in the probabilistic approximation of finite metric spaces [5, 32], and in physics in the context of a mean-field theory of spin glasses [65, 74].

Especially for phylogenetic trees (and dendrograms), where one tries to characterize the structure of an underlying evolutionary process or the difference between two such processes, it is important to have a meaningful method of comparison, i.e., a meaningful metric on $\mathcal{U}^{\mathrm{w}}$. However, it is evident from the definition of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and its relationship with $d_{\mathrm{GW},p}$ (see [60]), that the ultrametric structure of $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ is not taken into account in the computation of either $d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$ or $d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$,

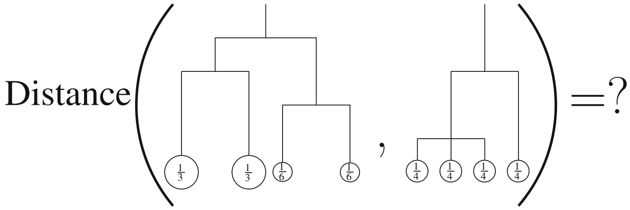

Fig. 1 Comparison between ultrametric measure spaces. Our objective is to devise methods for comparing ultrametric measure spaces that take into account their unique structure, as represented by dendrograms with weights on each of their leaves

$1 \leq p \leq \infty$. Hence, we suggest, just as for the ultrametric Gromov–Hausdorff distance, to adapt the definition of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see (4)) as well as the one of $d_{\mathrm{GW},p}$ (see (5)) and verify that this makes the comparisons of ultrametric measure spaces more sensitive and for $p = \infty$ leads to a *polynomial time* algorithm for the derivation of the proposed metrics.

1.1 The Proposed Approach

Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be ultrametric measure spaces. We aim to define meaningful distances for comparing them (see Fig. 1).

Reconsidering the definition of Sturm's Gromov–Wasserstein distance in (4), we propose to only infimize over ultrametric spaces (Z, u_Z). Thus, we define for $p \in [1, \infty]$ *Sturm's ultrametric Gromov–Wasserstein distance* of order p as

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) := \inf_{Z,\phi,\psi} d_{\mathrm{W},p}^{(Z,u_Z)}(\phi_\# \mu_X, \psi_\# \mu_Y), \tag{8}$$

where $\phi\colon X \to Z$, $\psi\colon Y \to Z$ are isometric embeddings into an ultrametric space (Z, u_Z).

In subsequent sections of this paper, we will establish several theoretically appealing properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$. Unfortunately, we will verify that, although an explicit formula for the Wasserstein distance of order p on ultrametric spaces exists [46], for $p \in [1, \infty)$ the calculation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ yields a highly non-trivial combinatorial optimization problem (see Sect. 3.1.1). Therefore, we demonstrate that an adaptation of the Gromov–Wasserstein distance defined in (5) yields a topologically equivalent and easily approximable distance on $\mathcal{U}^{\mathrm{w}}$. In order to define this adaption, we need to introduce some notation. For $a, b \geq 0$ and $1 \leq q < \infty$ let $\Lambda_q(a, b) := |a^q - b^q|^{1/q}$. Further define $\Lambda_\infty(a, b) := \max(a, b)$ whenever $a \neq b$ and $\Lambda_\infty(a, b) = 0$ if $a = b$.

Now, we can rewrite $d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, as follows:

$$d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \|\Lambda_1(d_X, d_Y)\|_{L^p(\mu \otimes \mu)}. \tag{9}$$

Considering the derivation of $d_{\mathrm{GW},p}$ in [60] and the results on the closely related ultrametric Gromov–Hausdorff distance studied in [64] and [63], this suggests replacing

Λ_1 in (9) with Λ_∞ in order to incorporate the ultrametric structures of (X, u_X, μ_X) and (Y, u_Y, μ_Y) into the comparison. Hence, we define the *p-ultra-distortion* of a coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ for $1 \leq p \leq \infty$ as

$$\mathrm{dis}_p^{\mathrm{ult}}(\mu) := \|\Lambda_\infty(u_X, u_Y)\|_{L^p(\mu\otimes\mu)}. \tag{10}$$

The *ultrametric Gromov–Wasserstein distance* of order $p \in [1, \infty]$, is given as

$$u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) := \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \mathrm{dis}_p^{\mathrm{ult}}(\mu). \tag{11}$$

Due to the structural similarity between $d_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}$, we expect that many properties of $d_{\mathrm{GW},p}$ extend to $u_{\mathrm{GW},p}$. In particular, we will establish that $u_{\mathrm{GW},p}$ can be approximated[2] via conditional gradient descent and also admits several polynomial time computable lower bounds which are useful in applications.

It is worth mentioning that Sturm [84] studied the family of so-called $L^{p,q}$-distortion distances similar to our construction of $u_{\mathrm{GW},p}$. In our language, for any $p, q \in [1, \infty)$, the $L^{p,q}$-distortion distance is constructed by infimizing over the (p, q)-distortion defined by replacing Λ_∞ with $(\Lambda_q)^q$ in (10). This distance shares many properties with $d_{\mathrm{GW},p}$.

1.2 Overview of Our Results

Section 2. We generalize the results of [18] on the relation between ultrametric spaces and dendrograms and establish a bijection between compact ultrametric spaces and *proper dendrograms* (see Definition 2.1). After recalling some results on the ultrametric Gromov–Hausdorff distance (see (7)), we use the connection between compact ultrametric spaces and dendrograms to reformulate the expression of the p-Wasserstein distance ($1 \leq p < \infty$) on ultrametric spaces derived by [46] in terms of proper dendrograms. This allows us to derive a formulation of the ∞-Wasserstein distance on ultrametric spaces and to study the Wasserstein distance on compact subspaces of the ultrametric space $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$, which will be relevant when studying lower bounds of $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$.

Section 3. We demonstrate that $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \leq p \leq \infty$, are p-metrics on the collection of ultrametric measure spaces $\mathcal{U}^{\mathrm{w}}$. We derive several alternative representations for $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and study the relation between the metrics $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$. In particular, we show that, while for $1 \leq p < \infty$ it holds in general that $u_{\mathrm{GW},p} \leq 2^{1/p} u_{\mathrm{GW},p}^{\mathrm{sturm}}$, both metrics coincide for $p = \infty$, i.e., $u_{\mathrm{GW},\infty} = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$. Furthermore, we show that an alternative representation of $u_{\mathrm{GW},\infty}$ leads to a *polynomial time algorithm* for the calculation of $u_{\mathrm{GW},\infty}$ (as well as $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$). Moreover, we study the topological properties of $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ and $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$, $1 \leq p \leq \infty$. Most importantly, we show that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ induce the same topology on $\mathcal{U}^{\mathrm{w}}$ which is also different from

[2] Here "approximation" is meant in the sense that one can write code which will locally minimize the functional. There are in general no theoretical guarantees that these algorithms will converge to a global minimum.

the one induced by $d_{\mathrm{GW},p}^{\mathrm{sturm}}/d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$. More precisely, the topology induced by $u_{\mathrm{GW},p}$ (resp. $u_{\mathrm{GW},p}^{\mathrm{sturm}}$) on $\mathcal{U}^{\mathrm{w}}$ is much finer than that induced by $d_{\mathrm{GW},p}$ (resp. $d_{\mathrm{GW},p}^{\mathrm{sturm}}$). As we will show in Sect. 5, the distance $u_{\mathrm{GW},p}$ is more sensitive to certain differences of the ultrametric measure spaces considered than $d_{\mathrm{GW},p}$. While we further prove that the metric spaces $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ and $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$, $1 \leq p < \infty$, are neither complete nor separable metric space, we demonstrate that the ultrametric space $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty}^{\mathrm{sturm}})$, which coincides with $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty})$, is complete. Finally, we establish that $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$ is a geodesic space.

Section 4. It seems impossible to derive a polynomial time algorithm for the calculation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$, $1 \leq p < \infty$. Consequently, based on easily computable invariant features, we derive several polynomial time computable lower bounds for $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$. Due to the structural similarity between $d_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}$, these are in a certain sense analogous to those derived in [59, 60] for $d_{\mathrm{GW},p}$. Among other things, we show that

$$u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) := \inf_{\gamma \in \mathcal{C}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)} \|\Lambda_\infty(u_X, u_Y)\|_{L^p(\gamma)}.$$

We verify that the lower bound $\mathbf{SLB}_p^{\mathrm{ult}}$ can be reformulated in terms of the Wasserstein distance on the ultrametric space $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ (we derive an explicit formula for $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}$ in Sect. 2.3). This allows us to efficiently compute $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$ in $O(\max(|X|, |Y|)^2)$ steps.

Section 5. We illustrate the behavior and relation between $u_{\mathrm{GW},1}$ (which can be approximated via conditional gradient descent) and $\mathbf{SLB}_1^{\mathrm{ult}}$ in a set of examples. We also carefully illustrate the differences between $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$, and $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ (see Sect. 4 for a definition), respectively.

1.3 Related Work

In order to better contextualize our contribution, we now describe related work, both in applied and computational geometry, and in phylogenetics (where notions of distance between trees have arisen naturally).

Metrics between trees: the phylogenetics perspective. In phylogenetics, where one chief objective is to infer the evolutionary relationship between species via methods that evaluate observable traits (such as DNA sequences), the need to be able to measure dissimilarity between different trees arises from the fact that the process of reconstruction of a phylogenetic tree may depend on the set of genes being considered. At the same time, even for the same set of genes, different reconstruction methods could be applied which would result in different trees. As such, this has led to the development of many different metrics for measuring distance between phylogenetic trees. Examples include the Robinson–Foulds metric [77], the subtree-prune and regraft distance [39], and the nearest-neighbor interchange distance [76].

As pointed out in Owen and Provan [69], many of these distances tend to quantify differences between tree topologies and often do not take into account edge lengths. A certain phylogenetic tree metric space which encodes for edge lengths was proposed in Billera et al. [6] and studied algorithmically in [69]. This tree space assumes that all trees have the same set of taxa. An extension to the case of trees over different underlying sets is given in Grindstaff and Owen [37]. Lafond et al. [51] considered one type of metrics on possibly *multilabeled* phylogenetic trees with a fixed number of leafs. As the authors pointed out, a multilabeled phylogenetic tree in which no leafs are repeated is just a standard phylogenetic tree, whereas a multilabeled phylogenetic tree in which all labels are equal defines a *tree shape*. The authors then proceeded to study the computational complexity associated to generalizations of some of the usual metrics for phylogenetic trees (such as the Robinson–Foulds distance) to the multilabeled case. Colijn and Plazzotta [23] studied a metric between (binary) phylogenetic tree shapes based on a bottom to top enumeration of specific connectivity structures. The authors applied their metric to compare evolutionary trees based on the HA protein sequences from human influenza collected in different regions.

Metrics between trees: the applied geometry perspective. From a different perspective, ideas from applied geometry and applied and computational topology have been applied to the comparison of tree shapes in applications in probability, clustering and applied and computational topology.

Metric trees are also considered in probability theory in the study of models for random trees together with the need to quantify their distance; Evans [30] described some variants of the Gromov–Hausdorff distance between metric trees. See also Greven et al. [36] for the case of metric measure space representations of trees and a certain Gromov–Prokhorov type of metric on the collection thereof.

Trees, in the form of dendrograms, are abundant in the realm of hierarchical clustering methods. In their study of the *stability* of hierarchical clustering methods, Carlsson and Mémoli [18] utilized the Gromov–Hausdorff distance between the ultrametric representation of dendrograms. Schmiedl [80] proved that computing the Gromov–Hausdorff distance between tree metric spaces is NP-hard. Liebscher [55] suggested some variants of the Gromov–Hausdorff distance that are applicable in the context of phylogenetic trees. As mentioned before, Zarichnyi [93] introduced the ultrametric Gromov–Hausdorff distance u_{GH} between compact ultrametric spaces (a special type of tree metric spaces). Certain theoretical properties such as precompactness of u_{GH} have been studied in Qiu [73]. In contrast with the NP-hardness of computing d_{GH}, Mémoli et al. [63] devised a polynomial time algorithm for computing u_{GH}.

In computational topology *merge trees* arise through the study of the sublevel sets of a given function [1, 75] with the goal of shape simplification. Morozov et al. [66] developed the notion of *interleaving distance* between merge trees which is related to the Gromov–Hausdorff distance between trees through bi-Lipschitz bounds. In Agarwal et al. [2], exploiting the connection between the interleaving distance and the Gromov–Hausdorff between metric trees, the authors approached the computation of the Gromov–Hausdorff distance between metric trees in general and provide certain approximation algorithms. Touli and Wang [87] devised fixed-parameter tractable (FPT) algorithms for computing the interleaving distance between metric trees. One can imply from their methods an FPT algorithm to compute a 2-approximation of the Gromov–Hausdorff distance between ultrametric spaces. Mémoli et al. [63] devised an FPT algorithm for computing the exact value of the Gromov–Hausdorff distances between ultrametric spaces.

2 Preliminaries

In this section we briefly summarize the basic notions and concepts required throughout the paper.

2.1 Ultrametric Spaces and Dendrograms

It is well known that ultrametric spaces possess tree-like structures. In particular, it was established in [18] that finite ultrametric spaces are equivalent to the so-called *dendrograms*. In this way, we generalize this equivalence to the case of *compact* ultrametric spaces.

We first introduce some definitions and some notation. Given a set X, a *partition* of X is a set $P_X = \{X_i\}_{i\in I}$ where I is any index set, $\emptyset \neq X_i \subseteq X$, $X_i \cap X_j = \emptyset$ for all $i \neq j \in I$ and $\bigcup_{i\in I} X_i = X$. We call each element X_i a *block* of the given partition P_X and denote by $\mathbf{Part}(X)$ the collection of all partitions of X. For two partitions P_X and P'_X we say that P_X is *finer* than P'_X, if for every block $X_i \in P_X$ there exists a block $X'_j \in P'_X$ such that $X_i \subseteq X'_j$.

Definition 2.1 (*Proper dendrogram*) Given a set X (not necessarily finite), a *proper dendrogram* $\theta_X : [0,\infty) \to \mathbf{Part}(X)$ is a map satisfying the following conditions:

(i) $\theta_X(s)$ is finer than $\theta_X(t)$ for any $0 \leq s < t < \infty$.
(ii) $\theta_X(0)$ is the finest partition consisting only singleton sets.
(iii) There exists $T > 0$ such that for any $t \geq T$, $\theta_X(t) = \{X\}$ is the trivial partition.
(iv) For each $t > 0$, there exists $\varepsilon > 0$ such that $\theta_X(t) = \theta_X(t')$ for all $t' \in [t, t+\varepsilon]$.
(v) For any distinct points $x, x' \in X$, there exists $T_{xx'} > 0$ such that x and x' belong to different blocks in $\theta_X(T_{xx'})$.
(vi) For each $t > 0$, $\theta_X(t)$ consists of only finitely many blocks.
(vii) Let $\{t_n\}_{n\in\mathbb{N}}$ be a decreasing sequence such that $\lim_{n\to\infty} t_n = 0$ and let $X_n \in \theta_X(t_n)$. If for any $1 \leq n < m$, $X_m \subseteq X_n$, then $\bigcap_{n\in\mathbb{N}} X_n \neq \emptyset$.

When X is finite, a function $\theta_X : [0,\infty) \to \mathbf{Part}(X)$ satisfying conditions (i) to (iv) will satisfy conditions (v), (vi) and (vii) automatically, and thus a proper dendrogram reduces to the usual dendrogram (see [18, Sect. 3.1] for a formal definition). Let θ_X be a proper dendrogram over a set X. For any $x \in X$ and $t \geq 0$, we denote by $[x]_t^X$ the block in $\theta(t)$ that contains $x \in X$ and abbreviate $[x]_t^X$ to $[x]_t$ when the underlying set X is clear from the context. Similarly to Carlsson and Mémoli [18], who considered the relation between finite ultrametric spaces and dendrograms, we will prove that there is a bijection between compact ultrametric spaces and proper dendrograms. In particular, one can show that the subsequent theorem generalizes [18, Thm. 9]. Since its proof depends on several concepts not yet introduced, we postpone it to the proof of Theorem 2.2. We remark that compact ultrametric spaces have been also characterized via other terminology such as *synchronized rooted tree* in [46] and *comb metric space* in [52]. We chose to work with dendrogram as it provides a succinct and illustrative description of ultrametric spaces.

Theorem 2.2 *Given a set X, denote by $\mathcal{U}(X)$ the collection of all compact ultrametrics on X and $\mathcal{D}(X)$ the collection of all proper dendrograms over X. For any $\theta \in \mathcal{D}(X)$, consider u_θ defined as follows:*

$$\forall x, x' \in X, \ \ u_\theta(x, x') := \inf\big\{t \geq 0 \mid x, x' \text{ belong to the same block of } \theta(t)\big\}.$$

Then, $u_\theta \in \mathcal{U}(X)$ and the map $\Upsilon_X : \mathcal{D}(X) \to \mathcal{U}(X)$ sending θ to u_θ is bijective.

Remark 2.3 From now on, we denote by θ_X the proper dendrogram corresponding to a given compact ultrametric u_X on X under the bijection given above. Note that a block $[x]_t$ in $\theta_X(t)$ is actually the closed ball $B_t(x)$ in X centered at x with radius t. So for each $t \geq 0$, $\theta_X(t)$ partitions X into a union of several closed balls in X with respect to u_X.

2.2 The Ultrametric Gromov–Hausdorff Distance

Both $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, are by construction closely related to the Gromov–Hausdorff distance. In a recent paper, Mémoli et al. [63] studied an ultrametric version of this distance, namely the *ultrametric Gromov–Hausdorff distance* (denoted as u_{GH}). Since we will demonstrate several connections between $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, and this distance, we briefly summarize some of the results in [63, 64]. We start by recalling the formal definition of u_{GH}.

Definition 2.4 Let (X, u_X) and (Y, u_Y) be two compact ultrametric spaces. Then, the *ultrametric Gromov–Hausdorff* between X and Y is defined as

$$u_{\mathrm{GH}}(X, Y) = \inf_{Z,\phi,\psi} d_{\mathrm{H}}^{(Z,u_Z)}(\phi(X), \psi(Y)),$$

where $\phi\colon X \to Z$ and $\psi\colon Y \to Z$ are isometric embeddings (distance preserving transformations) into the ultrametric space (Z, u_Z).

Zarichnyi [93] has shown that u_{GH} is an ultrametric on the isometry classes of compact ultrametric spaces, which is denoted by $\mathcal{U}$, and Mémoli and Wan [64] identified a structural theorem (cf. Thm. 2.5) that gives rise to a polynomial time algorithm for u_{GH}. More precisely, it was proven in [64] that u_{GH} can be calculated via so-called *quotient ultrametric spaces*, which we define next. Let (X, u_X) be any ultrametric space and let $t \geq 0$. We define an equivalence relation $\sim_t$ on X as follows: $x \sim_t x'$ iff $u_X(x, x') \leq t$. We denote by $[x]_t^X$ (resp. $[x]_t$) the equivalence class of x under $\sim_t$ and by X_t the set of all such equivalence classes. In fact, $[x]_t^X = \{x' \in X \mid u(x, x') \leq t\}$ is exactly the closed ball centered at x with radius t and corresponds to a block in the corresponding proper dendrogram $\theta_X(t)$ (see Remark 2.3). Thus, one can think of X_t as a "set representation" of $\theta_X(t)$. We define $u_{X_t}: X_t \times X_t \to \mathbb{R}_{\geq 0}$ as follows:

$$u_{X_t}([x]_t, [x']_t) := \begin{cases} u_X(x, x'), & [x]_t \neq [x']_t, \\ 0, & [x]_t = [x']_t. \end{cases}$$

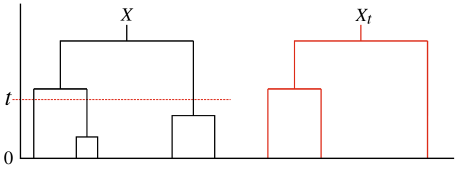

Fig. 2 Metric quotient: An ultrametric space (black) and its quotient at level t (red)

Then, (X_t, u_{X_t}) is an ultrametric space which we call the *quotient* of (X, u_X) at level t (see Fig. 2 for an illustration). It turns out that the quotient spaces characterize u_{GH} as follows.

Theorem 2.5 (Structural theorem for u_{GH}, [64, Thm. 5.7]) *Let (X, u_X) and (Y, u_Y) be two compact ultrametric spaces. Then,*

$$u_{\mathrm{GH}}(X, Y) = \inf\{t \geq 0 \mid X_t \cong Y_t\}.$$

Remark 2.6 Let (X, u_X) and (Y, u_Y) denote two finite ultrametric spaces and let $t \geq 0$. The quotient spaces X_t and Y_t can be considered as vertex weighted, rooted trees [63]. Hence, it is possible to check whether $X_t \cong Y_t$ in polynomial time [3]. Consequently, Theorem 2.5 induces a simple, polynomial time algorithm to calculate u_{GH} between two finite ultrametric spaces.

2.3 Wasserstein Distance on Ultrametric Spaces

Kloeckner [46] uses the representation of ultrametric spaces as so-called *synchronized rooted trees* to derive an explicit formula for the Wasserstein distance on ultrametric spaces. By the constructions of the proper dendrograms and of the synchronized rooted trees (see Sect. A.2.1), it is immediately clear how to reformulate the results of [46] on compact ultrametric spaces in terms of proper dendrograms. To this end, we need to introduce some notation. For a compact ultrametric space X, let θ_X be the associated proper dendrogram and let $V(X) := \bigcup_{t>0} \theta_X(t) = \{[x]_t \mid x \in X, t > 0\}$. $V(X)$ is in fact the collection of all closed balls in X except for singletons $\{x\}$ such that x is a cluster point[3] (see Lemma A.8). For $B \in V(X)$, we denote by B^* the smallest (under inclusion) element in $V(X)$ such that $B \subsetneq B^*$ (for the existence and uniqueness of B^* see Lemma A.1).

Theorem 2.7 (Wasserstein distance on ultrametric spaces, [46, Thm. 3.1]) *Let $X \in \mathcal{U}$. For all $\alpha, \beta \in \mathcal{P}(X)$ and $1 \leq p < \infty$, we have*

$$\left(d_{\mathrm{W},p}^X(\alpha,\beta)\right)^p = 2^{-1} \sum_{B \in V(X)\setminus\{X\}} \left(\mathrm{diam}\,(B^*)^p - \mathrm{diam}\,(B)^p\right) |\alpha(B) - \beta(B)|,$$

where $\mathrm{diam}\,(B)$ *denotes the diameter of the set B.*

[3] A cluster point x in a topological space X is such that any neighborhood of x contains countably many points in X.

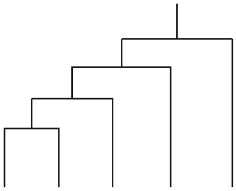

Fig. 3 **Illustration of** $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$**:** This is the dendrogram for a subspace of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ consisting of 5 arbitrary distinct points of $\mathbb{R}_+$

We extend Lemma 2.7 to the case $p = \infty$.

Lemma 2.8 *Let $X \in \mathcal{U}$. Then, for any $\alpha, \beta \in \mathcal{P}(X)$, we have*

$$d_{\mathrm{W},\infty}^X(\alpha, \beta) = \max_{\substack{B \in V(X)\setminus\{X\} \\ \alpha(B) \neq \beta(B)}} \operatorname{diam}(B^*). \tag{12}$$

The proof of Lemma 2.8 is technical and we postpone it to Sect. A.1.2.

2.3.1 Wasserstein Distance on $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$

The non-negative half real line $\mathbb{R}_{\geq 0}$ endowed with Λ_∞ turns out to be an ultrametric space (cf. [64, Example 2.7]). Finite subspaces of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ are of particular interest in this paper. These spaces possess a particular structure (see Fig. 3) and the computation of the Wasserstein distance on them can be further simplified.

Theorem 2.9 ($d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}$ between finitely supported measures) *Suppose α, β are two probability measures supported on a finite subset $\{x_0, \ldots, x_n\}$ of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ such that $0 \leq x_0 < x_1 < \cdots < x_n$. Denote $\alpha_i := \alpha(\{x_i\})$ and $\beta_i := \beta(\{x_i\})$. Then, we have for $p \in [1, \infty)$ that*

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}(\alpha, \beta) = 2^{-1/p}\left(\sum_{i=0}^{n-1}\left|\sum_{j=0}^{i}(\alpha_j - \beta_j)\right| \cdot |x_{i+1}^p - x_i^p| + \sum_{i=0}^{n} |\alpha_i - \beta_i| \cdot x_i^p\right)^{1/p}. \tag{13}$$

Let F_α and F_β denote the cumulative distribution functions of α and β, respectively. Then, for the case $p = \infty$ we obtain

$$d_{\mathrm{W},\infty}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}(\alpha, \beta) = \max\left(\max_{\substack{0 \leq i \leq n-1 \\ F_\alpha(x_i) \neq F_\beta(x_i)}} x_{i+1}, \max_{\substack{0 \leq i \leq n \\ \alpha_i \neq \beta_i}} x_i\right).$$

Proof Clearly, $V(X) = \{\{x_0, x_1, \ldots, x_i\} \mid i = 1, \ldots, n\} \cup \{\{x_i\} \mid i = 1, \ldots, n\}$ (recall that each set corresponds to a closed ball). Thus, we conclude the proof by applying Lemmas 2.7 and 2.8. □

Remark 2.10 (*Closed-form solution for* $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}$) As a closed-form solution for $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ is given by Theorem 2.9, we also note a classic closed-form solution for Wasserstein distance on $\mathbb{R}$ equipped with Euclidean distance Λ_1:

$$d_{\mathrm{W},p}^{(\mathbb{R},\Lambda_1)}(\alpha,\beta)=\left(\int_0^1 |F_\alpha^{-1}(t)-F_\beta^{-1}(t)|^p\,dt\right)^{1/p}, \tag{14}$$

where F_α and F_β are cumulative distribution functions of α and β, respectively. It turns out that closed-form solutions exist for more general $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}$, $q\in(1,\infty)$ and $q\leq p$, and we show more details in Sect. A.3.1.

Remark 2.11 (*Case* $p=1$) Note that when $p=1$, by combining (13) with (14), we obtain that for any finitely supported probability measures $\alpha,\beta\in\mathcal{P}(\mathbb{R}_{\geq 0})$,

$$\begin{aligned}d_{\mathrm{W},1}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha,\beta)&=\frac{1}{2}\left(d_{\mathrm{W},1}^{(\mathbb{R},\Lambda_1)}(\alpha,\beta)+\sum_{i=0}^n x_i\,|\alpha_i-\beta_i|\right)\\&=\frac{1}{2}\left(d_{\mathrm{W},1}^{(\mathbb{R},\Lambda_1)}(\alpha,\beta)+\int_{\mathbb{R}} x\,|\alpha-\beta|\,(dx)\right),\end{aligned}$$

where α_i, β_i and x_i are defined similarly as in Theorem 2.9 and we write the sum $\sum_{i=0}^n x_i\,|\alpha_i-\beta_i|$ into an integral for a succinct expression which requires no specification of the supports of the measures. The formula indicates that the 1-Wasserstein distance on $(\mathbb{R}_{\geq 0},\Lambda_\infty)$ is the average of the usual 1-Wasserstein distance on $(\mathbb{R}_{\geq 0},\Lambda_1)$ and a "weighted total variation distance". The weighted total variation like distance term is sensitive to difference of supports. For example, let $\alpha=\delta_{x_1}$ and $\beta=\delta_{x_2}$, then $\int_{\mathbb{R}} x\,|\alpha-\beta|\,(dx)=x_1+x_2$ if $x_1\neq x_2$.

Remark 2.12 (*Extension to compactly supported measures*) In fact, $X\subseteq(\mathbb{R}_{\geq 0},\Lambda_\infty)$ is compact if and only if it is either a finite set or a countable set containing zero and with zero being the unique cluster point (w.r.t. the usual Euclidean distance Λ_1) (see Lemma A.2). Hence, it is straightforward to extend Theorem 2.9 to compactly supported measures and we refer to Sect. A.3 for the missing details.

3 Ultrametric Gromov–Wasserstein Distances

In this section we investigate the properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ as well as $u_{\mathrm{GW},p}$, $1\leq p\leq\infty$, and study the relation between them.

3.1 Sturm's Ultrametric Gromov–Wasserstein Distance

We begin by establishing several basic properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1\leq p\leq\infty$, including a proof that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is indeed a metric (or more precisely a p-metric) on the collection of compact ultrametric measure spaces $\mathcal{U}^{\mathrm{w}}$.

The definition of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ given in (8) is clunky, technical and in general not easy to work with. Hence, the first observation to make is the fact that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \le p \le \infty$, shares a further property with $d_{\mathrm{GW},p}^{\mathrm{sturm}}$: $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ can be calculated by minimizing over pseudo-ultrametrics[4] instead of isometric embeddings.

Lemma 3.1 *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be two ultrametric measure spaces. Let $\mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ denote the collection of all pseudo-ultrametrics u on the disjoint union $X \sqcup Y$ such that $u|_{X\times X} = u_X$ and $u|_{Y\times Y} = u_Y$. Let $p \in [1, \infty]$. Then, it holds that*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = \inf_{u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)} d_{\mathrm{W},p}^{(X\sqcup Y,u)}(\mu_X, \mu_Y), \tag{15}$$

where $d_{\mathrm{W},p}^{(X\sqcup Y,u)}$ denotes the Wasserstein pseudometric *of order p defined in (28) (resp. in (29) for $p = \infty$) in Sect. B.5.1.*

Proof The claim follows by the same arguments as Lemma 3.3 (iii) in [83]. □

Remark 3.2 (*Wasserstein pseudometric*) The *Wasserstein pseudometric* is a natural extension of the Wasserstein distance to pseudometric spaces and has for example been studied in [85]. In Sect. B.5.1 we carefully show that it is closely related to the Wasserstein distance on a canonically induced metric space. We further establish that the Wasserstein distance and the Wasserstein pseudometric share many relevant properties. Hence, we do not notationally distinguish between these two concepts.

The representation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \le p \le \infty$, given by the above lemma is much more accessible and we first use it to establish the subsequent basic properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see Sect. B.1.1 for a full proof).

Proposition 3.3 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. Then, the following hold:*

(i) *For any $p \in [1, \infty]$, we always have that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) \ge d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$.*

(ii) *For any $1 \le p \le q \le \infty$, we have that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) \le u_{\mathrm{GW},q}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$.*

(iii) *It holds that $\lim_{p\to\infty} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$.*

We use Lemma 3.1 to prove that $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ is indeed a metric space.

Theorem 3.4 *$u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is a p-metric on the collection $\mathcal{U}^{\mathrm{w}}$ of compact ultrametric measure spaces. In particular, when $p = \infty$, $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$ is an ultrametric.*

In order to increase the readability of this section we postpone the proof of Theorem 3.4 to Sect. B.1.2. In the course of the proof, we will, among other things, verify the existence of optimal metrics and optimal couplings in (15) (see Proposition B.1). Furthermore, it is important to note that the topology induced on $\mathcal{U}^{\mathrm{w}}$ by $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \le p \le \infty$, is different from the one induced by $d_{\mathrm{GW},p}^{\mathrm{sturm}}$. This is well illustrated in the following example.

[4] A pseudo-ultrametric is a pseudometric which satisfies the strong triangle inequality (cf. (6)); see Sect. B.5.1 for the definition and further discussion on pseudometrics.

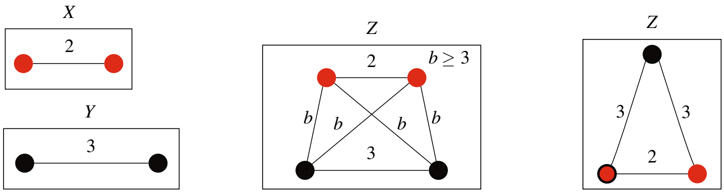

Fig. 4 Common ultrametric spaces: Representation of the two kinds of ultrametric spaces Z (middle and right) into which we can isometrically embed the spaces X and Y (left)

Example 3.5 ($u_{\mathrm{GW},p}^{\mathrm{sturm}}$ *and* $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ *induce different topologies*) This example is an adaptation from [64, Exam. 4.17]. For each $a > 0$, denote by $\Delta_2(a)$ the two-point metric space with interpoint distance a. Endow with $\Delta_2(a)$ the uniform probability measure μ_a and denote the corresponding ultrametric measure space $\widehat{\Delta}_2(a)$. Now, let $\mathcal{X} := \widehat{\Delta}_2(1)$ and let $\mathcal{X}_n := \widehat{\Delta}_2(1+1/n)$ for $n \in \mathbb{N}$. It is easy to check that for any $1 \le p \le \infty$, $d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{X}_n) = 1/(2n)$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{X}_n) = 2^{-1/p}(1+1/n)$ where we adopt the convention that $1/\infty = 0$. Hence, as n goes to infinity $\mathcal{X}_n$ will converge to $\mathcal{X}$ in the sense of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$, but not in the sense of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, for any $1 \le p \le \infty$.

3.1.1 Alternative Representations of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

In this subsection, we derive an alternative representation for $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ defined in (8). We mainly focus on the case $p < \infty$, however it turns out that the results also hold for $p = \infty$ (see Sect. 3.3).

Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ and recall the definition of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $p \in [1,\infty]$, given in (8), i.e.,

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = \inf_{Z,\phi,\psi} d_{\mathrm{W},p}^{(Z,u_Z)}(\varphi_\# \mu_Y, \psi_\# \mu_Y),$$

where $\phi\colon X \to Z$ and $\psi\colon Y \to Z$ are isometric embeddings into an ultrametric space (Z, u_Z). It turns out that we only need to consider relatively few possibilities of mapping two ultrametric spaces into a common ultrametric space. Exemplarily, this is shown in Fig. 4, where we see two finite ultrametric spaces and two possibilities for a common ultrametric space Z.

Indeed, it is straightforward to write down all reasonable embeddings and target spaces. We define the set

$$\mathcal{A} := \left\{ (A,\varphi) \,\middle|\, \begin{array}{l} \emptyset \neq A \subseteq X \text{ is closed and} \\ \varphi\colon A \hookrightarrow Y \text{ is an isometric embedding} \end{array} \right\}. \tag{16}$$

Clearly, $\mathcal{A} \neq \emptyset$, as it holds for each $x \in X$ that $\{(\{x\}, \varphi_y)\}_{y \in Y} \subseteq \mathcal{A}$, where φ_y is the map sending x to $y \in Y$. Another possibility to construct elements in $\mathcal{A}$ is illustrated in the subsequent example.

Example 3.6 Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ be finite spaces and let $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$. If $u^{-1}(0) \neq \emptyset$, we define $A := \pi_X(u^{-1}(0)) \subseteq X$, where $\pi_X\colon X \times Y \to X$ is the canonical projection.

Then, the map $\varphi\colon A \to Y$ defined by sending $x \in A$ to $y \in Y$ such that $u(x, y) = 0$ is an isometric embedding and $(A, \varphi) \in \mathcal{A}$.

Now, fix two compact spaces $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. Let $(A, \varphi) \in \mathcal{A}$ and let $Z_A = X \sqcup (Y\backslash\varphi(A)) \subseteq X \sqcup Y$. Furthermore, define $u_{Z_A}\colon Z_A \times Z_A \to \mathbb{R}_{\geq 0}$ as follows:

(i) $u_{Z_A}|_{X\times X} := u_X$ and $u_{Z_A}|_{Y\backslash\varphi(A)\times Y\backslash\varphi(A)} := u_Y|_{Y\backslash\varphi(A)\times Y\backslash\varphi(A)}$.
(ii) For any $x \in A$ and $y \in Y\backslash\varphi(A)$ define $u_{Z_A}(x, y) := u_Y(y, \varphi(x))$.
(iii) For $x \in X\backslash A$ and $y \in Y\backslash\varphi(A)$ let

$$u_{Z_A}(x, y) := \inf\{\max(u_X(x, a), u_Y(\varphi(a), y)) \mid a \in A\}.$$

(iv) For any $x \in X$ and $y \in Y\backslash\varphi(A)$, $u_{Z_A}(y, x) := u_{Z_A}(x, y)$.

Then, (Z_A, u_{Z_A}) is an ultrametric space such that X and Y can be mapped isometrically into Z_A (see [93, Lem. 1.1]). Let $\phi^X_{(A,\varphi)}$ and $\psi^Y_{(A,\varphi)}$ denote the corresponding isometric embeddings of X and Y, respectively. This allows us to derive the following statement, whose proof is postponed to Sect. B.1.3.

Theorem 3.7 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. Then, we have for each $p \in [1, \infty)$ that*

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \inf_{(A,\varphi)\in\mathcal{A}} d^{Z_A}_{\mathrm{W},p}\Big(\big(\phi^X_{(A,\varphi)}\big)_\# \mu_X, \big(\psi^Y_{(A,\varphi)}\big)_\# \mu_Y\Big). \tag{17}$$

Remark 3.8 Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ be finite spaces. The representation of $u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$, $1 \leq p \leq \infty$, given by Theorem 3.7 is very explicit and recasts the computation of $u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$, $1 \leq p \leq \infty$, as a combinatorial problem. In fact, the set $\mathcal{A}$ in (17) can be further reduced. More precisely, we demonstrate in Sect. B.1.3 (see Corollary B.7) that it is sufficient to infimize over the set of all *maximal pairs*, denoted by $\mathcal{A}^*$. Here, a pair $(A, \varphi_1) \in \mathcal{A}$ is denoted as *maximal*, if for all pairs $(B, \varphi_2) \in \mathcal{A}$ with $A \subseteq B$ and $\varphi_2|_A = \varphi_1$ it holds $A = B$. Using the ultrametric Gromov–Hausdorff distance (see (7)) it is possible to determine if two ultrametric spaces are isometric in polynomial time [63, Lem. 68]. However, this is clearly not sufficient to identify all $(A, \varphi) \in \mathcal{A}^*$ in polynomial time. Especially, for a given, viable $A \subseteq X$, there are usually multiple ways to define the corresponding map φ (see Example 3.9 right below this remark). Furthermore, for $1 \leq p < \infty$, we have neither been able to further restrict the set $\mathcal{A}^*$ nor to identify the optimal (A^*, φ^*). This just leaves a brute force approach which is computationally not feasible. On the other hand, for $p = \infty$ we are able to explicitly construct the optimal pair (A^*, φ^*) (see Theorem 3.23).

Example 3.9 Let $\{d_i\}_{i=1}^n$ be pairwise different real numbers with $0 \leq d_i < 1$. Let $X := \{x^i_j\}^{i=1,2}_{j=1,\dots,n}$ be a set with $2n$ points. Then, we define u_X as follows:

$$u_X(x^i_j, x^a_b) = \begin{cases} 0 & \text{if } i = a \text{ and } j = b; \\ d_i & \text{if } i \neq a \text{ and } j = b; \\ 1 & \text{if } j \neq b \end{cases}$$

u_X is obviously an ultrametric on X. Let $\{d_i^*\}_{i=1}^n$ be pairwise different real numbers with $0 \le d_i^* < 1$. Assume $d_i^* \ne d_j$ for all $i, j = 1, \ldots, n$. We similarly define an ultrametric space (Y, u_Y) w.r.t. $\{d_i^*\}_{i=1}^n$. Equip both spaces with some probability measures μ_X and μ_Y. Consider the minimization problem

$$\inf_{\{\varphi \,|\, (A,\varphi)\in\mathcal{A}^*\}} d_{\mathrm{W},p}^{Z_A}\Big(\big(\phi_{(A,\varphi)}^X\big)_\# \mu_X, \big(\psi_{(A,\varphi)}^Y\big)_\# \mu_Y\Big),$$

where we use the notation from Theorem 3.7 and Remark 3.8. If we let $A := \{x_j^1\}_{j=1,\ldots,n}$, then it is easy to see that $|\{\varphi \,|\, (A,\varphi)\in\mathcal{A}^*\}| = 2^n n!$, which suggests that it is not possible to solve the above minimization problem in polynomial time using a brute force approach.

3.2 The Ultrametric Gromov–Wasserstein Distance

In this section, we consider basic properties of $u_{\mathrm{GW},p}$ and prove the analogue of Theorem 3.4, i.e., we verify that also $u_{\mathrm{GW},p}$ is a p-metric, $1 \le p \le \infty$, on the collection of ultrametric measure spaces.

The subsequent proposition collects basic properties of $u_{\mathrm{GW},p}$ which are also shared by $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ (cf. Proposition 3.3). We refer to Sect. B.2.1 for its proof.

Proposition 3.10 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. Then, the following claims hold:*

(i) *For any $p \in [1, \infty]$, we always have that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \ge d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$.*
(ii) *For any $1 \le p \le q \le \infty$, it holds $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \le u_{\mathrm{GW},q}(\mathcal{X}, \mathcal{Y})$.*
(iii) *We have that $\lim_{p\to\infty} u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$.*

Next, we verify that $u_{\mathrm{GW},p}$ is indeed a metric on the collection of ultrametric measure spaces.

Theorem 3.11 *The ultrametric Gromov–Wasserstein distance $u_{\mathrm{GW},p}$ is a p-metric on the collection $\mathcal{U}^{\mathrm{w}}$ of compact ultrametric measure spaces. In particular, when $p = \infty$, $u_{\mathrm{GW},\infty}$ is an ultrametric.*

The full proof of Theorem 3.11, which is based on the existence of optimal couplings in (11) (see Proposition B.10), is postponed to Sect. B.2.2.

Remark 3.12 ($u_{\mathrm{GW},p}$ *and* $d_{\mathrm{GW},p}$ *induce different topologies*) Reconsidering Example 3.5, it is easy to verify that in this setting $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{X}_n) = 2^{-1/p}(1 + 1/n)$ while $d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{X}_n) = 1/(2^{1/p}n)$, $1 \le p \le \infty$. Hence, just like $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$, $u_{\mathrm{GW},p}$ and $d_{\mathrm{GW},p}$ induce different topologies on $\mathcal{U}^{\mathrm{w}}$. This result can also be obtained from Sect. 3.4 where we derive that $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ give rise to the same topology. Together with the fact that $u_{\mathrm{GW},p} \ge d_{\mathrm{GW},p}$, we know that $u_{\mathrm{GW},p}$ induces a finer topology than the one induced by $d_{\mathrm{GW},p}$. In this way, $u_{\mathrm{GW},p}$ is more sensitive to perturbations in ultrametric data sets. In particular, unlike the metric $d_{\mathrm{GW},p}$, $u_{\mathrm{GW},p}$ is able to differentiate between different types of perturbation. This point is further examined in Sect. 5.3 where we empirically show that $d_{\mathrm{GW},p}$ is indifferent to different types of noise whereas $u_{\mathrm{GW},p}$ is sensitive to 'large scale' perturbation.

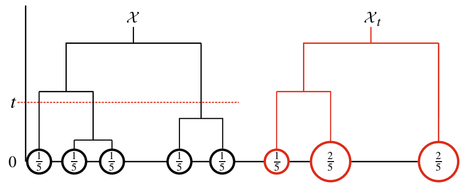

Fig. 5 Weighted Quotient: An ultrametric measure space (black) and its weighted quotient at level t (red)

Remark 3.13 As for the case of $d_{\mathrm{GW},p}$, $1 \le p < \infty$, [60, Sect. 7], it follows that for two finite ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ the computation of $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$, $1 \le p < \infty$, boils down to solving a (non-convex) quadratic program. This is in general NP-hard [71]. In contrast, for $p = \infty$, we will derive a polynomial time algorithm to determine $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$ (cf. Sect. 3.2.1).

3.2.1 Alternative Representations of $u_{\mathrm{GW},\infty}$

In this section, we will derive an alternative representation of $u_{\mathrm{GW},\infty}$ that resembles the one of u_{GH} derived in [64, Thm. 5.1]. It also leads to a polynomial time algorithm for the computation of $u_{\mathrm{GW},\infty}$. For this purpose, we define the *weighted quotient* of an ultrametric measure space. Let $\mathcal{X} = (X, u_X, \mu_X) \in \mathcal{U}^{\mathrm{w}}$ and let $t \ge 0$. Then, the *weighted quotient* of $\mathcal{X}$ at level t, is given as $\mathcal{X}_t = (X_t, u_{X_t}, \mu_{X_t})$, where (X_t, u_{X_t}) is the quotient of the ultrametric space (X, u_X) at level t (see Sect. 2.2) and $\mu_{X_t} \in \mathcal{P}(X_t)$ is the pushforward of μ_X under the canonical quotient map $Q_t : (X, u_X) \to (X_t, u_{X_t})$ sending x to $[x]_t$ for $x \in X$. Figure 5 illustrates the weighted quotient in a simple example.

Based on this definition, we show the following theorem, whose proof is postponed to Sect. B.2.3.

Theorem 3.14 *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be two compact ultrametric measure spaces. Then, it holds that*

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) = \min\{t \ge 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}. \tag{18}$$

Remark 3.15 The weighted quotients $\mathcal{X}_t$ and $\mathcal{Y}_t$ can be considered as vertex weighted, rooted trees and thus it is possible to verify whether $\mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t$ in polynomial time [3]. In consequence, we obtain a polynomial time algorithm for the calculation of $u_{\mathrm{GW},\infty}$. See p. 27 for the details.

The representations of u_{GH} in Theorem 2.5 and $u_{\mathrm{GW},\infty}$ in Theorem 3.14 strongly resemble themselves. As a direct consequence of both Theorems 2.5 and 3.14, we obtain the following comparison between the two metrics.

Corollary 3.16 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. Then, it holds that*

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) \ge u_{\mathrm{GH}}(X, Y). \tag{19}$$

The inequality (19) is sharp. Indeed, by [64, Cor. 5.3] we know that if the considered ultrametric spaces (X, u_X) and (Y, u_Y) have different diameters (w.l.o.g. $\operatorname{diam}(X) < \operatorname{diam}(Y)$), then $u_{\mathrm{GH}}(X, Y) = \operatorname{diam}(Y)$. The same statement also holds for $u_{\mathrm{GW},\infty}$

Corollary 3.17 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ be such that* $\operatorname{diam}(X) < \operatorname{diam}(Y)$. *Then,*

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) = \operatorname{diam}(Y) = u_{\mathrm{GH}}(X, Y).$$

Proof The first equality follows directly from [64, Cor. 5.3]. For the second equality, let $t := \operatorname{diam}(Y)$. It is obvious that $\mathcal{X}_t \cong_{\mathrm{w}} * \cong_{\mathrm{w}} \mathcal{Y}_t$, where $*$ denotes the one point ultrametric measure space. Let $s \in (\operatorname{diam}(X), \operatorname{diam}(Y))$, then $\mathcal{X}_t \cong_{\mathrm{w}} *$ whereas $\mathcal{Y} \ncong_{\mathrm{w}} *$. By Theorem 3.14, $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) = t = \operatorname{diam}(Y)$. □

3.3 The Relation Between $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

In this section, we study the relation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$, $1 \le p \le \infty$, and establish the topological equivalence between the two metrics.

3.3.1 Lipschitz Relation

We first study the Lipschitz relation between $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$. For this purpose, we have to distinguish the cases $p < \infty$ and $p = \infty$.

The case $p < \infty$. We start the consideration of this case by proving that it is essentially enough to consider the case $p = 1$ (see Theorem 3.18). To this end, we need to introduce some notation. For each $\alpha > 0$, we define a function $S_\alpha : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ by $x \mapsto x^\alpha$. Given an ultrametric space (X, u_X) and $\alpha > 0$, we abuse the notation and denote by $S_\alpha(X)$ the new space $(X, S_\alpha \circ u_X)$. It is obvious that $S_\alpha(X)$ is still an ultrametric space. This transformation of metric spaces is also known as the *snowflake transform* [24]. Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ denote two ultrametric measure spaces. Let $1 \le p < \infty$. We denote by $S_p(\mathcal{X})$ the ultrametric measure space $(X, S_p \circ u_X, \mu_X)$. The snowflake transform can be used to relate $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ as well as $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$ with $u_{\mathrm{GW},1}(S_p(\mathcal{X}), S_p(\mathcal{Y}))$ and $u_{\mathrm{GW},1}^{\mathrm{sturm}}(S_p(\mathcal{X}), S_p(\mathcal{Y}))$, respectively.

Theorem 3.18 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ and let $p \in [1, \infty)$. Then, we obtain*

$$\left(u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})\right)^p = u_{\mathrm{GW},1}(S_p(\mathcal{X}), S_p(\mathcal{Y})),$$
$$\left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})\right)^p = u_{\mathrm{GW},1}^{\mathrm{sturm}}(S_p(\mathcal{X}), S_p(\mathcal{Y})).$$

We give full proof of Theorem 3.18 in Sect. B.2.4. Based on this result, we can directly relate the metrics $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ by only considering the case $p = 1$ and prove the following Theorem 3.19 (see Sect. B.3.1 for its proof).

Theorem 3.19 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. Then, we have for $p \in [1, \infty)$ that*

$$u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \le 2^{1/p}\, u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}).$$

The subsequent example verifies that the coefficient in Theorem 3.19 is tight.

Example 3.20 For each $n \in \mathbb{N}$, let $\mathcal{X}_n$ be the three-point space $\Delta_3(1)$ (i.e. the 3-point space labeled by $\{x_1, x_2, x_3\}$ where all distances are 1) with a probability measure ${\mu_X}^n$ such that ${\mu_X}^n(x_1) = {\mu_X}^n(x_2) = 1/(2n)$ and ${\mu_X}^n(x_3) = 1 - 1/n$. Let $Y = *$ and μ_Y be the only probability measure on Y. Then, it is routine (using Proposition B.23 from Sect. B.5.3) to check that $u_{\mathrm{GW},1}(\mathcal{X}_n, \mathcal{Y}) = 2(1-3/(4n))/n$ and $u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X}_n, \mathcal{Y}) = 1/n$. Therefore, we have

$$\lim_{n\to\infty} \frac{u_{\mathrm{GW},1}(\mathcal{X}_n, \mathcal{Y})}{u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X}_n, \mathcal{Y})} = 2.$$

Example 3.21 ($u_{\mathrm{GW},p}^{\mathrm{sturm}}$ *and* $u_{\mathrm{GW},p}$ *are not bi-Lipschitz equivalent*) Following [60, Rem. 5.17], we verify in Sect. B.3.2 that for any positive integer n

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}\big(\widehat{\Delta}_n(1), \widehat{\Delta}_{2n}(1)\big) \geq \frac{1}{4} \quad \text{and} \quad u_{\mathrm{GW},p}\big(\widehat{\Delta}_n(1), \widehat{\Delta}_{2n}(1)\big) \leq \left(\frac{3}{2n}\right)^{1/p}.$$

Here, $\widehat{\Delta}_n(1)$ denotes the n-point metric measure space with interpoint distance 1 and the uniform probability measure. Thus, there exists no constant $C > 0$ such that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) \leq C \cdot u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ holds for every input spaces $\mathcal{X}$ and $\mathcal{Y}$. Hence, $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ are not bi-Lipschitz equivalent.

The case $p = \infty$. Next, we consider the relation between $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},\infty}$. By taking the limit $p \to \infty$ in Theorem 3.19, one might expect that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}} \geq u_{\mathrm{GW},\infty}$. In fact, we prove that the equality holds (see Sect. B.3.3).

Theorem 3.22 *Let* $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. *Then, it holds that*

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}).$$

One application of Theorem 3.22 is to explicitly derive the minimizing pair $(A, \phi) \in \mathcal{A}^*$ in (25) for $p = \infty$ (see Sect. B.3.4 for an explicit construction).

Theorem 3.23 *Let* $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$. *Let* $s := u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$ *and assume that* $s > 0$. *Then, there exists* $(A, \phi) \in \mathcal{A}$ *defined in (16) such that*

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y),$$

where Z_A *denotes the ultrametric space defined in Sect. 3.1.1.*

3.3.2 Topological Equivalence Between $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

Mémoli [60] proved the topological equivalence between $d_{\mathrm{GW},p}$ and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$. We establish an analogous result for $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$. To this end, we recall the *modulus of mass distribution*.

Definition 3.24 ([36, Defn. 2.9]) Given $\delta > 0$ we define *the modulus of mass distribution* of $\mathcal{X} \in \mathcal{U}^{\mathrm{w}}$ as

$$v_\delta(\mathcal{X}) := \inf\left\{\varepsilon > 0 \mid \mu_X(\{x : \mu_X(B^\circ_\varepsilon(x)) \le \delta\}) \le \varepsilon\right\},$$

where $B^\circ_\varepsilon(x)$ denotes the *open* ball centered at x with radius ε.

We note that $v_\delta(\mathcal{X})$ is non-decreasing, right-continuous and bounded above by 1. Furthermore, it holds that $\lim_{\delta \searrow 0} v_\delta(\mathcal{X}) = 0$ [36, Lem. 6.5]. With Definition 3.24 at hand, we derive the following theorem.

Theorem 3.25 *Let* $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$, $p \in [1, \infty)$ *and* $\delta \in (0, 1/2)$. *Then, whenever* $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) < \delta^5$ *we have*

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \le \left(4 \cdot \min(v_\delta(\mathcal{X}), v_\delta(\mathcal{Y})) + \delta\right)^{1/p} \cdot M,$$

where $M := 2 \cdot \max(\mathrm{diam}(X), \mathrm{diam}(Y)) + 54.$

Remark 3.26 Since it holds that $\lim_{\delta \searrow 0} v_\delta(\mathcal{X}) = 0$ and that $2^{-1/p} u^{\mathrm{sturm}}_{\mathrm{GW},p} \ge u_{\mathrm{GW},p}$ (see Theorem 3.19), the above theorem gives the topological equivalence between $u_{\mathrm{GW},p}$ and $u^{\mathrm{sturm}}_{\mathrm{GW},p}$, $1 \le p < \infty$ (the topological equivalence between $u^{\mathrm{sturm}}_{\mathrm{GW},\infty}$ and $u_{\mathrm{GW},\infty}$ holds trivially thanks to Theorem 3.22).

The proof of the Theorem 3.25 follows the same strategy used for proving [60, Prop. 5.3] and we refer to Sect. B.3.5 for the details.

3.4 Topological and Geodesic Properties

In this section, we consider the topology induced by $u_{\mathrm{GW},p}$ and $u^{\mathrm{sturm}}_{\mathrm{GW},p}$ on $\mathcal{U}^{\mathrm{w}}$ and discuss the geodesic properties of both $u_{\mathrm{GW},p}$ and $u^{\mathrm{sturm}}_{\mathrm{GW},p}$ for $1 \le p \le \infty$.
Completeness and separability. We derive the subsequent theorem whose proof is postponed to Sect. B.4.1.

Theorem 3.27

(i) *For* $p \in [1, \infty)$, $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$ *and* $(\mathcal{U}^{\mathrm{w}}, u^{\mathrm{sturm}}_{\mathrm{GW},p})$ *are neither complete nor separable.*
(ii) $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty}) = (\mathcal{U}^{\mathrm{w}}, u^{\mathrm{sturm}}_{\mathrm{GW},\infty})$ *is complete but not separable.*

Geodesic property. A *geodesic* in a metric space (X, d_X) is a continuous function $\gamma\colon [0, 1] \to X$ such that for each $s, t \in [0, 1]$, $d_X(\gamma(s), \gamma(t)) = |s - t| \cdot d_X(\gamma(0), \gamma(1))$. We say a metric space is geodesic if for any two distinct points $x, x' \in X$, there exists a geodesic $\gamma\colon [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = x'$. For any $p \in [1, \infty)$, the notion of p-geodesic is introduced in [64]: A p-geodesic in a metric space (X, d_X) is a continuous function $\gamma\colon [0, 1] \to X$ such that for each $s, t \in [0, 1]$, $d_X(\gamma(s), \gamma(t)) = |s - t|^{1/p} \cdot d_X(\gamma(0), \gamma(1))$. Similarly, we say a metric space is p-geodesic if for any two distinct points $x, x' \in X$, there exists a p-geodesic $\gamma\colon [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = x'$. Note that a 1-geodesic is a usual geodesic and a 1-geodesic space is a usual geodesic space. The subsequent theorem establishes (p-)geodesic properties of $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ for $p \in [1, \infty)$. A full proof is given in Sect. B.4.2.

Theorem 3.28 *For any $p \in [1, \infty)$, the space $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ is p-geodesic.*

Remark 3.29 Due to the fact that a p-geodesic space cannot be geodesic when $p > 1$ (cf. Lemma B.15), $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ is not geodesic for all $p > 1$.

Remark 3.30 Though the geodesic properties of $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$, $1 \le p < \infty$ are clear, we remark that geodesic properties of $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$, $1 \le p < \infty$, still remain unknown to us.

Remark 3.31 (*Case $p = \infty$*) Being an ultrametric space itself (cf. Theorem 3.11), $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty})$ $(= (\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty}^{\mathrm{sturm}}))$ is *totally disconnected*, i.e., any subspace with at least two elements is disconnected [81]. This in turn implies that each continuous curve in $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty})$ is constant. Therefore, $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty})$ is not a p-geodesic space for any $p \in [1, \infty)$.

4 Lower Bounds for $u_{\mathrm{GW},p}$

Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be two ultrametric measure spaces. The metrics $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ respect the ultrametric structure of the spaces $\mathcal{X}$ and $\mathcal{Y}$. Thus, one would hope that comparing ultrametric measure spaces with $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ or $u_{\mathrm{GW},p}$ is more meaningful than doing it with the usual Gromov–Wasserstein distance or Sturm's distance. Unfortunately, for $p < \infty$, the computation of both $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ is complicated and for $p = \infty$ both metrics are extremely sensitive to differences in the diameters of the considered spaces (see Corollary 3.17). Thus, it is not feasible to use these metrics in many applications. However, we can derive meaningful lower bounds for $u_{\mathrm{GW},p}$ (and hence also for $u_{\mathrm{GW},p}^{\mathrm{sturm}}$) that resemble those of the Gromov–Wasserstein distance. Naturally, the question arises whether these lower bounds are better/sharper than the ones of the usual Gromov–Wasserstein distance in this setting. This question is addressed throughout this section and will be readdressed in Sect. 5 as well as in the Supplementary Material.

In [60], the author introduced three lower bounds for $d_{\mathrm{GW},p}$ that are computationally less expensive than the calculation of $d_{\mathrm{GW},p}$. We will briefly review these three lower bounds and then define candidates for the corresponding lower bounds for $u_{\mathrm{GW},p}$. In the sequel, we always assume $p \in [1, \infty]$.

First lower bound. Let $s_{X,p} \colon X \to \mathbb{R}_{\ge 0}$, $x \mapsto \|u_X(x,\cdot)\|_{L^p(\mu_X)}$. Then, the first lower bound $\mathbf{FLB}_p(\mathcal{X}, \mathcal{Y})$ for $d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ is defined as follows:

$$\mathbf{FLB}_p(\mathcal{X}, \mathcal{Y}) := \frac{1}{2} \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \|\Lambda_1(s_{X,p}, s_{Y,p})\|_{L^p(\mu)}.$$

Following our intuition of replacing Λ_1 with Λ_∞, we define the ultrametric version of **FLB** as

$$\mathbf{FLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) := \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \|\Lambda_\infty(s_{X,p}, s_{Y,p})\|_{L^p(\mu)}.$$

Second lower bound. The second lower bound $\mathbf{SLB}_p(\mathcal{X},\mathcal{Y})$ for $d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$ is given as

$$\mathbf{SLB}_p(\mathcal{X},\mathcal{Y}) := \frac{1}{2}\inf_{\gamma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)} \|\Lambda_1(u_X, u_Y)\|_{L^p(\gamma)}.$$

Thus, we define the ultrametric second lower bound between two ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ as follows:

$$\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) := \inf_{\gamma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)} \|\Lambda_\infty(u_X, u_Y)\|_{L^p(\gamma)}.$$

Third lower bound. Before we introduce the final lower bound, we have to define several functions. First, let $\Gamma^1_{X,Y}\colon X\times Y\times X\times Y \to \mathbb{R}_{\geq 0}$, $(x,y,x',y')\mapsto \Lambda_1(u_X(x,x'), u_Y(y,y'))$ and let $\Omega^1_p\colon X\times Y\to\mathbb{R}_{\geq 0}$, $p\in[1,\infty]$, be given by

$$\Omega^1_p(x,y) := \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \|\Gamma^1_{X,Y}(x,y,\cdot,\cdot)\|_{L^p(\mu)}.$$

Then, the third lower bound $\mathbf{TLB}_p$ is given as

$$\mathbf{TLB}_p(\mathcal{X},\mathcal{Y}) := \frac{1}{2}\inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \|\Omega^1_p\|_{L^p(\mu)}.$$

Analogously to the definition of previous ultrametric versions, we define $\Gamma^\infty_{X,Y}\colon X\times Y\times X\times Y\to\mathbb{R}_{\geq 0}$, $(x,y,x',y')\mapsto\Lambda_\infty(u_X(x,x'),u_Y(y,y'))$. Further, for $p\in[1,\infty]$, let $\Omega^\infty_p\colon X\times Y\to\mathbb{R}_{\geq 0}$ be given by

$$\Omega^\infty_p(x,y) := \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \|\Gamma^\infty_{X,Y}(x,y,\cdot,\cdot)\|_{L^p(\mu)}.$$

Then, the ultrametric third lower bound between two ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ is defined as

$$\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) := \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \|\Omega^\infty_p\|_{L^p(\mu)}.$$

4.1 Properties and Computation of the Lower Bounds

Next, we examine the quantities $\mathbf{FLB}^{\mathrm{ult}}$, $\mathbf{SLB}^{\mathrm{ult}}$ and $\mathbf{TLB}^{\mathrm{ult}}$ more closely. Since $\Lambda_\infty(a,b)\geq\Lambda_1(a,b)=|a-b|$ for any $a,b\geq 0$, it is easy to conclude that

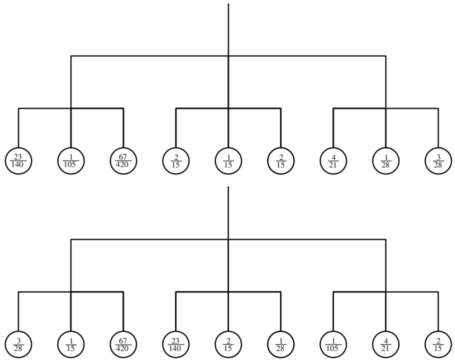

Fig. 6 $\mathbf{TLB}_p^{\mathrm{ult}}$ **and** $u_{\mathrm{GW},p}$ **induce different topologies.** For the two non-isomorphic ultrametric measure spaces depicted above, $\mathbf{TLB}_p^{\mathrm{ult}}$ equals 0 whereas $u_{\mathrm{GW},p} > 0$. This example is the same as the one given in [60, Fig. 8]

$\mathbf{FLB}_p^{\mathrm{ult}} \geq \mathbf{FLB}_p$, $\mathbf{SLB}_p^{\mathrm{ult}} \geq \mathbf{SLB}_p$ and $\mathbf{TLB}_p^{\mathrm{ult}} \geq \mathbf{TLB}_p$. Moreover, the three ultrametric lower bounds satisfy the following theorem (for a complete proof see Sect. C.1.1).

Theorem 4.1 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ and let $p \in [1, \infty]$.*

(i) $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{FLB}_\infty^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$.
(ii) $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$.

Remark 4.2 Interestingly, it turns out that $\mathbf{FLB}_p^{\mathrm{ult}}$ is not a lower bound of $u_{\mathrm{GW},p}$ in general when $p < \infty$. For example, let $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$ and define u_X such that $u_X(x_1, x_2) = 1$ and $u_X(x_i, x_j) = 2\delta_{i \neq j}$ for $(i, j) \neq (1, 2)$, $(i, j) \neq (2, 1)$ and $i, j = 1, \ldots, n$. Let $u_Y(y_i, y_j) = 2\delta_{i \neq j}$, $i, j = 1, \ldots, n$, and let μ_X and μ_Y be uniform measures on X and Y, respectively. Then, $u_{\mathrm{GW},1}(\mathcal{X}, \mathcal{Y}) \leq 4/n^2$ whereas $\mathbf{FLB}_1^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) = (4n - 4)/n^2$ which is greater than $u_{\mathrm{GW},1}(\mathcal{X}, \mathcal{Y})$ as long as $n > 2$. Moreover, we have in this case that $\mathbf{FLB}_1^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) = O(1/n)$ whereas $u_{\mathrm{GW},1}(\mathcal{X}, \mathcal{Y}) = O(1/n^2)$. Hence, there exists no constant $C > 0$ such that $\mathbf{FLB}_1^{\mathrm{ult}} \leq C \cdot u_{\mathrm{GW},1}$ in general.

Remark 4.3 There exist ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ such that $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$ equals 0 whereas $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) > 0$ (an example is given in [60, Fig. 8] and see Fig. 6 for an illustration). Furthermore, there are spaces $\mathcal{X}$ and $\mathcal{Y}$ such that $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) = 0$ whereas $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) > 0$ (see Sect. C.1.3). The analogous statement is true for $\mathbf{TLB}_p$ and $\mathbf{SLB}_p$, which are nevertheless useful in applications (see e.g. [34]).

From the structure of $\mathbf{SLB}_p^{\mathrm{ult}}$ and $\mathbf{TLB}_p^{\mathrm{ult}}$ it is obvious that their computation leads to different optimal transport problems (see e.g. [90]). However, in analogy to [21, Thm. 3.1] we can rewrite $\mathbf{SLB}_p^{\mathrm{ult}}$ and $\mathbf{TLB}_p^{\mathrm{ult}}$ in order to further simplify their computation. The full proof of the subsequent proposition is given in Sect. C.1.2.

Proposition 4.4 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ and let $p \in [1, \infty]$. Then, we find that*

(i) $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) = d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}(u_X)_\# ((\mu_X \otimes \mu_X), (u_Y)_\# (\mu_Y \otimes \mu_Y))$.

(ii) *For each* $x, y \in X \times Y$, $\Omega_p^\infty(x, y) = d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}(u_X(x, \cdot)_\# \mu_X, u_Y(y, \cdot)_\# \mu_Y)$.

Remark 4.5 Since Theorem 2.9 gives an explicit formula for the Wasserstein distance on $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ between finitely supported probability measures, these alternative representations of the lower bound $\mathbf{SLB}_p^{\mathrm{ult}}$ and the cost functional Ω_p^∞ drastically reduce the computation time of $\mathbf{SLB}_p^{\mathrm{ult}}$ and $\mathbf{TLB}_p^{\mathrm{ult}}$, respectively. In particular, we note that this allows us to compute $\mathbf{SLB}_p^{\mathrm{ult}}$, $1 \leq p \leq \infty$, between finite ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ in $O(\max(|X|, |Y|)^2)$ steps.

Proposition 4.4 allows us to directly compare $\mathbf{SLB}_1^{\mathrm{ult}}$ and $\mathbf{SLB}_1$.

Corollary 4.6 *For any finite ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$, we have that*

$$\begin{aligned} \mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) = \mathbf{SLB}_1(\mathcal{X}, \mathcal{Y}) \\ + \frac{1}{2} \int_{\mathbb{R}} t \left|(u_X)_\# (\mu_X \otimes \mu_X) - (u_Y)_\# (\mu_Y \otimes \mu_Y)\right| (dt). \end{aligned} \tag{20}$$

Proof The claim follows from Proposition 4.4 and Remark 2.11. □

This corollary implies that $\mathbf{SLB}_p^{\mathrm{ult}}$ is more rigid than $\mathbf{SLB}_p$, since the second summand on the right-hand side of (20) is sensitive to distance perturbations. This is also illustrated very well in the subsequent example.

Example 4.7 Recall notations from Example 3.5. For any $d, d' > 0$, we let $X := \Delta_2(d)$ and let $Y := \Delta_2(d')$. Assume that X and Y have underlying sets $\{x_1, x_2\}$ and $\{y_1, y_2\}$, respectively. Define $\mu_X \in \mathcal{P}(X)$ and $\mu_Y \in \mathcal{P}(Y)$ as follows. Let $\alpha_1, \alpha_2 \geq 0$ be such that $\alpha_1 + \alpha_2 = 1$. Let $\mu_X(x_1) = \mu_Y(y_1) := \alpha_1$ and let $\mu_X(x_2) = \mu_Y(y_2) := \alpha_2$. Then, it is easy to verify that

(i) $u_{\mathrm{GW},1}(\mathcal{X}, \mathcal{Y}) = \mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) = 2\alpha_1\alpha_2\Lambda_\infty(d, d')$.
(ii) $d_{\mathrm{GW},1}(\mathcal{X}, \mathcal{Y}) = \mathbf{SLB}_1(\mathcal{X}, \mathcal{Y}) = \alpha_1\alpha_2\Lambda_1(d, d') = \alpha_1\alpha_2|d - d'|$.
(iii) $\frac{1}{2} \int_{\mathbb{R}} t \left|(u_X)_\# (\mu_X \otimes \mu_X) - (u_Y)_\# (\mu_Y \otimes \mu_Y)\right| (dt) = \alpha_1\alpha_2(d + d')\,\delta_{d \neq d'}$.

From (i) and (ii) we observe that both second lower bounds are tight. Moreover, since we obviously have that $(d + d')\delta_{d \neq d'} + |d - d'| = 2\Lambda_\infty(d, d')$, we have also verified (20) through this example. Unlike $\mathbf{SLB}_1(\mathcal{X}, \mathcal{Y})$ being proportional to $|d - d'|$, as long as $d \neq d'$, even if $|d - d'|$ is small, $\Lambda_\infty(d, d') = \max(d, d')$ which results in a large value of $\mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$ when d and d' are large numbers. This example illustrates that $\mathbf{SLB}_1^{\mathrm{ult}}$ (and hence $u_{\mathrm{GW},1}$) is rigid with respect to distance perturbation.

5 Computational Aspects

In this section, we investigate algorithms for approximating/calculating $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$. Furthermore, we evaluate for $p < \infty$ the performance of the computationally efficient lower bound $\mathbf{SLB}_p^{\mathrm{ult}}$ introduced in Sect. 4 and compare our findings to the results of the classical Gromov–Wasserstein distance $d_{\mathrm{GW},p}$ (see (5)). Matlab implementations of the presented algorithms and comparisons are available at https://github.com/ndag/uGW.

5.1 Algorithms

Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be two finite ultrametric measure spaces with cardinalities m and n, respectively.

The case $p < \infty$. Recall Remark 3.13 which highlights that the exact calculation of $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ for $p < \infty$ is infeasible. However, in many practical applications it is sufficient to work with good approximations of this metric. Therefore, we propose to approximate (local minima of) $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ for $p < \infty$ via conditional gradient descent. To this end, we note that the gradient G that arises from (10) can in the present setting be expressed with the following partial derivative with respect to $\mu \in \mathcal{C}(\mu_X, \mu_Y)$

$$G_{i,j} = 2 \sum_{k=1}^{m} \sum_{l=1}^{n} \big(\Lambda_\infty(u_X(x_i, x_k), u_Y(y_j, y_l))\big)^p \mu_{kl}, \tag{21}$$

for all $1 \le i \le m$ and $1 \le j \le n$. As we deal with a non-convex minimization problem, the performance of the gradient descent strongly depends on the starting coupling $\mu^{(0)}$. Therefore, we follow the suggestion of Chowdhury and Needham [22] and employ a Markov Chain Monte Carlo Hit-And-Run sampler to obtain multiple random start couplings. Running the gradient descent from each point in this ensemble greatly improves the approximation in many cases. For a precise description of the proposed procedure, we refer to Algorithm 1.

Algorithm 1 $u_{\mathrm{GW},p}(X, Y, p, N, L)$

//Create a list of random couplings
couplings =CreateRandomCouplings(N)
stat_points = cell(N)
for i=1:N **do**
  $\mu^{(0)}$ =couplings$\{i\}$
  for j=1:L **do**
    G = Gradient from (21) w.r.t. $\mu^{(j-1)}$
    $\widetilde{\mu}^{(j)}$ = Solve OT with ground loss G
    $\gamma^{(j)} = \frac{2}{j+2}$ *//Alt. find* $\gamma \in [0,1]$ *that minimizes* $\mathrm{dis}_p^{\mathrm{ult}}\big(\mu^{(j-1)} + \gamma(\widetilde{\mu}^{(j)} - \mu^{(j-1)})\big)$
    $\mu^{(j)} = (1 - \gamma^{(j)})\mu^{(j-1)} + \gamma^{(j)}\widetilde{\mu}^{(j)}$
  end for
  stat_points$\{i\}$= $\mu^{(L)}$
end for
Find μ^* in stat_points that minimizes $\mathrm{dis}_p^{\mathrm{ult}}(\mu)$
result = $\mathrm{dis}_p^{\mathrm{ult}}(\mu^*)$

The case $p = \infty$. In what follows, we present the details of the polynomial time algorithm for the computation of $u_{\mathrm{GW},\infty}(= u_{\mathrm{GW},\infty}^{\mathrm{sturm}})$ hinted at in Remark 3.15. Let $\mathrm{spec}(X) := \{u_X(x, x') \mid x, x' \in X\}$ denote the spectrum of X. Then, it is evident that in order to find the minimum in (18), we only have to check $\mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t$ for each $t \in \mathrm{spec}(X) \cup \mathrm{spec}(Y)$, starting from the largest to the smallest and $u_{\mathrm{GW},\infty}$ is given as the smallest t such that $\mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t$. This can be done in polynomial time by considering

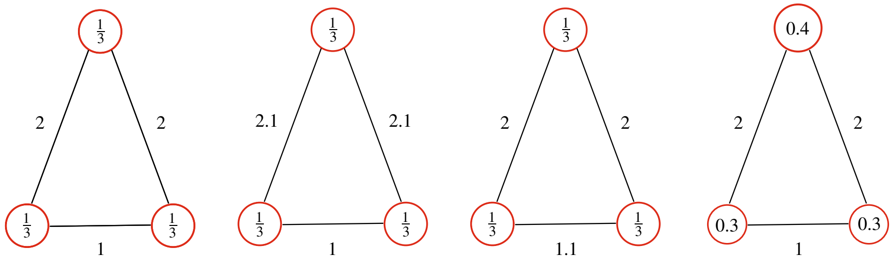

Fig. 7 Ultrametric measure spaces: Four non-isomorphic ultrametric measure spaces denoted (from left to right) as $\mathcal{X}_i = (X_i, d_{X_i}, \mu_{X_i})$, $1 \leq i \leq 4$

$\mathcal{X}_t$ and $\mathcal{Y}_t$ as weighted rooted trees and by solving a tree isomorphism problem (e.g., by using a slight modification of the algorithm in [3, Exam. 3.2]. Roughly, this algorithm assigns codes to any two given trees, respectively, by summarizing neighborhood information of vertices of the trees in a bottom-up manner, and then ascertains whether these two trees are isomorphic by comparing their codes.). This gives rise to a simple algorithm (see Algorithm 2) to calculate $u_{\mathrm{GW},\infty}$.[5]

Algorithm 2 $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$

```
spec = sort(spec(X) ∪ spec(Y), 'descent')
for i = 1: length(spec) do
   t = spec(i)
   if 𝒳_t ≇_w 𝒴_t then
      return spec(i − 1)
   end if
end for
return 0
```

5.2 The Relation Between $u_{\mathrm{GW},1}$, $u_{\mathrm{GW},\infty}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$

In order to understand how $u_{\mathrm{GW},p}$ (or at least its approximation), $u_{\mathrm{GW},\infty}$ and $\mathbf{SLB}_p^{\mathrm{ult}}$ are influenced by small structural changes of the considered ultrametric measure spaces, we exemplarily consider the ultrametric measure spaces $\mathcal{X}_i = (X_i, d_{X_i}, \mu_{X_i})$, $1 \leq i \leq 4$, displayed in Fig. 7. These differ only by one characteristic (e.g. one side length or the equipped measure). Exemplarily, we calculate $u_{\mathrm{GW},1}(\mathcal{X}_i, \mathcal{X}_j)$ (approximated with Algorithm 1, where $L = 5000$ and $N = 40$), $\mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{X}_i, \mathcal{X}_j)$ and $u_{\mathrm{GW},\infty}(\mathcal{X}_i, \mathcal{X}_j)$, $1 \leq i, j \leq 4$. In particular, note that we use Algorithm 1 to determine $u_{\mathrm{GW},1}(\mathcal{X}_i, \mathcal{X}_i)$, $1 \leq i \leq 4$. First of all, we observe that $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$ are influenced by the change in the diameter of the spaces the most and attain (up to differences of order 10^{-7}) the same value for the comparison of the spaces $\mathcal{X}_i$, $1 \leq i \leq 3$ (see Tables 1 and 2 in Sect. D.1 for the complete results). The picture changes for the comparisons of $\mathcal{X}_i$,

---

[5] The algorithm can be sped up via a binary search process which we do not include for simplicity of presentation.

$1 \le i \le 3$ with $\mathcal{X}_4$. Here, $\mathbf{SLB}_1^{\mathrm{ult}}$ attains significantly lower values than $u_{\mathrm{GW},1}$. While we cannot be completely sure that we approximate the global minima of $u_{\mathrm{GW},1}$ for these comparisons, the other results (and especially the approximation of $u_{\mathrm{GW},1}(\mathcal{X}_4, \mathcal{X}_4)$) imply that we should be reasonably close. All in all, this suggests that changes in metric influence $\mathbf{SLB}_1^{\mathrm{ult}}$ in a similar fashion as $u_{\mathrm{GW},1}$, while changes in the measure have less impact on $\mathbf{SLB}_1^{\mathrm{ult}}$. We can also conclude that the proposed algorithm for the approximation of $u_{\mathrm{GW},1}$ works reasonably well in this simple setting.

Further, we observe that $u_{\mathrm{GW},\infty}$ attains for almost all comparisons the maximal possible value. Only the comparison of $\mathcal{X}_1$ with $\mathcal{X}_3$, where the only small scale structure of the space was changed, yields a value that is smaller than the maximum of the diameters of the considered spaces.

5.3 Comparison of $u_{\mathrm{GW},1}$, $\mathbf{SLB}_1^{\mathrm{ult}}$, $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$

In the remainder of this section, we will demonstrate the differences between $u_{\mathrm{GW},1}$, $\mathbf{SLB}_1^{\mathrm{ult}}$, $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$. To this end, we first compare the metric measure spaces in Fig. 7 based on $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$. We observe that $d_{\mathrm{GW},1}$ (approximated in the same manner as $u_{\mathrm{GW},1}$) and $\mathbf{SLB}_1$ are hardly influenced by the differences between the ultrametric measure spaces $\mathcal{X}_i$, $1 \le i \le 4$. In particular, it is remarkable that $d_{\mathrm{GW},1}$ is affected the most by the changes made to the measure and not the metric structure (see Table 3 in Sect. D.2 for the complete results).

Next, we consider the differences between the aforementioned quantities more generally. For this purpose, we generate four ultrametric spaces Z_k, $1 \le k \le 4$, with totally different dendrogram structures, whose diameters are between 0.5 and 0.6 (for the precise construction of these spaces see Sect. D.2). For each $t = 0, 0.2, 0.4, 0.6$, we perturb each Z_k independently to generate 15 ultrametric spaces $Z_{k,t}^i$, $1 \le i \le 15$, such that $(Z_{k,t}^i)_t \equiv (Z_k)_t$ for all i. The spaces $Z_{k,t}^i$ are called *perturbations of Z_k at level t* (see Fig. 8 for an illustration and see Sect. D.2 for more details). The spaces $Z_{k,t}^i$ are endowed with the uniform probability measure and we obtain a collection of ultrametric measure spaces $\mathcal{Z}_{k,t}^i$. Naturally, we refer to k as the class of the ultrametric measure spaces $\mathcal{Z}_{k,t}^i$. We compute for each t the quantities $u_{\mathrm{GW},1}$ (approximated with Algorithm 1, $L = 100$, $N = 5$), $\mathbf{SLB}_1^{\mathrm{ult}}$, $d_{\mathrm{GW},1}$ (approximated with conditional gradient descent, 100 gradient steps) and $\mathbf{SLB}_1$ among the resulting 60 ultrametric measure spaces. Note that it is shown in Sect. B of the Supplementary Material that Algorithm 1 approximates $u_{\mathrm{GW},1}$ reasonably well in this setting. The results, where the spaces have been ordered lexicographically by (k, i), are visualized in Fig. 9. As previously, we observe that $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$ as well as $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ behave in a similar manner. More precisely, we see that both $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ discriminate well between the different classes and that their behavior does not change too much for an increasing level of perturbation. On the other hand, $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$ are very sensitive to the level of perturbation. For small t they discriminate better than $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ between the different classes and pick up clearly that the perturbed spaces differ. However, if the level of perturbation becomes too large both quantities start to discriminate between spaces from the same class (see Fig. 9).

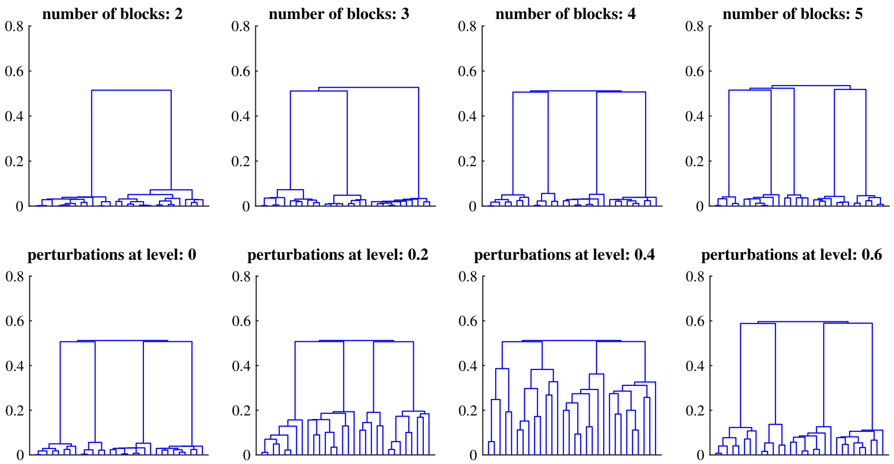

Fig. 8 Randomly sampled ultrametric measure spaces: Illustration of Z_k for $k = 2, 3, 4, 5$ (top row) and instances for perturbations of Z_4 with respect to perturbation level $t \in \{0, 0.2, 0.4, 0.6\}$ (bottom row)

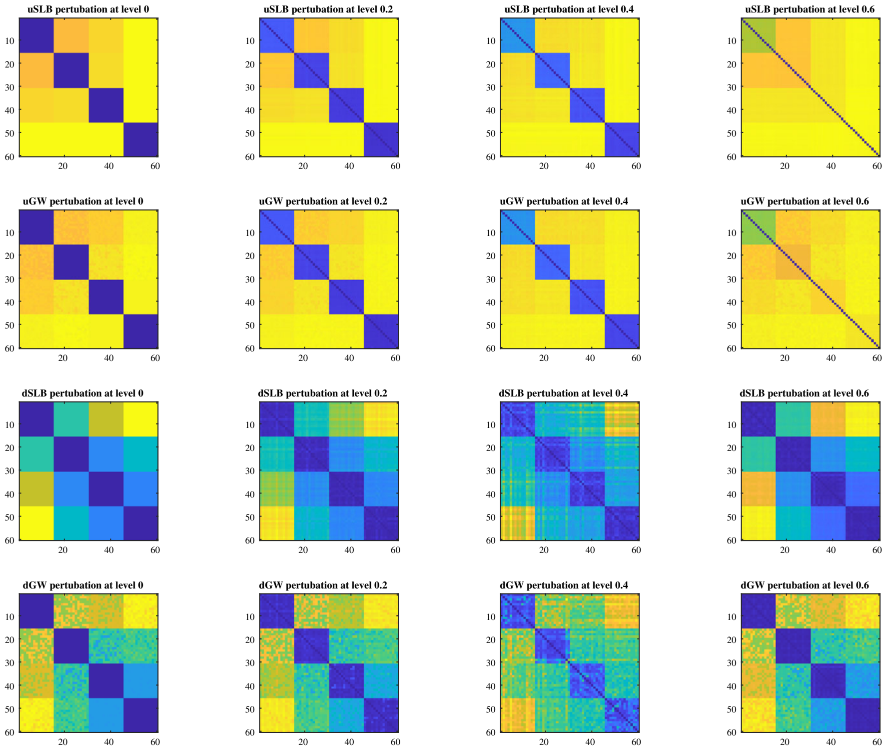

Fig. 9 $u_{\mathrm{GW},1}/\mathbf{SLB}_1^{\mathrm{ult}}$ **and** $d_{\mathrm{GW},1}/\mathbf{SLB}_1$ **among randomly generated ultrametric measure spaces:** Heatmap representations of $\mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{Z}_{k,t}^i, \mathcal{Z}_{k',t}^{i'})$ (top row), $u_{\mathrm{GW},1}(\mathcal{Z}_{k,t}^i, \mathcal{Z}_{k',t}^{i'})$ (second row), $\mathbf{SLB}_1(\mathcal{Z}_{k,t}^i, \mathcal{Z}_{k',t}^{i'})$ (third row) and $d_{\mathrm{GW},1}(\mathcal{Z}_{k,t}^i, \mathcal{Z}_{k',t}^{i'})$ (bottom row), $k, k' \in \{1, \ldots, 4\}$ and $i, i' \in \{1, \ldots, 15\}$

In conclusion, $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$ are sensitive to differences in the large scales of the considered ultrametric measure spaces. While this leads (from small t) to good discrimination in the above example, it also highlights that they are (different from $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$) susceptible to large scale noise.

5.4 Phylogenetic Tree Shape Comparison

In Sect. C of the Supplementary Material, we apply our lower bound $\mathbf{SLB}_1^{\mathrm{ult}}$, as well as $\mathbf{SLB}_1$ and the tree shape metric $d_{\mathrm{CP},2}$ introduced in [23, Eq. (4)], to the task of *phylogenetic tree shape comparison*: we use these distances to compare two sets of phylogenetic tree shapes based on the HA protein sequences from human influenza collected in different world regions. It turns out that (i) both $\mathbf{SLB}_1^{\mathrm{ult}}$ and $\mathbf{SLB}_1$ are able to detect some more refined clustering structure than $d_{\mathrm{CP},2}$ and (ii) $\mathbf{SLB}_1^{\mathrm{ult}}$ is more discriminating than $\mathbf{SLB}_1$ between tree shapes from different clusters.

6 Concluding Remarks

Since we suspect that computing $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ for finite p leads to NP-hard problems, it seems interesting to identify suitable collections of ultrametric measure spaces where these distances can be computed in polynomial time as done for the Gromov–Hausdorff distance in [63].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00454-023-00583-0.

Acknowledgements F.M. and A.M. thank the Mathematisches Forschungsinstitut Oberwolfach. Conversations which eventually led to this project were initiated during the 2019 workshop “Statistical and Computational Aspects of Learning with Complex Structure”.

Data Availability The code and datasets generated during and/or analyzed during the current study are available in https://github.com/ndag/uGW and from http://dx.doi.org/10.5061/dryad.3r8v1.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

A Technical Details from Sect. 2

A.1 Proofs from Sect. 2

In this section we give the proofs of various results form Sect. 2.

A.1.1 Proof of Theorem 2.2

Recall that for a given $\theta \in \mathcal{D}(X)$, we define $u_\theta : X \times X \to \mathbb{R}_{\geq 0}$ as follows:

$$u_\theta(x, x') := \inf\left\{t \geq 0 \mid x \text{ and } x' \text{ belong to the same block of } \theta(t)\right\}.$$

It is easy to verify that u_θ is an ultrametric. For any Cauchy sequence $\{x_n\}_{n\in\mathbb{N}}$ in (X, u_θ), let $D_i := \sup_{m,n\geq i} u_\theta(x_m, x_n)$ for each $i \in \mathbb{N}$. Then, each $D_i < \infty$ and $\lim_{i\to\infty} D_i = 0$. By definition of u_θ, for each $i \in \mathbb{N}$ the set $\{x_n\}_{n=i}^{\infty}$ is contained in the block $[x_i]_{D_i} \in \theta(D_i)$. Let $X_i := [x_i]_{D_i}$ for each $i \in \mathbb{N}$. Then, obviously we have that $X_j \subseteq X_i$ for any $1 \leq i < j$. By condition (vii) in Definition 2.1, we have that $\bigcap_{i\in\mathbb{N}} X_i \neq \emptyset$. Choose $x_* \in \bigcap_{i\in\mathbb{N}} X_i$, then it is easy to verify that $x_* = \lim_{n\to\infty} x_n$ and thus (X, u_θ) is a complete space. To prove that (X, u_θ) is a compact space, we need to verify that for each $t > 0$, X_t is a finite space (cf. Lemma A.7). Since $\theta(t)$ is finite by condition (vi) in Definition 2.1, we have that $X_t = \{[x]_t \mid x \in X\} = \theta(t)$ is finite and thus X is compact. Therefore, we have proved that $u_\theta \in \mathcal{U}(X)$. Based on this, the map $\Upsilon_X : \mathcal{D}(X) \to \mathcal{U}(X)$ defined by $\theta \mapsto u_\theta$ is well defined.

Now given $u \in \mathcal{U}(X)$, we define a map $\theta_u : [0, \infty) \to \mathbf{Part}(X)$ as follows: for each $t \geq 0$, consider the equivalence relation $\sim_t$ with respect to u, i.e., $x \sim_t x'$ iff $u(x, x') \leq t$. This is actually the same equivalence relation defined in Sect. 2.2 for introducing quotient ultrametric spaces. We then let $\theta_u(t)$ to be the partition induced by $\sim_t$, i.e., $\theta_u(t) = X_t$. It is not hard to show that θ_u satisfies conditions (i)–(v) in Definition 2.1. Since X is compact, then $\theta_u(t) = X_t$ is finite for each $t > 0$ and thus θ_u satisfies condition (vi) in Definition 2.1. Now, let $\{t_n\}_{n\in\mathbb{N}}$ be a decreasing sequence such that $\lim_{n\to\infty} t_n = 0$ and let $X_n \in \theta_X(t_n)$ be such that for any $1 \leq n < m$, $X_m \subseteq X_n$. Since each $X_n = [x_n]_{t_n}$ for some $x_n \in X$, X_n is a compact subset of X. Since X is also complete, we have that $\bigcap_{n\in\mathbb{N}} X_n \neq \emptyset$. Therefore, θ_u satisfies condition (vii) in Definition 2.1 and thus $\theta_u \in \mathcal{D}(X)$. Then, we define the map $\Delta_X : \mathcal{U}(X) \to \mathcal{D}(X)$ by $u \mapsto \theta_u$.

It is easy to check that Δ_X is the inverse of Υ_X and thus we have established that $\Upsilon_X : \mathcal{D}(X) \to \mathcal{U}(X)$ is bijective.

A.1.2 Proof of Lemma 2.8

First of all, we prove that the following supremum is attained to verify that the right-hand side of (12) is well defined

$$\sup_{\substack{B\in V(X)\backslash\{X\} \\ \alpha(B)\neq\beta(B)}} \operatorname{diam}(B^*).$$

Fix any $B_0 \in V(X)\backslash\{X\}$ such that $\alpha(B_0) \neq \beta(B_0)$. Then, it is obvious that $\operatorname{diam}(B_0^*) > 0$. By Lemma A.7, $X_{\operatorname{diam}(B_0^*)}$ is finite. So there are only finitely many $B \in V(X)\backslash\{X\}$ such that $\operatorname{diam}(B) \geq \operatorname{diam}(B_0^*)$ and thus $\operatorname{diam}(B^*) \geq \operatorname{diam}(B_0^*)$.

This implies that the supremum above is attained and thus

$$\sup_{\substack{B\in V(X)\backslash\{X\}\\ \alpha(B)\neq\beta(B)}} \operatorname{diam}(B^*) = \max_{\substack{B\in V(X)\backslash\{X\}\\ \alpha(B)\neq\beta(B)}} \operatorname{diam}(B^*). \tag{22}$$

Let B_1 denote the maximizer in (22) and let $\delta := \operatorname{diam}(B_1^*)$. It is easy to see that for any $x \in X$, $\alpha([x]_\delta) = \beta([x]_\delta)$.

By Strassen's theorem (see for example [28, Thm. 11.6.2]),

$$d_{\mathrm{W},\infty}(\alpha,\beta) = \inf\left\{r \geq 0 \mid \text{for any closed subset } A \subseteq X,\ \alpha(A) \leq \beta(A^r)\right\}, \tag{23}$$

where $A^r := \{x \in X \mid u_X(x, A) \leq r\}$.

Since $\alpha(B_1) \neq \beta(B_1)$, we assume without loss of generality that $\alpha(B_1) > \beta(B_1)$. By definition of B_1^*, it is obvious that $(B_1)^\delta = B_1^*$ (recall: $\delta := \operatorname{diam}(B_1^*)$) and $(B_1)^r = B_1$ for all $0 \leq r < \delta$. Therefore, $\alpha(B_1) \leq \beta((B_1)^r)$ only when $r \geq \delta$. By (23), this implies that $d_{\mathrm{W},\infty}(\alpha,\beta) \geq \delta$. Conversely, for any closed set A, we have that $A^\delta = \bigcup_{x\in A}[x]_\delta$. For two closed balls in ultrametric spaces, either one includes the other or they have no intersection. Therefore, there exists a subset $S \subseteq A$ such that $[x]_\delta \cap [x']_\delta = \emptyset$ for all $x, x' \in S$ and $x \neq x'$, and that $A^\delta = \bigsqcup_{x\in S}[x]_\delta$. Then, $\alpha(A) \leq \alpha(A^\delta) = \sum_{x\in S}\alpha([x]_\delta) = \sum_{x\in S}\beta([x]_\delta) = \beta(A^\delta)$. Hence, $d_{\mathrm{W},\infty}(\alpha,\beta) \leq \delta$ and thus we conclude the proof.

A.2 Technical Details from Sect. 2

In this section, we address various technical issues from Sect. 2.

A.2.1 Synchronized Rooted Trees

A *synchronized rooted tree*, is a combinatorial tree $T = (V, E)$ with a root $o \in V$ and a height function $h\colon V \to [0,\infty)$ such that $h^{-1}(0)$ coincides with the leaf set and $h(v) < h(v^*)$ for each $v \in V\backslash\{o\}$, where v^* is the parent of v. Similarly as in Theorem 2.2 that there exists a correspondence between ultrametric spaces and dendrograms, an ultrametric space X uniquely determines a synchronized rooted tree T_X [46].

Given $(X, u_X) \in \mathcal{U}$, recall from Sect. 2.3 that $V(X) := \bigcup_{t>0}\theta_X(t)$ and that for each $B \in V(X)\backslash\{X\}$, B^* denotes the smallest element in $V(X)$ containing B. The existence of B^* is guaranteed by the following lemma:

Lemma A.1 *Let $X \in \mathcal{U}$. For each $B \in V(X)$ such that $B \neq X$, there exists $B^* \in V(X)$ such that $B^* \neq B$ and $B^* \subseteq B'$ for all $B' \in V(X)$ with $B \subsetneq B'$.*

Proof Let $\delta := \operatorname{diam}(B)$. Let $x \in B$, then $B = [x]_\delta$. By Lemma A.7, X_δ is a finite set. Consider $\delta^* := \min\{u_{X_\delta}([x]_\delta, [x']_\delta) \mid [x']_\delta \neq [x]_\delta\}$. Let $B^* := [x]_{\delta^*}$, then B^* is the smallest element in $V(X)$ containing B under inclusion. Indeed, $B^* \neq B$ and if $B \subseteq B'$ for some $B' \in V(X)$, then $B' = [x]_r$ for some $r > \delta$. It is easy to see that for all $\delta < r < \delta^*$, $[x]_r = [x]_\delta$. Therefore, if $B' \neq B$, we must have that $r \geq \delta^*$ and thus $B^* = [x]_{\delta^*} \subseteq [x]_r = B'$. □

Now, we construct the synchronized rooted tree T_X corresponding to X via the proper dendrogram θ_X associated with u_X. We first define a combinatorial tree $T_X = (V_X, E_X)$ as follows: we let $V_X := V(X)$; for any distinct $B, B' \in V_X$, we let $(B, B') \in E_X$ iff either $B = (B')^*$ or $B' = B^*$. We choose $X \in V_X$ to be the root of T_X, then any $B \neq X$ in V_X has a unique parent B^*. We define $h_X : V_X \to [0, \infty)$ such that $h_X(B) := \mathrm{diam}(B)/2$ for any $B \in V_X$. Now, T_X endowed with the root X and the height function h_X is a synchronized rooted tree. It is easy to see that X can be isometrically identified with $h_X^{-1}(0)$ of the so-called *metric completion* of T_X (see [46, Sect. 2.3] for details). With this construction Lemma 2.7 follows directly from [46, Lem. 3.1].

A.3 $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ Between Compactly Supported Measures

Next, we demonstrate that Theorem 2.9 extends naturally to the case of compactly supported probability measures in $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$. For this purpose, it is important to note that compact subsets of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ have a very particular structure as shown by the next lemma.

Lemma A.2 *Let $X \subseteq (\mathbb{R}_{\geq 0}, \Lambda_\infty)$. X is a compact subset iff X is either a finite set or a countable set containing 0 and with 0 being the unique cluster point (w.r.t. the usual Euclidean distance Λ_1).*

Proof If X is finite, then obviously X is compact. Assume that X is a countable set with 0 being the unique cluster point (w.r.t. the usual Euclidean distance Λ_1). If $\{x_n\}_{n\in\mathbb{N}} \subseteq X$ is a Cauchy sequence with respect to Λ_∞, then either x_n is a constant when n is large or $\lim_{n\to\infty} x_n = 0$. In either case, the limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to X and thus X is complete. Now for any $\varepsilon > 0$, by Lemma A.7, X_ε is a finite set. Denote $X_\varepsilon = \{[x_1]_\varepsilon, \ldots, [x_n]_\varepsilon\}$. Then, $\{x_1, \ldots, x_n\}$ is a finite ε-net of X. Therefore, X is totally bounded and thus X is compact.

Now, assume that X is compact. Then, for any $\varepsilon > 0$, X_ε is a finite set. Suppose $X_\varepsilon = \{[x_1]_\varepsilon, \ldots, [x_n]_\varepsilon\}$ where $0 \leq x_1 < x_2 < \cdots < x_n$. Further, we have that $\Lambda_\infty(x_i, x_j) = x_j$ whenever $1 \leq i < j \leq n$. This implies that

(i) $x_i > \varepsilon$ for all $2 \leq i \leq n$;
(ii) $[x_i]_\varepsilon = \{x_i\}$ for all $2 \leq i \leq n$.

Therefore, $X \cap (\varepsilon, \infty) = \{x_2, \ldots, x_n\}$ is a finite set. Since $\varepsilon > 0$ is arbitrary, X is at most countable and has no cluster point (w.r.t. the Euclidean distance Λ_1) other than 0. If X is countable, then 0 must be a cluster point and by compactness of X, we have that $0 \in X$. ◻

Based on the special structure of compact subsets of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$, we derive the following extension of Theorem 2.9.

Theorem A.3 ($d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ between compactly supported measures) *Let $X := \{0\} \cup \{x_i \mid i \in \mathbb{N}\} \subseteq \mathbb{R}_{\geq 0}$ such that $0 < \ldots < x_n < x_{n-1} < \ldots < x_1$ and 0 is the only cluster point w.r.t. the usual Euclidean distance. Let $\alpha, \beta \in \mathcal{P}(X)$. Let $\alpha_i := \alpha(\{x_i\})$*

for $i \in \mathbb{N}$ *and* $\alpha_0 := \alpha(\{0\})$. *Similarly, let* $\beta_i := \beta(\{x_i\})$ *and* $\beta_0 := \beta(\{0\})$. *Then for* $p \in [1, \infty)$,

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha,\beta) = 2^{-1/p}\left(\sum_{i=2}^{\infty}\left|\sum_{j=i}^{\infty}(\alpha_j-\beta_j)+\alpha_0-\beta_0\right|\cdot|x_{i-1}^p - x_i^p| + \sum_{i=1}^{\infty}|\alpha_i-\beta_i|\cdot x_i^p\right)^{1/p}.$$

Let F_α *and* F_β *be the cumulative distribution functions of* α *and* β, *respectively. Then,*

$$d_{\mathrm{W},\infty}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha,\beta) = \max\Bigg(\max_{\substack{2\leq i<\infty \\ F_\alpha(x_i)\neq F_\beta(x_i)}} x_{i-1},\ \max_{\substack{1\leq i<\infty \\ \alpha_i\neq\beta_i}} x_i\Bigg).$$

Proof Note that $V(X) = \{\{0\}\cup\{x_j \mid j\geq i\} \mid i\in\mathbb{N}\}\cup\{\{x_i\}\mid i\in\mathbb{N}\}$ (recall that each set corresponds to a closed ball). Thus, we conclude by applying Lemmas 2.7 and 2.8. □

A.3.1 Closed-Form Solution for $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}$

In this section, we will derive the subsequent theorem.

Theorem A.4 *Given* $1 \leq p, q < \infty$ *and two compactly supported probability measures* α *and* β *on* $\mathbb{R}_{\geq 0}$, *we have that*

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta) \leq \left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t), F_\beta^{-1}(t))^p\,dt\right)^{1/p}.$$

When $q \leq p$, *the equality holds whereas when* $q > p$, *the equality does not hold in general.*

One important ingredient for the proof is the following direct adaptation of [67, Lem. 1].

Lemma A.5 *Let* X, Y *be two Polish metric spaces and let* $f: X \to \mathbb{R}$ *and* $g: Y \to \mathbb{R}$ *be measurable maps. Denote by* $f \times g: X \times Y \to \mathbb{R}^2$ *the map* $(x, y) \mapsto (f(x), g(y))$. *Then, for any* $\mu_Y \in \mathcal{P}(X)$ *and* $\mu_Y \in \mathcal{P}(Y)$

$$(f\times g)_\# \mathcal{C}(\mu_X,\mu_Y) = \mathcal{C}(f_\#\mu_Y, g_\#\mu_Y).$$

Based on Lemma A.5, we show the following auxiliary result.

Lemma A.6 *Let* $1 \leq q \leq p < \infty$. *Assume that* α *and* β *are compactly supported probability measures on* $\mathbb{R}_{\geq 0}$. *Then,*

$$\left(d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)\right)^p = \left(d_{\mathrm{W},p/q}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_\#\alpha,(S_q)_\#\beta)\right)^{p/q},$$

where $S_q: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *taking* x *to* x^q *is the* q*-snowflake transform defined in Sect. 3.3.*

Proof Since $p/q \geq 1$ and by Lemma A.5 we have that

$$\begin{aligned}\left(d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)\right)^p &= \inf_{\mu\in\mathcal{C}(\alpha,\beta)} \int_{\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}} (\Lambda_q(x,y))^p\,\mu(dx\times dy)\\ &= \inf_{\mu\in\mathcal{C}(\alpha,\beta)} \int_{\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}} |S_q(x)-S_q(y)|^{p/q}\,\mu(dx\times dy)\\ &= \inf_{\mu\in\mathcal{C}(\alpha,\beta)} \int_{\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}} |s-t|^{p/q}\,(S_q\times S_q)_\#\,\mu(ds\times dt)\\ &= \left(d_{\mathrm{W},p/q}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_\#\,\alpha,(S_q)_\#\,\beta)\right)^{p/q}.\end{aligned}$$

□

With Lemma A.6 at our disposal, we can demonstrate Theorem A.4.

Proof of Theorem A.4 We first note that

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta) = \inf_{(\xi,\eta)} (\mathbb{E}(\Lambda_q(\xi,\eta)^p))^{1/p},$$

where ξ and η are two random variables with marginal distributions α and β, respectively. Moreover, let ζ be the random variable uniformly distributed on $[0,1]$, then $F_\alpha^{-1}(\zeta)$ has distribution function F_α and $F_\beta^{-1}(\zeta)$ has distribution function F_β (see for example [88]). Let $\xi = F_\alpha^{-1}(\zeta)$ and $\eta = F_\beta^{-1}(\zeta)$, then we have

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta) \leq (\mathbb{E}(\Lambda_q(\xi,\eta)^p))^{1/p} = \left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t),F_\beta^{-1}(t))^p\,dt\right)^{1/p}.$$

Next, we assume that $q \leq p$. By Lemma A.6, we have that

$$\left(d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)\right)^p = \left(d_{\mathrm{W},p/q}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_\#\,\alpha,(S_q)_\#\,\beta)\right)^{p/q}.$$

Then,

$$\left(d_{\mathrm{W},p/q}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_\#\,\alpha,(S_q)_\#\,\beta)\right)^{p/q} = \int_0^1 |F_{\alpha,q}^{-1}(t)-F_{\beta,q}^{-1}(t)|^{p/q}\,dt,$$

where $F_{\alpha,q}$ and $F_{\beta,q}$ are distribution functions of $(S_q)_\#\,\alpha$ and $(S_q)_\#\,\beta$, respectively. It is easy to verify that $F_{\alpha,q}(t) = (F_\alpha^{-1}(t))^q$ and $F_{\beta,q}(t) = (F_\beta^{-1}(t))^q$. Therefore,

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta) = \left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t),F_\beta^{-1}(t))^p\,dt\right)^{1/p}.$$

Finally, we demonstrate that for $q > p$ the equality does not hold in general. We first consider the extreme case $p = 1$ and $q = \infty$ (though we require $q < \infty$ in the assumptions of the theorem, we relax this for now). Let $\alpha_0 = \delta_1/2 + \delta_2/2$ and

$\beta_0 = \delta_2/2 + \delta_3/2$ where δ_x means the Dirac measure at point $x \in \mathbb{R}_{\geq 0}$. Then, we have that

$$d_{\mathrm{W},1}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha_0, \beta_0) = \frac{3}{2} < \frac{5}{2} = \int_0^1 \Lambda_\infty(F_\alpha^{-1}(t), F_\beta^{-1}(t))\, dt.$$

It is not hard to see that both $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha_0, \beta_0)$ and

$$\left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t), F_\beta^{-1}(t))^p\, dt\right)^{1/p}$$

are continuous with respect to $p \in [1, \infty)$ and $q \in [1, \infty]$. Then, for p close to 1 and $q < \infty$ large enough, and in particular, $p < q$, we have that

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha_0, \beta_0) < \left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t), F_\beta^{-1}(t))^p\, dt\right)^{1/p}.$$

□

A.3.2 Miscellaneous

In the remainder of this section, we collect several technical results that find implicit or explicit usage throughout Sect. 2.

Lemma A.7 *A complete ultrametric space X is compact iff for any $t > 0$, X_t is finite.*

Proof Wan [92, Lem. 2.3] proves that whenever X is compact, X_t is finite for any $t > 0$.

Conversely, we assume that X_t is finite for any $t > 0$. We only need to prove that X is totally bounded. For any $\varepsilon > 0$, X_ε is a finite set and thus there exist $x_1, \ldots, x_n \in X$ such that $X_\varepsilon = \{[x_1]_\varepsilon, \ldots, [x_n]_\varepsilon\}$. Now, for any $x \in X$, there exists x_i for some $i = 1, \ldots, n$ such that $x \in [x_i]_\varepsilon$. This implies that $u_X(x, x_i) \leq \varepsilon$. Therefore, the set $\{x_1, \ldots, x_n\} \subseteq X$ is an ε-net of X. Then, X is totally bounded and thus compact. □

Lemma A.8 *$V(X)$ is the collection of all closed balls in X except for singletons $\{x\}$ such that x is a cluster point in X.*

Proof Given any $t > 0$ and $x \in X$, $[x]_t = B_t(x) = \{x' \in X \mid u_X(x, x') \leq t\}$. Therefore, $V(X)$ is a collection of closed balls in X. On the contrary, any closed ball $B_t(x)$ with positive radius $t > 0$ coincides with $[x]_t \in \theta_X(t)$ and thus belongs to $V(X)$. Now, for any singleton $\{x\} = B_0(x)$, if x is not a cluster point, then there exists $t > 0$ such that $B_t(x) = \{x\}$ which implies that $\{x\} \in V(X)$. If x is a cluster point, then for any $t > 0$, $\{x\} \subsetneq B_t(x) = [x]_t$. This implies that $\{x\} \neq [x]_t$ for all $t > 0$ and thus $\{x\} \notin V(X)$. This concludes the proof. □

B Technical Details from Sect. 3

B.1 Proofs from Sect. 3.1

Next, we give the missing proofs of the results stated in Sect. 3.1.

B.1.1 Proof of Proposition 3.3

Part 1. This directly follows from the definitions of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see (8) and (4)).
Part 2. This simply follows from Jensen's inequality.
Part 3. By Part 2, $\{u_{\mathrm{GW},n}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\}_{n\in\mathbb{N}}$ is an increasing sequence with a finite upper bound $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$. Therefore, $L := \lim_{n\to\infty} u_{\mathrm{GW},n}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$ exists and $L \leq u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$.

Next, we come to the opposite inequality. By Proposition B.1, there exist $u_n \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $\mu_n \in \mathcal{C}(\mu_X, \mu_Y)$ such that

$$\left(\int_{X\times Y} (u_n(x,y))^n \mu_n(dx\times dy)\right)^{1/n} = u_{\mathrm{GW},n}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}).$$

By Lemmas B.19 and B.21, the sequence $\{u_n\}_{n\in\mathbb{N}}$ uniformly converges to some $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $\{\mu_n\}_{n\in\mathbb{N}}$ weakly converges to some $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ (after taking appropriate subsequences of both sequences). Let

$$M := \sup_{(x,y)\in\mathrm{supp}(\mu)} u(x,y).$$

Let $\varepsilon > 0$ and let $U = \{(x,y) \in X\times Y \mid u(x,y) > M-\varepsilon\}$. Then, $\mu(U) > 0$. Since U is open, it follows that there exists a small $\varepsilon_1 > 0$ such that $\mu_n(U) > \mu(U) - \varepsilon_1 > 0$ for all n large enough (see e.g. [7, Thm. 2.1]). Moreover, by uniform convergence of the sequence $\{u_n\}_{n\in\mathbb{N}}$, we have $|u(x,y) - u_n(x,y)| \leq \varepsilon$ for any $(x,y) \in X\times Y$ when n is large enough. Therefore, we obtain for n large enough

$$\begin{aligned}\left(\int_{X\times Y} (u_n(x,y))^n \mu_n(dx\times dy)\right)^{1/n} &\geq (\mu_n(U))^{1/n}(M-2\varepsilon)\\ &\geq (\mu(U)-\varepsilon_1)^{1/n}(M-2\varepsilon).\end{aligned}$$

Letting $n \to \infty$, we obtain $L \geq M - 2\varepsilon$. Since $\varepsilon > 0$ is arbitrary, $L \geq M \geq u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$.

B.1.2 Proof of Theorem 3.4

In this section, we devote to prove Theorem 3.4. To this end, we will first verify the existence of optimal metrics and optimal couplings in (15).

Proposition B.1 (Existence of optimal couplings) *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Then, there always exist $u \in \mathcal{D}^{\text{ult}}(u_X, u_Y)$ and $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ such that for $1 \le p \le \infty$,*

$$u_{\text{GW},p}^{\text{sturm}}(\mathcal{X}, \mathcal{Y}) = \|u\|_{L^p(\mu)}.$$

Proof The following proof is a suitable adaptation from proof of [83, Lem. 3.3]. We will only prove the claim for the case $p < \infty$ since the case $p = \infty$ can be shown in a similar manner. Let $u_n \in \mathcal{D}^{\text{ult}}(u_X, u_Y)$ and $\mu_n \in \mathcal{C}(\mu_X, \mu_Y)$ be such that

$$\left(\int_{X\times Y} (u_n(x, y))^p \mu_n(dx \times dy)\right)^{1/p} \le u_{\text{GW},p}^{\text{sturm}}(\mathcal{X}, \mathcal{Y}) + \frac{1}{n}.$$

By Lemma B.19, $\{\mu_n\}_{n\in\mathbb{N}}$ weakly converges (after taking an appropriate subsequence) to some $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. By Lemma B.21, $\{u_n\}_{n\in\mathbb{N}}$ uniformly converges (after taking an appropriate subsequence) to some $u \in \mathcal{D}^{\text{ult}}(u_X, u_Y)$. Then, it is easy to verify that

$$\left(\int_{X\times Y} (u(x, y))^p \mu(dx \times dy)\right)^{1/p} \le u_{\text{GW},p}^{\text{sturm}}(\mathcal{X}, \mathcal{Y}).$$

□

As a direct consequence of the proposition, we get the subsequent result.

Corollary B.2 *Fix $1 \le p \le \infty$. Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Then, there exist a compact ultrametric space Z and isometric embeddings $\phi\colon X \hookrightarrow Z$ and $\psi\colon Y \hookrightarrow Z$ such that*

$$u_{\text{GW},p}^{\text{sturm}}(\mathcal{X}, \mathcal{Y}) = d_{\text{W},p}^{Z}(\phi_\# \mu_X, \psi_\# \mu_Y).$$

Before we come to the proof of Theorem 3.4, it remains to establish another auxiliary result. We ensure that the Wasserstein pseudometric of order p on a compact pseudo-ultrametric space (X, u_X) is for $p \in [1, \infty)$ a p-pseudometric and for $p = \infty$ a pseudo-ultrametric, i.e., we prove for $1 \le p < \infty$ that for all $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{P}(X)$,

$$d_{\text{W},p}^{(X,u_X)}(\mu_1, \mu_3) \le \left(\left(d_{\text{W},p}^{(X,u_X)}(\mu_1, \mu_2)\right)^p + \left(d_{\text{W},p}^{(X,u_X)}(\mu_2, \mu_3)\right)^p\right)^{1/p}$$

and for $p = \infty$ that for all $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{P}(X)$

$$d_{\text{W},p}^{(X,u_X)}(\mu_1, \mu_3) \le \max\left(d_{\text{W},p}^{(X,u_X)}(\mu_1, \mu_2), d_{\text{W},p}^{(X,u_X)}(\mu_2, \mu_3)\right).$$

Lemma B.3 *Let (X, u_X) be a compact pseudo-ultrametric space. Then, for $1 \le p \le \infty$ the p-Wasserstein metric $d_{\text{W},p}^{(X,u_X)}$ is a p-pseudometric on $\mathcal{P}(X)$. In particular, when $p = \infty$, it is a pseudo-ultrametric on $\mathcal{P}(X)$.*

Proof We prove the statement by adapting the proof of the triangle inequality for the p-Wasserstein distance (see e.g., [90, Thm. 7.3]). We only prove the case when $p<\infty$ whereas the case $p=\infty$ follows by analogous arguments.

Let $\alpha_1,\alpha_2,\alpha_3\in\mathcal{P}(X)$, denote by μ_{12} an optimal transport plan between α_1 and α_2 and by μ_{23} an optimal transport plan between α_2 and α_3 (see [91, Thm. 4.1] for the existence of μ_{12} and μ_{23}). Furthermore, let X_i be the support of α_i, $1\le i\le 3$. Then, by the Gluing Lemma [90, Lem. 7.6] there exists a measure $\mu\in\mathcal{P}(X_1\times X_2\times X_3)$ with marginals μ_{12} on $X_1\times X_2$ and μ_{23} on $X_2\times X_3$. Clearly, we obtain

$$\begin{aligned}\big(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_1,\alpha_3)\big)^p &\le \int_{X_1\times X_2\times X_3} {u_X}^p(x,z)\,\mu(dx\times dy\times dz)\\ &\le \int_{X_1\times X_2\times X_3}\big({u_X}^p(x,y)+{u_X}^p(y,z)\big)\,\mu(dx\times dy\times dz).\end{aligned}$$

Here, we used that u_X is an ultrametric, i.e., in particular a p-metric [64, Prop. 2.11]. With this we obtain that

$$\begin{aligned}\big(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_1,\alpha_2)\big)^p &\le \int_{X_1\times X_2} {u_X}^p(x,y)\,\mu_{12}\,(dx\times dy)\\ &\quad+\int_{X_2\times X_3} {u_X}^p(y,z)\,\mu_{23}\,(dy\times dz)\\ &=\big(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_1,\alpha_2)\big)^p+\big(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_2,\alpha_3)\big)^p.\end{aligned}$$

□

With Proposition B.1 and Lemma B.3 at our disposal we are now ready to prove Theorem 3.4 which states that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is indeed a p-metric on $\mathcal{U}^{\mathrm{w}}$.

Proof of Theorem 3.4 It is clear that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is symmetric and that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})=0$ if $\mathcal{X}\cong_{\mathrm{w}}\mathcal{Y}$. Furthermore, we remark that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\ge d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$ by Proposition 3.3. Since $d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})=0$ implies that $\mathcal{X}\cong_{\mathrm{w}}\mathcal{Y}$ ([84]), we have that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})=0$ implies that $\mathcal{X}\cong_{\mathrm{w}}\mathcal{Y}$. It remains to verify the p-triangle inequality. To this end, we only prove the case when $p<\infty$ whereas the case $p=\infty$ follows by analogous arguments.

Let $\mathcal{X},\mathcal{Y},\mathcal{Z}\in\mathcal{U}^{\mathrm{w}}$. Suppose $u_{XY}\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and $u_{YZ}\in\mathcal{D}^{\mathrm{ult}}(u_Y,u_Z)$ are optimal metric couplings such that

$$\begin{aligned}\big(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\big)^p &= \big(d_{\mathrm{W},p}^{(X\sqcup Y,u_{XY})}(\mu_X,\mu_Y)\big)^p,\\ \big(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{Y},\mathcal{Z})\big)^p &= \big(d_{\mathrm{W},p}^{(Y\sqcup Z,u_{YZ})}(\mu_Y,\mu_Z)\big)^p.\end{aligned}$$

Further, define u_{XYZ} on $X \sqcup Y \sqcup Z$ as

$$u_{XYZ}(x_1, x_2) = \begin{cases} u_{XY}(x_1, x_2), & x_1, x_2 \in X \sqcup Y, \\ u_{YZ}(x_1, x_2), & x_1, x_2 \in Y \sqcup Z, \\ \inf\{\max(u_{XY}(x_1, y), u_{YZ}(y, x_2)) \mid y \in Y\}, & x_1 \in X,\ x_2 \in Z, \\ \inf\{\max(u_{XY}(x_2, y), u_{YZ}(y, x_1)) \mid y \in Y\}, & x_1 \in Z,\ x_2 \in X. \end{cases}$$

Then, by [93, Lem. 1.1] u_{XYZ} is a pseudo-ultrametric on $X \sqcup Y \sqcup Z$ that coincides with u_{XY} on $X \sqcup Y$ and with u_{YZ} on $Y \sqcup Z$. Thus by Lemma B.3 we obtain that

$$\begin{aligned} \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Z})\right)^p &\le \left(d_{\mathrm{W},p}^{(X\sqcup Y\sqcup Z, u_{XYZ})}(\mu_X, \mu_Z)\right)^p \\ &\le \left(d_{\mathrm{W},p}^{(X\sqcup Y\sqcup Z, u_{XYZ})}(\mu_X, \mu_Y)\right)^p + \left(d_{\mathrm{W},p}^{(X\sqcup Y\sqcup Z, u_{XYZ})}(\mu_Y, \mu_Z)\right)^p \\ &= \left(d_{\mathrm{W},p}^{(X\sqcup Y, u_{XY})}(\mu_X, \mu_Y)\right)^p + \left(d_{\mathrm{W},p}^{(Y\sqcup Z, u_{YZ})}(\mu_Y, \mu_Z)\right)^p \\ &= \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})\right)^p + \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{Y}, \mathcal{Z})\right)^p. \end{aligned}$$

This gives the claim for $p < \infty$. □

B.1.3 Proof of Theorem 3.7

In order to proof Theorem 3.7, we will first establish the statement for *finite* ultrametric measure spaces. For this purpose, we need to introduce some notation. Given $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$, let $\mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X, u_Y)$ denote the collection of all admissible pseudo-ultrametrics on $X \sqcup Y$, where $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ is called *admissible*, if there exists no $u^* \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ such that $u^* \neq u$ and $u^*(x, y) \le u(x, y)$ for all $x, y \in X \sqcup Y$.

Lemma B.4 *For any $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$, $\mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X, u_Y) \neq \emptyset$. Moreover,*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = \inf_{u \in \mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X, u_Y)} d_{\mathrm{W},p}^{(X\sqcup Y, u)}(\mu_X, \mu_Y).$$

Proof If $\{u_n\}_{n\in\mathbb{N}} \subseteq \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ is a decreasing sequence (with respect to pointwise inequality), it is easy to verify that $u := \inf_{n\in\mathbb{N}} u_n \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and thus u is a lower bound of $\{u_n\}_{n\in\mathbb{N}}$. Then, by Zorn's lemma $\mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X, u_Y) \neq \emptyset$. Therefore, we obtain the claim. □

Combined with Example 3.6, the following result implies that each $u \in \mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X, u_Y)$ gives rise to an element in $\mathcal{A}$.

Lemma B.5 *Given* finite *spaces $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$, for each $u \in \mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X, u_Y)$, $u^{-1}(0) \neq \emptyset$.*

Proof Assume otherwise that $u^{-1}(0) = \emptyset$. Let $(x_0, y_0) \in X \times Y$ be such that $u(x_0, y_0) = \min_{x\in X, y\in Y} u(x, y)$. The existence of (x_0, y_0) is due to the finiteness of X and Y. We define $u_{(x_0,y_0)} : X \sqcup Y \times X \sqcup Y \to \mathbb{R}_{\ge 0}$ as follows:

(i) $u_{(x_0,y_0)}|_{X\times X} := u_X$ and $u_{(x_0,y_0)}|_{Y\times Y} := u_Y$.

(ii) For $(x, y) \in X \times Y$,

$$u_{(x_0,y_0)}(x, y) := \min\left(u(x, y), \max\left(u_X(x, x_0), u_Y(y, y_0)\right)\right).$$

(iii) For any $(y, x) \in Y \times X$, $u_{(x_0,y_0)}(y, x) := u_{(x_0,y_0)}(x, y)$.

It is easy to verify that $u_{(x_0,y_0)} \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$. Further, it is obvious that $u_{(x_0,y_0)}(x_0, y_0) = 0 < u(x_0, y_0)$ and that $u_{(x_0,y_0)}(x, y) \le u(x, y)$ for all $x, y \in X \sqcup Y$ which contradicts with $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$. Therefore, $u^{-1}(0) \neq \emptyset$. □

Theorem B.6 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ be* finite *spaces. Then, we have for each $p \in [1, \infty)$ that*

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \inf_{(A,\varphi)\in\mathcal{A}} d^{Z_A}_{\mathrm{W},p}\big((\phi^X_{(A,\varphi)})_\# \mu_X, (\psi^Y_{(A,\varphi)})_\# \mu_Y\big). \tag{24}$$

Proof By Lemma B.4 suffices to prove that $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$ induces $(A, \varphi) \in \mathcal{A}$ such that

$$d^{(X\sqcup Y,u)}_{\mathrm{W},p}(\mu_X, \mu_Y) \ge d^{Z_A}_{\mathrm{W},p}\big((\phi^X_{(A,\varphi)})_\# \mu_X, (\psi^Y_{(A,\varphi)})_\# \mu_Y\big).$$

Let $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$. Define $A_0 := \{x \in X \mid \exists\, y \in Y \text{ such that } u(x, y) = 0\}$ ($A_0 \neq \emptyset$ by Lemma B.5). By Example 3.6, the map $\varphi_0 \colon A_0 \to Y$ taking x to y such that $u(x, y) = 0$ is a well-defined isometric embedding. This means in particular that $(A_0, \varphi_0) \in \mathcal{A}$.

If $u(x, y) \ge u_{Z_{A_0}}(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y))$ holds for all $(x, y) \in X \times Y$, then we set $A := A_0$ and $\varphi := \varphi_0$. This gives

$$d^{(X\sqcup Y,u)}_{\mathrm{W},p}(\mu_X, \mu_Y) \ge d^{Z_A}_{\mathrm{W},p}\big((\phi^X_{(A,\varphi)})_\# \mu_X, (\psi^Y_{(A,\varphi)})_\# \mu_Y\big).$$

Otherwise, there exists $(x, y) \in X\backslash A_0 \times Y\backslash \varphi_0(A_0)$ such that

$$u(x, y) < u_{Z_{A_0}}\big(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y)\big)$$

(if $x \in A_0$ or $y \in \varphi_0(A_0)$, then $u(x, y) \ge u_{Z_{A_0}}\big(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y)\big)$ must hold). Let $(x_1, y_1) \in X\backslash A_0 \times Y\backslash \varphi_0(A_0)$ be such that

$$u(x_1, y_1) = \min\left\{u(x, y) \;\middle|\; \begin{array}{l}(x, y) \in X\backslash A_0 \times Y\backslash\varphi_0(A_0) \text{ and}\\ u(x, y) < u_{Z_{A_0}}\big(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y)\big)\end{array}\right\} > 0.$$

The existence of (x_1, y_1) follows from finiteness of X and Y. It is easy to check that φ_0 extends to an isometry from $A_0 \cup \{x_1\}$ to $\varphi_0(A_0) \cup \{y_1\}$ by taking x_1 to y_1. We denote the new isometry φ_1 and set $A_1 := A_0 \cup \{x_1\}$. If for any $(x, y) \in X \times Y$, we have that $u(x, y) \ge u_{Z_{A_1}}(\phi^X_{(A_1,\varphi_1)}(x), \psi^Y_{(A_1,\varphi_1)}(y))$, then we define $A := A_1$ and $\varphi := \varphi_1$. Otherwise, we continue the process to obtain $A_2, A_3, \ldots$. This process will eventually stop since we are considering finite spaces. Suppose the process stops at

A_n, then $A := A_n$ and $\varphi := \varphi_n$ satisfy that $u(x, y) \geq u_{Z_A}(\phi^X_{(A,\varphi)}(x), \psi^Y_{(A,\varphi)}(y))$ for any $(x, y) \in X \times Y$. Therefore,

$$d_{\mathrm{W},p}^{(X \sqcup Y,u)}(\mu_X, \mu_Y) \geq d_{\mathrm{W},p}^{Z_A}\big((\phi^X_{(A,\varphi)})_\# \mu_X, (\psi^Y_{(A,\varphi)})_\# \mu_Y\big).$$

Since $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$ is arbitrary, this gives the claim. □

As a direct consequence of Theorem B.6, we obtain that it is sufficient, as claimed in Remark 3.8, for finite spaces to infimize in (24) over the collection of all maximal pairs $\mathcal{A}^* \subseteq \mathcal{A}$. Recall that a pair $(A, \varphi_1) \in \mathcal{A}$ is denoted as *maximal*, if for all pairs $(B, \varphi_2) \in \mathcal{A}$ with $A \subseteq B$ and $\varphi_2|_A = \varphi_1$ it holds $A = B$.

Corollary B.7 *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ be* finite *spaces. Then, we have for each $p \in [1, \infty]$ that*

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \inf_{(A,\varphi)\in\mathcal{A}^*} d^{Z_A}_{\mathrm{W},p}\big((\phi^X_{(A,\varphi)})_\# \mu_X, (\psi^Y_{(A,\varphi)})_\# \mu_Y\big). \tag{25}$$

By proving Theorem B.6, we have verified Theorem 3.7 for finite ultrametric measure spaces. Then, we will use Theorem B.6 and weighted quotients to demonstrate Theorem 3.7. However, before we come to this, we need to establish the following two auxiliary results.

Lemma B.8 *Let $X \in \mathcal{U}$ be a compact ultrametric space. Let $t > 0$ and let $p \in [1, \infty)$. Then, for any $\alpha, \beta \in \mathcal{P}(X)$, we have that*

$$\big(d^{X_t}_{\mathrm{W},p}(\alpha_t, \beta_t)\big)^p \geq \big(d^X_{\mathrm{W},p}(\alpha, \beta)\big)^p - t^p,$$

where α_t is the push forward of α under the canonical quotient map $Q_t \colon X \to X_t$ taking $x \in X$ to $[x]_t \in X_t$.

Proof For any $\mu_t \in \mathcal{C}(\alpha_t, \beta_t)$, it is easy to see that there exists $\mu \in \mathcal{C}(\alpha, \beta)$ such that $\mu_t = (Q_t \times Q_t)_\# \mu$ where $Q_t \times Q_t \colon X \times X \to X_t \times X_t$ maps $(x, x') \in X \times X$ to $([x]_t, [x']_t)$. For example, suppose $X_t = \{[x_1]_t, \ldots, [x_n]_t\}$, then one can let

$$\mu := \sum_{i,j=1}^n \mu_t(([x_i]_t, [x_j]_t)) \, \frac{\alpha|_{[x_i]_t}}{\alpha([x_i]_t)} \otimes \frac{\beta|_{[x_j]_t}}{\beta([x_j]_t)},$$

where $\alpha|_{[x_i]_t}$ is the restriction of α on $[x_i]_t$.

For any $x, x' \in X$, we have that $(u_X(x, x'))^p \leq (u_{X_t}([x]_t, [x']_t))^p + t^p$. Then

$$\begin{aligned}\big(d^X_{\mathrm{W},p}(\alpha, \beta)\big)^p &\leq \int_{X\times X} (u_X(x, x'))^p \, \mu(dx \times dx') \\ &\leq \int_{X\times X} \big((u_{X_t}([x]_t, [x']_t))^p + t^p\big) \, \mu(dx \times dx') \\ &= \int_{X\times X} (u_X(Q_t(x), Q_t(x')))^p \, \mu(dx \times dx') + t^p\end{aligned}$$

$$= \int_{X_t \times X_t} \left(u_{X_t}([x]_t, [x']_t)\right)^p \mu_t(d[x]_t \times d[x']_t) + t^p$$

Infimizing over all $\mu_t \in \mathcal{C}(\alpha_t, \beta_t)$, we obtain the claim. □

Lemma B.9 *Let $\mathcal{X} \in \mathcal{U}^{\mathrm{w}}$ and let $p \in [1,\infty]$. Then, for any $t > 0$, we have that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \le t$. In particular,* $\lim_{t\to 0} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) = 0$.

Proof It is obvious that $(\mathcal{X}_t)_t \cong_{\mathrm{w}} \mathcal{X}_t$. Hence, it holds by Theorem 3.14 that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \le t$. By Proposition 3.3 we have that for any $p \in [1,\infty]$,

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \le u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \le t.$$

□

With Lemmas B.8 and B.9 available, we can come to the proof of Theorem 3.7.

Proof of Theorem 3.7 It follows from the definition of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see (8)) that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) \le \inf_{(A,\varphi)\in\mathcal{A}} d_{\mathrm{W},p}^{Z_A}\left((\phi_{(A,\varphi)}^X)_\# \mu_X, (\psi_{(A,\varphi)}^Y)_\# \mu_Y\right).$$

Hence, we focus on proving the opposite inequality. Given any $t > 0$, by Lemma A.7, both $\mathcal{X}_t$ and $\mathcal{Y}_t$ are finite spaces. By Theorem B.6 we have that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{Y}_t) = \inf_{(A_t,\varphi_t)\in\mathcal{A}_t} d_{\mathrm{W},p}^{Z_{A_t}}\left((\phi_{(A_t,\varphi_t)}^{X_t})_\# (\mu_X)_t, (\psi_{(A_t,\varphi_t)}^{Y_t})_\# (\mu_Y)_t\right),$$

where $\mathcal{A}_t := \{(A_t, \varphi_t) \mid \emptyset \ne A_t \subseteq X_t$ is closed and $\varphi_t \colon A_t \hookrightarrow Y_t$ is an isometricembedding $\}$.

For any $(A_t, \varphi_t) \in \mathcal{A}_t$, assume that $A_t = \{[x_1]_t^X, \ldots, [x_n]_t^X\}$ and that $\varphi_t([x_i]_t) = [y_i]_t \in Y_t$ for all $i = 1, \ldots, n$. Let $A := \{x_1, \ldots, x_n\}$. Then, the map $\varphi\colon A \to Y$ defined by $x_i \mapsto y_i$ for $i = 1, \ldots, n$ is an isometric embedding. Therefore, $(A, \varphi) \in \mathcal{A}$.

Claim 1 $((Z_A)_t, u_{(Z_A)_t}) \cong (Z_{A_t}, u_{Z_{A_t}})$.

Proof of Claim 1 We define a map $\Psi\colon (Z_A)_t \to Z_{A_t}$ by $[x]_t^{Z_A} \mapsto [x]_t^X$ for $x \in X$ and $[y]_t^{Z_A} \mapsto [y]_t^Y$ for $y \in Y\backslash\varphi(A)$. We first show that Ψ is well defined. For any $x' \in X$, if $u_{Z_A}(x, x') \le t$, then obviously we have that $u_X(x, x') = u_{Z_A}(x, x') \le t$ and thus $[x]_t^X = [x']_t^X$. Now, assume that there exists $y \in Y\backslash\varphi(A)$ such that $u_{Z_A}(x, y) \le t$, i.e., $[x]_t^{Z_A} = [y]_t^{Z_A}$. Then, by finiteness of A and definition of Z_A, there exists $x_i \in A$ such that $u_{Z_A}(x, y) = \max(u_X(x, x_i), u_Y(\varphi(x_i), y)) \le t$. This gives that

$$u_{Z_{A_t}}\left([x]_t^X, [y]_t^Y\right) \le \max\left(u_{X_t}\left([x]_t^X, [x_i]_t^X\right), u_{Y_t}\left([\varphi(x_i)]_t^Y, [y]_t^Y\right)\right) \le t.$$

However, this happens only if $u_{Z_{A_t}}([x]_t^X, [y]_t^Y) = 0$, that is, $[x]_t^X$ is identified with $[y]_t^Y$ under the map φ_t. Therefore, Ψ is well defined. It is easy to see from the definition that

Ψ is surjective. Thus, it suffices to show that Ψ is an isometric embedding to finish the proof. For any $x, x' \in X$ such that $u_X(x, x') > t$, we have that

$$\begin{aligned} u_{(Z_A)_t}\big([x]_t^{Z_A}, [x']_t^{Z_A}\big) &= u_{Z_A}(x, x') \\ &= u_X(x, x') = u_{X_t}\big([x]_t^X, [x']_t^X\big) = u_{Z_{A_t}}\big([x]_t^X, [x']_t^X\big). \end{aligned}$$

Similarly, for any $y, y' \in Y \backslash \varphi(A)$ such that $u_Y(y, y') > t$, we have that

$$u_{(Z_A)_t}\big([y]_t^{Z_A}, [y']_t^{Z_A}\big) = u_{Z_{A_t}}\big([y]_t^Y, [y']_t^Y\big).$$

Now, consider $x \in X$ and $y \in Y \backslash \varphi(A)$. Assume that $u_{Z_A}(x, y) > t$ (otherwise $[x]_t^{Z_A} = [y]_t^{Z_A}$). Then, we have that

$$u_{Z_A}(x, y) = \min_{i=1,\dots,n} \max(u_X(x, x_i), u_Y(\varphi(x_i), y)) > t.$$

This implies that

$$\begin{aligned} u_{Z_{A_t}}\big([x]_t^X, [y]_t^Y\big) &= \min_{i=1,\dots,n} \max\big(u_{X_t}\big([x]_t^X, [x_i]_t^X\big), u_{Y_t}\big(\varphi_t([x_i]_t^X), [y]_t^Y\big)\big) \\ &= \min_{i=1,\dots,n} \max(u_X(x, x_i), u_Y(\varphi(x_i), y)) \\ &= u_{Z_A}(x, y) = u_{(Z_A)_t}\big([x]_t^{Z_A}, [y]_t^{Z_A}\big). \end{aligned}$$

Therefore, Ψ is an isometric embedding and thus we conclude the proof. $\square$

By Lemma B.8 we have that

$$\begin{aligned} &\Big(d_{\mathrm{W},p}^{Z_{A_t}}\big(\big(\phi_{(A_t,\varphi_t)}^{X_t}\big)_\# (\mu_X)_t, \big(\psi_{(A_t,\varphi_t)}^{Y_t}\big)_\# (\mu_Y)_t\big)\Big)^p \\ &\qquad \geq \Big(d_{\mathrm{W},p}^{Z_A}\big(\big(\phi_{(A,\varphi)}^{X}\big)_\# \mu_X, \big(\psi_{(A,\varphi)}^{Y}\big)_\# \mu_Y\big)\Big)^p - t^p \end{aligned}$$

Therefore,

$$\begin{aligned} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{Y}_t) &= \inf_{(A_t,\varphi_t)\in\mathcal{A}_t} d_{\mathrm{W},p}^{Z_{A_t}}\big(\big(\phi_{(A_t,\varphi_t)}^{X_t}\big)_\# (\mu_X)_t, \big(\psi_{(A_t,\varphi_t)}^{Y_t}\big)_\# (\mu_Y)_t\big) \\ &\geq \inf_{(A,\varphi)\in\mathcal{A}} \Big(\Big(d_{\mathrm{W},p}^{Z_A}\big(\big(\phi_{(A,\varphi)}^{X}\big)_\# \mu_X, \big(\psi_{(A,\varphi)}^{Y}\big)_\# \mu_Y\big)\Big)^p - t^p\Big)^{1/p}. \end{aligned}$$

Notice that the last inequality already holds when we only consider (A, φ) corresponding to $(A_t, \varphi_t) \in \mathcal{A}_t$. By Lemma B.9, we have that

$$\begin{aligned} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) &= \lim_{t\to 0} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{Y}_t) \\ &\geq \inf_{(A,\varphi)\in\mathcal{A}} d_{\mathrm{W},p}^{Z_A}\big(\big(\phi_{(A,\varphi)}^{X}\big)_\# \mu_X, \big(\psi_{(A,\varphi)}^{Y}\big)_\# \mu_Y\big), \end{aligned}$$

which concludes the proof. $\square$

B.2 Proofs from Sect. 3.2

In this section, we give the complete proofs of the results stated in Sect. 3.2.

B.2.1 Proof of Proposition 3.10

Part 1. This follows directly from the definitions of $u_{\mathrm{GW},p}$ and $d_{\mathrm{GW},p}$ (see (11) and (5)).
Part 2. By Jensen's inequality we have that $\mathrm{dis}_p^{\mathrm{ult}}(\mu) \leq \mathrm{dis}_q^{\mathrm{ult}}(\mu)$ for any $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. Therefore, $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \leq u_{\mathrm{GW},q}(\mathcal{X}, \mathcal{Y})$.
Part 3. By Part 2 we know that $\{u_{\mathrm{GW},n}(\mathcal{X}, \mathcal{Y})\}_{n\in\mathbb{N}}$ is an increasing sequence with a finite upper bound $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$. Therefore, $L := \lim_{n\to\infty} u_{\mathrm{GW},n}(\mathcal{X}, \mathcal{Y})$ exists and it holds $L \leq u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$.

To prove the opposite inequality, by Proposition B.10, there exists for each $n \in \mathbb{N}$, $\mu_n \in \mathcal{C}(\mu_X, \mu_Y)$ such that

$$\|\Gamma^{\infty}_{X,Y}\|_{L^n(\mu_n\otimes\mu_n)} = u_{\mathrm{GW},n}(\mathcal{X}, \mathcal{Y}).$$

By Lemma B.19, $\{\mu_n\}_{n\in\mathbb{N}}$ weakly converges (after taking an appropriate subsequence) to some $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. Let

$$M := \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \Lambda_\infty(u_X(x, x'), u_Y(y, y'))$$

and for a given $\varepsilon > 0$ let

$$U = \left\{((x, y), (x', y')) \in X \times Y \times X \times Y \mid \Lambda_\infty(u_X(x, x'), u_Y(y, y')) > M - \varepsilon\right\}.$$

Then, we have $\mu\otimes\mu(U) > 0$. As μ_n weakly converges to μ, we have that $\mu_n\otimes\mu_n$ weakly converges to $\mu\otimes\mu$. Since U is open, there exists a small $\varepsilon_1 > 0$ such that $\mu_n\otimes\mu_n(U) > \mu\otimes\mu(U) - \varepsilon_1 > 0$ for n large enough (see e.g. [7, Thm. 2.1]). Therefore,

$$\|\Gamma^{\infty}_{X,Y}\|_{L^n(\mu_n\otimes\mu_n)} \geq (\mu_n\otimes\mu_n(U))^{1/n}(M - \varepsilon) \geq (\mu\otimes\mu(U) - \varepsilon_1)^{1/n}(M - \varepsilon).$$

Letting $n \to \infty$, we obtain $L \geq M - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we obtain $L \geq M \geq u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$.

B.2.2 Proof of Theorem 3.11

One main step to verify Theorem 3.11 is to demonstrate the existence of optimal couplings.

Proposition B.10 *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Then, for any $p \in [1, \infty]$, there always exists an optimal coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ such that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \mathrm{dis}_p^{\mathrm{ult}}(\mu)$.*

Proof We will only prove the claim for the case $p < \infty$ since the case $p = \infty$ can be proven in a similar manner. Let $\mu_n \in \mathcal{C}(\mu_X, \mu_Y)$ be such that

$$\|\Lambda_\infty(u_X, u_Y)\|_{L^p(\mu_n \otimes \mu_n)} \leq u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) + \frac{1}{n}.$$

By Lemma B.19, $\{\mu_n\}_{n\in\mathbb{N}}$ weakly converges to some $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ (after taking an appropriate subsequence). Then, by the boundedness and continuity of $\Lambda_\infty(u_X, u_Y)$ on $X \times Y \times X \times Y$ (cf. Lemma B.22) as well as the weak convergence of $\mu_n \otimes \mu_n$, we have that

$$\mathrm{dis}_p^{\mathrm{ult}}(\mu) = \lim_{n\to\infty} \mathrm{dis}_p^{\mathrm{ult}}(\mu_n) \leq u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}).$$

Hence, $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \mathrm{dis}_p^{\mathrm{ult}}(\mu)$. □

Based on Proposition B.10, it is straightforward to prove Theorem 3.11.

Proof of Theorem 3.11 It is clear that $u_{\mathrm{GW},p}$ is symmetric and that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = 0$ if $\mathcal{X} \cong_{\mathrm{w}} \mathcal{Y}$. Furthermore, we remark that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ by Proposition 3.10. Since $d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = 0$ implies that $\mathcal{X} \cong_{\mathrm{w}} \mathcal{Y}$ (see [60]), we have that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = 0$ implies that $\mathcal{X} \cong_{\mathrm{w}} \mathcal{Y}$. It remains to verify the p-triangle inequality. To this end, we only prove the case when $p < \infty$ whereas the case $p = \infty$ follows by analogous arguments.

Now let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be three ultrametric measure spaces. Let $\mu_{XY} \in \mathcal{C}(\mu_X, \mu_Y)$ and $\mu_{YZ} \in \mathcal{C}(\mu_Y, \mu_Z)$ be optimal (cf. Proposition B.10). By the Gluing Lemma [90, Lem. 7.6], there exists a measure $\mu_{XYZ} \in \mathcal{P}(X \times Y \times Z)$ with marginals μ_{XY} on $X \times Y$ and μ_{YZ} on $Y \times Z$. Further, we define $\mu_{XZ} = (\pi_{XZ})_\# \mu \in \mathcal{P}(X \times Z)$, where π_{XZ} denotes the canonical projection $X \times Y \times Z \to X \times Z$. Then

$$\begin{aligned}
(u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Z}))^p &\leq \|\Lambda_\infty(u_X, u_Z)\|^p_{L^p(\mu_{XZ}\otimes\mu_{XZ})} \\
&= \|\Lambda_\infty(u_X, u_Z)\|^p_{L^p(\mu_{XYZ}\otimes\mu_{XYZ})} \\
&\leq \|\Lambda_\infty(u_X, u_Y)\|^p_{L^p(\mu_{XYZ}\otimes\mu_{XYZ})} + \|\Lambda_\infty(u_Y, u_Z)\|^p_{L^p(\mu_{XYZ}\otimes\mu_{XYZ})} \\
&= \|\Lambda_\infty(u_X, u_Y)\|^p_{L^p(\mu_{XY}\otimes\mu_{XY})} + \|\Lambda_\infty(u_Y, u_Z)\|^p_{L^p(\mu_{YZ}\otimes\mu_{YZ})} \\
&= (u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}))^p + (u_{\mathrm{GW},p}(\mathcal{Y}, \mathcal{Z}))^p,
\end{aligned}$$

where the second inequality follows from the fact that Λ_∞ in an ultrametric on $\mathbb{R}_{\geq 0}$ (cf. [64, Exam. 2.7]) and the observation that an ultrametric is automatically a p-metric for any $p \in [1, \infty]$ [64, Prop. 2.11]. □

B.2.3 Proof of Theorem 3.14

We first prove that

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) = \inf\{t \geq 0 \mid \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\} \tag{26}$$

and then show that the infimum is attainable.

Since $\mathcal{X}_0 \cong_{\mathrm{w}} \mathcal{X}$ and $\mathcal{Y}_0 \cong_{\mathrm{w}} \mathcal{Y}$, if $\mathcal{X}_0 \cong_{\mathrm{w}} \mathcal{Y}_0$, then $\mathcal{X} \cong_{\mathrm{w}} \mathcal{Y}$ and thus by Theorem 3.11

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = 0 = \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}.$$

Now, assume that for some $t > 0$, $\mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t$. By Lemma A.7, for some $n \in \mathbb{N}$ we can write $X_t = \{[x_1]_t,\dots,[x_n]_t\}$ and $Y_t = \{[y_1]_t,\dots,[y_n]_t\}$ such that $u_{X_t}([x_i]_t,[x_j]_t) = u_{Y_t}([y_i]_t,[y_j]_t)$ and $\mu_X([x_i]_t) = \mu_Y([y_i]_t)$. Let $\mu_X{}^i := \mu_X|_{[x_i]_t}$ and $\mu_Y{}^i := \mu_Y|_{[y_i]_t}$ for all $i = 1,\dots,n$. Let $\mu := \sum_{i=1}^n \mu_X{}^i \otimes \mu_Y{}^i$. It is easy to check that $\mu \in \mathcal{C}(\mu_X,\mu_Y)$ and $\mathrm{supp}(\mu) = \bigcup_{i=1}^n [x_i]_t \times [y_i]_t$. Assume $(x,y) \in [x_i]_t \times [y_i]_t$ and $(x',y') \in [x_j]_t \times [y_j]_t$. If $i \neq j$, then $u_{X_t}([x_i]_t,[x_j]_t) = u_{Y_t}([y_i]_t,[y_j]_t)$ and thus

$$\Lambda_\infty(u_X(x,x'),u_Y(y,y')) = \Lambda_\infty(u_{X_t}([x_i]_t,[x_j]_t),u_{Y_t}([y_i]_t,[y_j]_t)) = 0.$$

If $i = j$, then $u_X(x,x'),u_Y(y,y') \leq t$ and thus $\Lambda_\infty(u_X(x,x'),u_Y(y,y')) \leq t$. In either case, we have that

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) \leq \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \Lambda_\infty(u_X(x,x'),u_Y(y,y')) \leq t.$$

Therefore, $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) \leq \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}$.

Conversely, suppose $\mu \in \mathcal{C}(\mu_X,\mu_Y)$ and let

$$t := \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \Lambda_\infty(u_X(x,x'),u_Y(y,y')).$$

By [60, Lem. 2.2], we know that $\mathrm{supp}(\mu)$ is a correspondence between X and Y. We define a map $f_t\colon X_t \to Y_t$ by taking $[x]_t^X \in X_t$ to $[y]_t^Y \in Y_t$ such that $(x,y) \in \mathrm{supp}(\mu)$. It is easy to check that f_t is well defined and moreover f_t is an isometry (see for example the proof of [64, Thm. 5.1]). Next, we prove that f_t is actually an isomorphism between $\mathcal{X}_t$ and $\mathcal{Y}_t$. For any $[x]_t^X \in X_t$, let $y \in Y$ be such that $(x,y) \in \mathrm{supp}(\mu)$ (in this case, $[y]_t^Y = f_t([x]_t^X)$). If there exists $(x',y') \in \mathrm{supp}(\mu)$ such that $x' \in [x]_t^X$ and $y' \notin [y]_t^Y$, then $\Lambda_\infty(u_X(x,x'),u_Y(y,y')) = u_Y(y,y') > t$, which is impossible. Consequently, $\mu([x]_t^X \times (Y\backslash[y]_t^Y)) = 0$ and similarly, $\mu((X\backslash[x]_t^X)\times[y]_t^Y) = 0$. This yields that

$$\mu_X([x]_t^X) = \mu([x]_t^Y \times Y) = \mu([x]_t^X \times [y]_t^Y) = \mu(X \times [y]_t^Y) = \mu_Y([y]_t^Y).$$

Therefore, f_t is an isomorphism between $\mathcal{X}_t$ and $\mathcal{Y}_t$. Hence, we have that $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) \geq \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}$ and hence $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}$.

Finally, we show that the infimum of $\inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}$ is attainable. Let $\delta := \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}$. If $\delta > 0$, let $\{t_n\}_{n\in\mathbb{N}}$ be a decreasing sequence converging to δ such that $\mathcal{X}_{t_n} \cong_{\mathrm{w}} \mathcal{Y}_{t_n}$ for all t_n. Since $\mathcal{X}_\delta$ and $\mathcal{Y}_\delta$ are finite, $\mathcal{X}_{t_n} = \mathcal{X}_\delta$ and $\mathcal{Y}_{t_n} = \mathcal{Y}_\delta$ when n is large enough. This immediately implies that $\mathcal{X}_\delta \cong_{\mathrm{w}} \mathcal{Y}_\delta$. Now, if $\delta = 0$, then by (26) we have that $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})=\delta=0$. By Theorem 3.11, $\mathcal{X}\cong_{\mathrm{w}}\mathcal{Y}$. This is equivalent to $\mathcal{X}_\delta\cong_{\mathrm{w}}\mathcal{Y}_\delta$. Therefore, the infimum of $\inf\{t\geq 0\,|\,\mathcal{X}_t\cong_{\mathrm{w}}\mathcal{Y}_t\}$ is always attainable.

B.2.4 Proof of Theorem 3.18

An important observation for the proof of Theorem 3.18 is that the snowflake transform relates the p-Wasserstein pseudometric on a pseudo-ultrametric space X with the 1-Wasserstein pseudometric on the space $S_p(X)$, $1\leq p<\infty$.

Lemma B.11 *Given a pseudo-ultrametric space (X,u_X) and $p\geq 1$, we have for any $\alpha,\beta\in\mathcal{P}(X)$ that $d_{\mathrm{W},p}^{(X,u_X)}(\alpha,\beta)=(d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta))^{1/p}$.*

Remark B.12 Since $S_p\circ u_X$ and u_X induce the same topology and thus the same Borel sets on X, $\mathcal{P}(X)=\mathcal{P}(S_p(X))$ and thus the expression $d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta)$ in the lemma is well defined.

Proof of Lemma B.11 Suppose $\mu_1,\mu_2\in\mathcal{C}(\alpha,\beta)$ are optimal for $d_{\mathrm{W},p}^{X}(\alpha,\beta)$ and $d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta)$, respectively (see Sect. B.5.1 for the existence of μ_1 and μ_2). Then,

$$\begin{aligned}\left(d_{\mathrm{W},p}^{(X,u_X)}(\alpha,\beta)\right)^p&=\int_{X\times X}(u_X(x,y))^p\,\mu_1(dx\times dy)\\&=\int_{X\times X}S_p(u_X)(x,y)\,\mu_1(dx\times dy)\geq d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta),\end{aligned}$$

and

$$\begin{aligned}d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta)&=\int_{X\times X}S_p(u_X)(x,y)\,\mu_2(dx\times dy)\\&=\int_{X\times X}(u_X(x,y))^p\,\mu_2(dx\times dy)\geq\left(d_{\mathrm{W},p}^{(X,u_X)}(\alpha,\beta)\right)^p.\end{aligned}$$

Therefore, $d_{\mathrm{W},p}^{(X,u_X)}(\alpha,\beta)=(d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta))^{1/p}$. □

With Lemma B.11 at our disposal we can prove Theorem 3.18.

Proof of Theorem 3.18 Let $\mu\in\mathcal{C}(\mu_X,\mu_Y)$. Then,

$$\|\Lambda_\infty(u_X,u_Y)\|^p_{L^p(\mu\times\mu)}=\|\Lambda_\infty(u_X{}^p,u_Y{}^p)\|_{L^1(\mu\times\mu)}.$$

By infimizing over $\mu\in\mathcal{C}(\mu_X,\mu_Y)$ on both sides, we obtain that $(u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}))^p=u_{\mathrm{GW},1}(S_p(\mathcal{X}),S_p(\mathcal{Y}))$.

To prove the second part of the claim, let $u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$. By Lemma B.11 we have that

$$\left(d_{\mathrm{W},p}^{(X\sqcup Y,u)}(\mu_X,\mu_Y)\right)^p=d_{\mathrm{W},1}^{(S_p(X)\sqcup S_p(Y),S_p(u))}(\mu_X,\mu_Y).$$

Finally, infimizing over $u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ yields

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})^p = u_{\mathrm{GW},1}^{\mathrm{sturm}}(S_p(\mathcal{X}),S_p(\mathcal{Y}))$$

□

As a direct consequence of Theorem 3.18, we obtain the following relation between $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$ and $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ for $p \in [1,\infty)$.

Corollary B.13 *For each $p \in [1,\infty)$, the metric space $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$ is isometric to the snowflake transform of $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$, i.e., $S_p(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}}) \cong (\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$.*

Proof Consider the snowflake transform map $S_p\colon \mathcal{U}^{\mathrm{w}} \to \mathcal{U}^{\mathrm{w}}$ sending $X \in \mathcal{U}^{\mathrm{w}}$ to $S_p(X) \in \mathcal{U}^{\mathrm{w}}$. It is obvious that S_p is bijective. By Theorem 3.18, S_p is an isometry from $S_p(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ to $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$. Therefore, $S_p(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}}) \cong (\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$. □

B.3 Proofs from Sect. 3.3

Throughout the following, we demonstrate the open claims from Sect. 3.3.

B.3.1 Proof of Theorem 3.19

First, we focus on the statement for $p = 1$, i.e., on showing

$$u_{\mathrm{GW},1}(\mathcal{X},\mathcal{Y}) \leq 2u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}). \tag{27}$$

Let $u \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and $\mu \in \mathcal{C}(\mu_X,\mu_Y)$ be such that

$$u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = \int u(x,y)\,\mu(dx \times dy).$$

The existence of u and μ follows from Proposition B.1.

Claim 1 For any $(x,y),(x',y') \in X \times Y$, we have

$$\Lambda_\infty(u_X(x,x'),u_Y(y,y')) \leq \max(u(x,y),u(x',y')) \leq u(x,y)+u(x',y').$$

Proof of Claim 1 We only need to show that

$$\Lambda_\infty(u_X(x,x'),u_Y(y,y')) \leq \max(u(x,y),u(x',y')).$$

If $u_X(x,x') = u_Y(y,y')$, then there is nothing to prove. Otherwise, we assume without loss of generality that $u_X(x,x') < u_Y(y,y')$. If $\max(u(x,y),u(x',y')) < u_Y(y,y')$, then by the strong triangle inequality we must have $u(x,y') = u_Y(y,y') = u(x',y)$. However, $u(x',y) \leq \max(u_X(x,x'),u(x,y)) < u_Y(y,y')$, which leads to a contradiction. Therefore,

$$\Lambda_\infty(u_X(x,x'),u_Y(y,y')) \leq \max(u(x,y),u(x',y')).$$

□

By Claim 1, we have that

$$\iint_{X\times Y\times X\times Y} \Lambda_\infty(u_X(x,x'),u_Y(y,y'))\,\mu(dx\times dy)\,\mu(dx'\times dy')$$
$$\leq \int_{X\times Y} u(x,y)\,\mu(dx\times dy)+\int_{X\times Y} u(x',y')\,\mu(dx'\times dy')\leq 2u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}).$$

Therefore, $u_{\mathrm{GW},1}(\mathcal{X},\mathcal{Y})\leq 2u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$.

Applying Theorem 3.18 and (27), yields that for any $p\in[1,\infty)$

$$\begin{aligned} u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) &= (u_{\mathrm{GW},1}(S_p(\mathcal{X}),S_p(\mathcal{Y})))^{1/p}\\ &\leq (2u_{\mathrm{GW},1}^{\mathrm{sturm}}(S_p(\mathcal{X}),S_p(\mathcal{Y})))^{1/p}=2^{1/p}\,u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}). \end{aligned}$$

B.3.2 Proof of Results in Example 3.21

It follows from [60, Rem. 5.17] that

$$d_{\mathrm{GW},p}^{\mathrm{sturm}}\big(\widehat{\Delta}_n(1),\widehat{\Delta}_{2n}(1)\big)\geq\frac{1}{4},$$
$$d_{\mathrm{GW},p}\big(\widehat{\Delta}_n(1),\widehat{\Delta}_{2n}(1)\big)\leq\frac{1}{2}\left(\frac{3}{2n}\right)^{1/p}.$$

Then, by Proposition 3.3, we have that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}\big(\widehat{\Delta}_n(1),\widehat{\Delta}_{2n}(1)\big)\geq d_{\mathrm{GW},p}^{\mathrm{sturm}}\big(\widehat{\Delta}_n(1),\widehat{\Delta}_{2n}(1)\big)\geq\frac{1}{4}.$$

Let μ_n denote the uniform probability measure of $\widehat{\Delta}_n(1)$. Since $\widehat{\Delta}_n(1)$ has the constant interpoint distance 1, it is obvious that for any coupling $\mu\in\mathcal{C}(\mu_n,\mu_{2n})$, $\mathrm{dis}_p(\mu)=\mathrm{dis}_p^{\mathrm{ult}}(\mu)$ This implies that $u_{\mathrm{GW},p}(\widehat{\Delta}_n(1),\widehat{\Delta}_{2n}(1))=2\,d_{\mathrm{GW},p}(\widehat{\Delta}_n(1),\widehat{\Delta}_{2n}(1))\leq(3/(2n))^{1/p}$.

B.3.3 Proof of Theorem 3.22

First, we prove that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\geq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$. Indeed, for any $u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and $\mu\in\mathcal{C}(\mu_X,\mu_Y)$, we have that

$$\begin{aligned} \sup_{(x,y)\in\mathrm{supp}(\mu)} u(x,y) &= \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \max(u(x,y),u(x',y'))\\ &\geq \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \Lambda_\infty(u_X(x,x'),u_Y(y,y'))\geq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}), \end{aligned}$$

where the first inequality follows from Claim 1 in the proof of Theorem 3.19. Then, by a standard limit argument, we conclude that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\geq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$.

Next, we prove that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq \min\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}$. Let $t > 0$ be such that $\mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t$ and let $\varphi\colon \mathcal{X}_t \to \mathcal{Y}_t$ denote such an isomorphism. Then, we define a function $u\colon X \sqcup Y \times X \sqcup Y \to \mathbb{R}_{\geq 0}$ as follows:

1. $u|_{X\times X} := u_X$ and $u|_{Y\times Y} := u_Y$;
2. for any $(x, y) \in X \times Y$,

$$u(x,y) := \begin{cases} u_{Y_t}(\varphi([x]_t^X), [y]_t^Y), & \text{if } \varphi([x]_t^X) \neq [y]_t^Y, \\ t, & \text{if } \varphi([x]_t^X) = [y]_t^Y; \end{cases}$$

3. for any $(y, x) \in Y \times X$, $u(y, x) := u(x, y)$.

Then, it is easy to verify that $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and that u is actually an ultrametric. Let $Z := (X \sqcup Y, u)$. By Lemma 2.8, we have

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq d_{\mathrm{W},\infty}^{Z}(\mu_X,\mu_Y) = \max_{\substack{B\in V(Z)\setminus\{Z\}\\ \mu_X(B)\neq\mu_Y(B)}} \mathrm{diam}(B^*).$$

We verify that $d_{\mathrm{W},\infty}^{Z}(\mu_X,\mu_Y) \leq t$ next. It is obvious that $Z_t \cong X_t \cong Y_t$. Write $X_t = \{[x_i]_t^X\}_{i=1}^n$ and $Y_t = \{[y_i]_t^Y\}_{i=1}^n$ such that $[y_i]_t^Y = \varphi([x_i]_t^X)$ for each $i = 1,\ldots,n$. Then, $[x_i]_t^Z = [y_i]_t^Z$ and $Z_t = \{[x_i]_t^Z \,|\, i = 1,\ldots,n\}$. Since φ is an isomorphism, for any $i = 1,\ldots,n$ we have that $\mu_X([x_i]_t^X) = \mu_Y([y_i]_t^Y)$ and thus $\mu_X([x_i]_t^Z) = \mu_Y([y_i]_s^Z) = \mu_Y([x_i]_t^Z)$ when μ_X and μ_Y are regarded as pushforward measures under the inclusion map $X \hookrightarrow Z$ and $Y \hookrightarrow Z$, respectively. Now for any $B \in V(Z)$ (cf. Sect. 2.3), if $\mathrm{diam}(B) \geq t$, then B is the union of certain $[x_i]_t^Z$'s in Z_t and thus $\mu_X(B) = \mu_Y(B)$. If $\mathrm{diam}(B) < t$ and $\mathrm{diam}(B^*) > t$, then there exists some x_i such that $B = [x_i]_s^Z$ and $[x_i]_s^Z = [x_i]_t^Z$ where $s := \mathrm{diam}(B)$. This implies that $\mu_X(B) = \mu_Y(B)$. In consequence, we have that $d_{\mathrm{W},\infty}^{Z}(\mu_X,\mu_Y) \leq t$ and thus $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq d_{\mathrm{W},\infty}^{(X\sqcup Y,u)}(\mu_X,\mu_Y) \leq t$. Therefore, $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_{\mathrm{w}} \mathcal{Y}_t\}$.

Finally, by invoking Theorem 3.14, we conclude that

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}).$$

B.3.4 Proof of Theorem 3.23

We prove the result via an explicit construction. By Theorem 3.22, we have $s = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$. By Theorem 3.14, there exists an isomorphism $\varphi\colon \mathcal{X}_s \to \mathcal{Y}_s$. Since $s > 0$, by Lemma A.7, both $\mathcal{X}_s$ and $\mathcal{Y}_s$ are finite spaces. Let $X_s = \{[x_1]_s^X,\ldots,[x_n]_s^X\}$, $Y_s = \{[y_1]_s^Y,\ldots,[y_n]_s^Y\}$ and assume $[y_i]_s^Y = \varphi([x_i]_s^X)$ for each $i = 1,\ldots,n$. Let $A := \{x_1,\ldots,x_n\}$ and define $\phi\colon A \to Y$ by sending x_i to y_i for each $i = 1,\ldots,n$. We prove that (A,ϕ) satisfies the conditions in the statement.

Since φ is an isomorphism, for any $1 \leq i < j \leq n$,

$$\begin{aligned} u_Y(y_i,y_j) &= u_{Y_s}([y_i]_s^Y,[y_j]_s^Y) \\ &= u_{Y_s}(\varphi([x_i]_s^X),\varphi([x_j]_s^X)) = u_{X_s}([x_i]_s^X,[x_j]_s^X) = u_X(x_i,x_j). \end{aligned}$$

This implies that $\phi\colon A \to Y$ is an isometric embedding and thus $(A, \phi) \in \mathcal{A}$.

It is obvious that $(Z_A)_s$ is isometric to both X_s and Y_s. In fact, $[x_i]_s^{Z_A} = [y_i]_s^{Z_A}$ in Z_A for each $i = 1, \dots, n$ and $(Z_A)_s = \{[x_i]_s^{Z_A} \,|\, i = 1, \dots, n\}$. Since φ is an isomorphism, for any $i = 1, \dots, n$ we have that $\mu_X([x_i]_s^X) = \mu_Y([y_i]_s^Y)$ and thus $\mu_X([x_i]_s^{Z_A}) = \mu_Y([y_i]_s^{Z_A}) = \mu_Y([x_i]_s^{Z_A})$ when μ_X and μ_Y are regarded as pushforward measures under the inclusion maps $X \to Z_A$ and $Y \to Z_A$, respectively. Now for any $B \in V(Z_A)$ (cf. Sect. 2.3), if $\mathrm{diam}(B) \geq s$, then B is the union of certain $[x_i]_s^{Z_A}$'s and thus $\mu_X(B) = \mu_Y(B)$. If otherwise $\mathrm{diam}(B) < s$ and $\mathrm{diam}(B^*) > s$, then there exists x_i such that $B = [x_i]_t^{Z_A}$ and $[x_i]_t^{Z_A} = [x_i]_s^{Z_A}$ where $t := \mathrm{diam}(B)$. This implies that $\mu_X(B) = \mu_Y(B)$. By Lemma 2.8, we have $d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y) \leq s$ and thus $d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y) = s$ since $d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y)$ is an upper bound for $s = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$ due to (8).

B.3.5 Proof of Theorem 3.25

In this section, we prove Theorem 3.25 by modifying the proof of [60, Prop. 5.3].

Lemma B.14 *Let (X, u_X) and (Y, u_Y) be compact ultrametric spaces and let $S \subseteq X \times Y$ be non-empty. Assume that $\sup_{(x,y),(x',y') \in S} \Lambda_\infty(u_X(x, x'), u_Y(y, y')) \leq \eta$. Define $u_S\colon X \sqcup Y \times X \sqcup Y \to \mathbb{R}_{\geq 0}$ as follows:*

(i) $u_S|_{X\times X} := u_X$ *and* $u_S|_{Y\times Y} := u_Y$*;*
(ii) *for any* $(x, y) \in X \times Y$, $u_S(x, y) := \inf_{(x',y') \in S} \max(u_X(x, x'), u_Y(y, y'), \eta)$*;*
(iii) *for any* $(x, y) \in X \times Y$, $u_S(y, x) := u_S(x, y)$.

Then, $u_S \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $u_S(x, y) \leq \eta$ for all $(x, y) \in S$.

Proof That $u_S \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ essentially follows by [93, Lem. 1.1]. It remains to prove the second half of the statement. For $(x, y) \in S$, we set $(x', y') := (x, y)$. This yields

$$u_S(x, y) \leq \max(u_X(x, x'), u_Y(y, y'), \eta) = \max(0, 0, \eta) = \eta.$$

□

Proof of Theorem 3.25 Let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ be a coupling s.t. $\|\Gamma_{X,Y}^\infty\|_{L^p(\mu\otimes\mu)} < \delta^5$. Set $\varepsilon := 4v_\delta(X) \leq 4$. By [60, Claim 10.1], there exist a positive integer $N \leq \lceil 1/\delta \rceil$ and points $x_1, \dots, x_N$ in X such that $\min_{i\neq j} u_X(x_i, x_j) \geq \varepsilon/2$, $\min_i \mu_X(B_\varepsilon^X(x_i)) > \delta$ and $\mu_X\big(\bigcup_{i=1}^N B_\varepsilon^X(x_i)\big) \geq 1 - \varepsilon$.

Claim 1 For every $i = 1, \dots, N$ there exists $y_i \in Y$ such that

$$\mu\big(B_\varepsilon^X(x_i) \times B_{2(\varepsilon+\delta)}^Y(y_i)\big) \geq (1 - \delta^2)\, \mu_X(B_\varepsilon^X(x_i)).$$

Proof of Claim 1 Assume the claim is false for some i and let

$$Q_i(y) = B_\varepsilon^X(x_i) \times (Y \backslash B_{2(\varepsilon+\delta)}^Y(y)).$$

Then, as $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ it holds

$$\begin{aligned}\mu_X(B_\varepsilon^X(x_i)) &= \mu\big(B_\varepsilon^X(x_i) \times Y\big)\\ &= \mu\big(B_\varepsilon^X(x_i) \times B^Y_{2(\varepsilon+\delta)}(y)\big) + \mu\big(B_\varepsilon^X(x_i) \times \big(Y \backslash B^Y_{2(\varepsilon+\delta)}(y)\big)\big).\end{aligned}$$

Consequently, we have that $\mu(Q_i(y)) \geq \delta^2 \mu_X(B_\varepsilon^X(x_i))$. Further, let

$$\mathcal{Q}_i := \big\{(x, y, x', y') \in X \times Y \times X \times Y \mid x, x' \in B_\varepsilon^X(x_i),\ u_Y(y, y') \geq 2(\varepsilon + \delta)\big\}.$$

Clearly, it holds for $(x, y, x', y') \in \mathcal{Q}_i$ that

$$\Gamma^\infty_{X,Y}(x, y, x', y') = \Lambda_\infty(u_X(x, x'), u_Y(y, y')) = u_Y(y, y') \geq 2\delta.$$

Further, we have that $\mu \otimes \mu(\mathcal{Q}_i) \geq \delta^4$. Indeed, it holds

$$\begin{aligned}\mu \otimes \mu(\mathcal{Q}_i) &= \int_{B_\varepsilon^X(x_i)\times Y} \int_{Q_i(y)} 1\, \mu(dx' \times dy')\, \mu(dx \times dy)\\ &= \int_{B_\varepsilon^X(x_i)\times Y} \mu(Q_i(y))\, \mu(dx \times dy)\\ &= \mu_X(B_\varepsilon^X(x_i)) \int_Y \mu(Q_i(y))\, \mu_Y(dy) \geq (\mu_X(B_\varepsilon^X(x_i)))^2 \delta^2 \geq \delta^4.\end{aligned}$$

However, this yields that

$$\begin{aligned}\|\Gamma^\infty_{X,Y}\|_{L^p(\mu\otimes\mu)} &\geq \|\Gamma^\infty_{X,Y}\|_{L^1(\mu\otimes\mu)}\\ &\geq \|\Gamma^\infty_{X,Y} \mathbb{1}_{\mathcal{Q}_i}\|_{L^1(\mu\otimes\mu)} \geq 2\delta \cdot \mu \otimes \mu(\mathcal{Q}_i) \geq 2\delta^5,\end{aligned}$$

which contradicts $\|\Gamma^\infty_{X,Y}\|_{L^p(\mu\otimes\mu)} < \delta^5$. $\square$

Define for each $i = 1, \ldots, N$, $S_i := B_\varepsilon^X(x_i) \times B^Y_{2(\varepsilon+\delta)}(y_i)$. Then, by Claim 1, $\mu(S_i) \geq \delta(1 - \delta^2)$, for all $i = 1, \ldots, N$.

Claim 2 $\Gamma^\infty_{X,Y}(x_i, y_i, x_j, y_j) \leq 6(\varepsilon + \delta)$ for all $i, j = 1, \ldots, N$.

Proof of Claim 2 Assume the claim fails for some (i_0, j_0), i.e.,

$$\Lambda_\infty(u_X(x_{i_0}, x_{j_0}), u_Y(y_{i_0}, y_{j_0})) > 6(\varepsilon + \delta) > 0.$$

Then, we have $\Lambda_\infty(u_X(x_{i_0}, x_{j_0}), u_Y(y_{i_0}, y_{j_0})) = \max(u_X(x_{i_0}, x_{j_0}), u_Y(y_{i_0}, y_{j_0}))$. We assume without loss of generality that

$$u_X(x_{i_0}, x_{j_0}) = \Lambda_\infty(u_X(x_{i_0}, x_{j_0}), u_Y(y_{i_0}, y_{j_0})) > u_Y(y_{i_0}, y_{j_0}).$$

Consider any $(x, y) \in S_{i_0}$ and $(x', y') \in S_{j_0}$. By the strong triangle inequality and the fact that $u_X(x_{i_0}, x_{j_0}) > 6(\varepsilon+\delta) > \varepsilon$, it is easy to verify that $u_X(x, x') = u_X(x_{i_0}, x_{j_0})$. Moreover,

$$\begin{aligned} u_Y(y, y') &\le \max\big(u_Y(y, y_{i_0}), u_Y(y_{i_0}, y_{j_0}), u_Y(y_{j_0}, y')\big) \\ &< \max\big(2(\varepsilon+\delta), u_X(x_{i_0}, x_{j_0}), 2(\varepsilon+\delta)\big) = u_X(x_{i_0}, x_{j_0}) = u_X(x, x'). \end{aligned}$$

Therefore, $\Gamma^{\infty}_{X,Y}(x, y, x', y') = u_X(x, x') = u_X(x_{i_0}, x_{j_0}) = \Gamma^{\infty}_{X,Y}(x_{i_0}, y_{i_0}, x_{j_0}, y_{j_0}) > 6(\varepsilon+\delta) > 2\delta$. Consequently, we have that

$$\begin{aligned} \|\Gamma^{\infty}_{X,Y}\|_{L^p(\mu\otimes\mu)} &\ge \|\Gamma^{\infty}_{X,Y}\|_{L^1(\mu\otimes\mu)} \\ &\ge \|\Gamma^{\infty}_{X,Y}\mathbb{1}_{S_{i_0}}\mathbb{1}_{S_{j_0}}\|_{L^1(\mu\otimes\mu)} \ge 2\delta\mu(S_{i_0})\mu(S_{j_0}) > 2\delta(\delta(1-\delta^2))^2. \end{aligned}$$

However, for $\delta \le 1/2$, $2\delta(\delta(1-\delta^2))^2 \ge 2\delta^5$. This leads to a contradiction. □

Consider $S \subseteq X \times Y$ given by $S := \{(x_i, y_i) \,|\, i = 1, \ldots, N\}$. Let u_S be the ultrametric on $X \sqcup Y$ given by Lemma B.14. By Claim 2,

$$\sup_{(x,y),(x',y')\in S} \Gamma^{\infty}_{X,Y}(x, y, x', y') \le 6(\varepsilon+\delta).$$

Then, for all $i = 1, \ldots, N$ we have that $u_S(x_i, y_i) \le 6(\varepsilon+\delta)$ and for any $(x, y) \in X \times Y$ we have that $u_S(x, y) \le \max(\mathrm{diam}(X), \mathrm{diam}(Y), 6(\varepsilon+\delta)) \le \max(\mathrm{diam}(X), \mathrm{diam}(Y), 27) =: M'$. Here in the second inequality we use the assumption that $\delta < 1/2$ and the fact that $\varepsilon = 4v_\delta(X) \le 4$.

Claim 3 Fix $i \in \{1, \ldots, N\}$. Then, for all $(x, y) \in S_i$, it holds $u_S(x, y) \le 6(\varepsilon+\delta)$.

Proof of Claim 3 Let $(x, y) \in S_i$. Then, $u_X(x, x_i) \le \varepsilon$ and $u_Y(y, y_i) \le 2(\varepsilon+\delta)$. Then, by the strong triangle inequality for u_S we obtain

$$\begin{aligned} u_S(x, y) &\le \max\{u_X(x, x_i), u_Y(y, y_i), u_S(x_i, y_i)\} \\ &\le \max\{\varepsilon, 2(\varepsilon+\delta), 6(\varepsilon+\delta)\} \le 6(\varepsilon+\delta). \end{aligned}$$

□

Let $L := \bigcup_{i=1}^N S_i$. The next step is to estimate the mass of μ in the complement of L.

Claim 4 $\mu(X \times Y \backslash L) \le \varepsilon + \delta$.

Proof of Claim 4 For each $i = 1, \ldots, N$, let

$$A_i := B^X_\varepsilon(x_i) \times (Y \backslash B^Y_{2(\varepsilon+\delta)}(y_i)).$$

Then,

$$A_i = \big(B^X_\varepsilon(x_i) \times Y\big) \backslash \big(B^X_\varepsilon(x_i) \times B^Y_{2(\varepsilon+\delta)}(y_i)\big) = \big(B^X_\varepsilon(x_i) \times Y\big) \backslash S_i.$$

Hence, $\mu(A_i) = \mu(B_\varepsilon^X(x_i)\times Y) - \mu(S_i) = \mu_X(B_\varepsilon^X(x_i)) - \mu(S_i)$, where the last equality follows from the fact that $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. By Claim 1, we have that $\mu(S_i) \geq \mu_X(B_\varepsilon^X(x_i))(1-\delta^2)$. Consequently, $\mu(A_i) \leq \mu_X(B_\varepsilon^X(x_i))\delta^2$. Notice that

$$X\times Y\backslash L \subseteq \left(X \Bigm\backslash \bigcup_{i=1}^N B_\varepsilon^X(x_i)\right)\times Y \cup \left(\bigcup_{i=1}^N A_i\right).$$

Hence,

$$\begin{aligned}\mu(X\times Y\backslash L) &\leq \mu_X\left(X \Bigm\backslash \bigcup_{i=1}^N B_\varepsilon^X(x_i)\right) + \sum_{i=1}^N \mu(A_i)\\ &\leq 1-\mu_X\left(\bigcup_{i=1}^N B_\varepsilon^X(x_i)\right) + \sum_{i=1}^N \delta^2\mu_X(B_\varepsilon^X(x_i)) \leq \varepsilon + N\cdot\delta^2 \leq \varepsilon+\delta.\end{aligned}$$

Here, the third inequality follows from the choice of the points x_is at the beginning of this section and from the fact that $N \leq [1/\delta]$. $\square$

Now,

$$\begin{aligned}\int_{X\times Y}(u_S(x,y))^p\,\mu(dx\times dy) &= \left(\int_L + \int_{X\times Y\backslash L}\right)(u_S(x,y))^p\,\mu(dx\times dy)\\ &\leq (6(\varepsilon+\delta))^p + M'^p\cdot(\varepsilon+\delta).\end{aligned}$$

Since we have for any $a, b \geq 0$ and $p \geq 1$ that $a^{1/p} + b^{1/p} \geq (a+b)^{1/p}$, we obtain

$$\begin{aligned}u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) &\leq (\varepsilon+\delta)^{1/p}\big(6(\varepsilon+\delta)^{1-1/p}+M'\big)\\ &\leq (\varepsilon+\delta)^{1/p}(27+M') \leq (4v_\delta(\mathcal{X})+\delta)^{1/p}\cdot M,\end{aligned}$$

where we used $\varepsilon = 4v_\delta(\mathcal{X})$ and $M := 2\max(\mathrm{diam}(X), \mathrm{diam}(Y)) + 54 \geq M'+27$. Since the roles of $\mathcal{X}$ and $\mathcal{Y}$ are symmetric, we have $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq (4\min(v_\delta(\mathcal{X}), v_\delta(Y))+\delta)^{1/p}\cdot M$. $\square$

B.4 Proofs from Sect. 3.4

The subsequent section contains the full proofs of the statements in Sect. 3.4.

B.4.1 Proof of Theorem 3.27

Part 1. We first prove that $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$ is non-separable for each $p \in [1,\infty]$. Recall notation in Example 3.5 and consider the family $\{\widehat{\Delta}_2(a)\}_{a\in[1,2]}$.

Claim 1 For all $a \neq b \in [1,2]$, $u_{\mathrm{GW},p}(\widehat{\Delta}_2(a), \widehat{\Delta}_2(b)) = 2^{-1/p}\Lambda_\infty(a,b) \geq 2^{-1/p}$, where $2^{-1/\infty} := 1$.

Proof of Claim 1 First note by Theorem 4.1 that

$$u_{\mathrm{GW},p}(\widehat{\Delta}_2(a), \widehat{\Delta}_2(b)) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\widehat{\Delta}_2(a), \widehat{\Delta}_2(b)).$$

It is easy to verify that $\mathbf{SLB}_p^{\mathrm{ult}}(\widehat{\Delta}_2(a), \widehat{\Delta}_2(b)) = 2^{-1/p}\Lambda_\infty(a,b)$. On the other hand, consider the diagonal coupling between μ_a and μ_b, then for $p \in [1,\infty)$

$$u_{\mathrm{GW},p}(\widehat{\Delta}_2(a), \widehat{\Delta}_2(b)) \leq \left(2 \cdot \Lambda_\infty(a,b)^p \cdot \frac{1}{2} \cdot \frac{1}{2}\right)^{1/p} = 2^{-1/p}\Lambda_\infty(a,b),$$

and for $p = \infty$, $u_{\mathrm{GW},\infty}(\widehat{\Delta}_2(a), \widehat{\Delta}_2(b)) \leq \Lambda_\infty(a,b)$. This concludes the proof. □

By Claim 1, we have that $\{\widehat{\Delta}_2(a)\}_{a\in[1,2]}$ is an uncountable subset of $\mathcal{U}^{\mathrm{w}}$ with pairwise distance greater than $2^{-1/p}$, which implies that $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$ is non-separable.

Now for $p \in [1,\infty)$, we show that $u_{\mathrm{GW},p}$ is not complete. Consider the family $\{\Delta_{2^n}(1)\}_{n\in\mathbb{N}}$ of 2^n-point spaces with unitary interpoint distances. Endow each space $\Delta_{2^n}(1)$ with the uniform measure μ_n and denote the corresponding ultrametric measure space by $\widehat{\Delta}_{2^n}(1)$. It is proven in [84, Exam. 2.2] that $\{\widehat{\Delta}_{2^n}(1)\}_{n\in\mathbb{N}}$ is a Cauchy sequence with respect to $d_{\mathrm{GW},p}$ without a compact metric measure space as limit. It is not hard to check that

$$u_{\mathrm{GW},p}(\widehat{\Delta}_{2^m}(1), \widehat{\Delta}_{2^n}(1)) = 2d_{\mathrm{GW},p}(\widehat{\Delta}_{2^m}(1), \widehat{\Delta}_{2^n}(1)), \quad \text{for all } \ n,m \in \mathbb{N}.$$

Therefore, $\{\widehat{\Delta}_{2^n}(1)\}_{n\in\mathbb{N}}$ is a Cauchy sequence with respect to $u_{\mathrm{GW},p}$ without limit in $\mathcal{U}^{\mathrm{w}}$. This implies that $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$ is not complete.

By Theorem 3.19 and that $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p})$ is not separable, $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ is not separable. As for completeness, consider the subset $X := \{1-1/n\}_{n\in\mathbb{N}} \subseteq (\mathbb{R}_{\geq 0}, \Lambda_\infty)$. By Lemma A.2, X is not a compact ultrametric space. Let $\mu_0 \in \mathcal{P}(X)$ be a probability defined as follows:

$$\mu_0\left(\left\{1-\frac{1}{n}\right\}\right) := 2^{-n}, \quad \text{for all } \ n \in \mathbb{N}.$$

For each $N \in \mathbb{N}$, let $X_N := \{1-1/n \mid n = 1,\ldots,N\}$. Since each X_N is finite, (X_N, Λ_∞) is a compact ultrametric space. Let $\mu_N \in \mathcal{P}(X_N)$ be a probability defined as follows:

$$\mu_N\left(\left\{1-\frac{1}{n}\right\}\right) := \begin{cases} 2^{-n}, & 1 \leq n < N, \\ 2^{-N+1} & n = N. \end{cases}$$

Then, it is easy to verify (e.g. via Theorem 3.7) that $\{(X_N, \Lambda_\infty, \mu_N)\}_{N\in\mathbb{N}}$ is a $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ Cauchy sequence with $(X, \Lambda_\infty, \mu_0)$ being the limit. Since the set X is not compact, $(X, \Lambda_\infty, \mu_0) \notin \mathcal{U}^{\mathrm{w}}$ and thus $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ is not complete.

Part 2. That $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},\infty})$ is non-separable is already proved in Part 1. We prove completeness next. Given a Cauchy sequence $\{\mathcal{X}_n = (X_n, u_n, \mu_n)\}_{n\in\mathbb{N}}$ with respect to $u_{\mathrm{GW},\infty}$, we have that the underlying ultrametric spaces $\{X_n\}_{n\in\mathbb{N}}$ form a Cauchy sequence w.r.t. u_{GH} due to Corollary 3.16. Since $(\mathcal{U}, u_{\mathrm{GH}})$ is complete (see [93, Prop. 2.1]), there exists a compact ultrametric space (X, u_X) such that $\lim_{n\to\infty} u_{\mathrm{GH}}(X_n, X) = 0$.

Let $\{\delta_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 such that $\delta_n \geq u_{\mathrm{GH}}(X_n, X)$. By Theorem 2.5, we have that $(X_n)_{\delta_n} \cong X_{\delta_n}$. Denote by $\widehat{\mu}_n \in \mathcal{P}(X_{\delta_n})$ the pushforward of $(\mu_n)_{\delta_n}$ under the isometry. Furthermore, we have by Lemma A.7 that X_{δ_n} is finite and we let $X_{\delta_n} = \{[x_1]_{\delta_n}, \ldots, [x_k]_{\delta_n}\}$ for $x_1, \ldots, x_k \in X$. Based on this, we define $\nu_n := \sum_{i=1}^{k} \widehat{\mu}_n([x_i]_{\delta_n}) \cdot \delta_{x_i} \in \mathcal{P}(X)$, where δ_{x_i} is the Dirac measure at x_i. Since X is compact, $\mathcal{P}(X)$ is weakly compact. Therefore, the sequence $\{\nu_n\}_{n\in\mathbb{N}}$ has a cluster point $\nu \in \mathcal{P}(X)$.

Now we show that $\mathcal{X} := (X, u_X, \nu)$ is a $u_{\mathrm{GW},\infty}$ cluster point of $\{\mathcal{X}_n\}_{n\in\mathbb{N}}$ and thus the limit of $\{\mathcal{X}_n\}_{n\in\mathbb{N}}$ (since $\{\mathcal{X}_n\}_{n\in\mathbb{N}}$ is Cauchy). Without loss of generality, we assume that $\{\nu_n\}_{n\in\mathbb{N}}$ weakly converges to ν. Fix any $\varepsilon > 0$, we need to show that $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{X}_n) \leq \varepsilon$ when n is large enough. For any fixed $x_* \in X$, $[x_*]_\varepsilon$ is both an open and closed ball in X. Therefore, $\nu([x_*]_\varepsilon) = \lim_{n\to\infty} \nu_n([x_*]_\varepsilon)$ (see e.g. [7, Thm. 2.1]). Since $\delta_n \to 0$ as $n \to \infty$, there exists $N_1 > 0$ such that for any $n > N_1$, $\delta_n < \varepsilon$. We specify an isometry $\varphi_n\colon (X_n)_{\delta_n} \to X_{\delta_n}$ that gives rise to the construction of ν_n. Then, we let $\psi_n\colon (X_n)_\varepsilon \to X_\varepsilon$ be the isometry such that the following diagram commutes:

$$\begin{tikzcd}
(X_n)_{\delta_n} \arrow[r, "\varphi_n"] \arrow[d, "\varepsilon\text{-quotient}"'] & X_{\delta_n} \arrow[d, "\varepsilon\text{-quotient}"] \\
(X_n)_\varepsilon \arrow[r, "\psi_n"] & X_\varepsilon
\end{tikzcd}$$

Assume that $[x_*]_\varepsilon^X = \bigcup_{i=1}^{l} [x_i]_{\delta_n}^X$. Let $x_*^n \in X_n$ be such that $\psi_n([x_*^n]_\varepsilon^{X_n}) = [x_*]_\varepsilon^X$ and let $x_1^n, \ldots, x_l^n \in X_n$ be such that $\varphi_n([x_i^n]_{\delta_n}^{X_n}) = [x_i]_{\delta_n}^X$ for each $i = 1, \ldots, l$. Then, $[x_*^n]_\varepsilon^{X_n} = \bigcup_{i=1}^{l} [x_i^n]_{\delta_n}^{X_n}$. Therefore,

$$\begin{aligned}
\nu_n([x_*]_\varepsilon^X) &= \sum_{i=1}^{l} \nu_n([x_i]_{\delta_n}^X) \\
&= \sum_{i=1}^{l} \widehat{\mu}_n([x_i]_{\delta_n}^X) = \sum_{i=1}^{l} \mu_n\big([x_i^n]_{\delta_n}^{X_n}\big) = \mu_n\big([x_*^n]_\varepsilon^{X_n}\big).
\end{aligned}$$

Since $\mathcal{X}_n$ is a Cauchy sequence, there exists $N_2 > 0$ such that $u_{\mathrm{GW},\infty}(\mathcal{X}_n, \mathcal{X}_m) < \varepsilon$ when $n, m > N_2$. Then, by Theorem 3.14, $(\mathcal{X}_n)_\varepsilon \cong_{\mathrm{w}} (\mathcal{X}_m)_\varepsilon$ for all $n, m > N_2$. By Lemma A.7, $(X_n)_\varepsilon$ is finite, then $(X_n)_\varepsilon$ has cardinality independent of n when $n > N_2$. For all $n > N_2$, we define the finite set $A_n := \{\mu_n([x^n]_\varepsilon^{X_n}) \,|\, x^n \in X_n\}$. A_n is independent of n since $(\mathcal{X}_n)_\varepsilon \cong_{\mathrm{w}} (\mathcal{X}_m)_\varepsilon$ for all $n, m > N_2$. This implies that $\mu_n([x_*^n]_\varepsilon^{X_n})$ only takes value in a finite set A_n. Combining with the fact that $\lim_{n\to\infty} \mu_n([x_*^n]_\varepsilon^{X_n}) = \lim_{n\to\infty} \nu_n([x]_\varepsilon^X) = \nu([x_*]_\varepsilon^X)$ exists, there exists $N_3 > 0$ such that when $n > N_3$, $\mu_n([x_*^n]_\varepsilon) \equiv C$ for some constant C. This implies that $\nu([x_*]_\varepsilon^X) = \mu_n([x_*^n]_\varepsilon^{X_n})$, when $n > \max(N_1, N_2, N_3)$. Since X_ε is finite, there exists a common $N > 0$ such that for all $n > N$ and for all $[x_*]_\varepsilon \in X_\varepsilon$ we have $\nu([x_*]_\varepsilon^X) = \mu_n([x_*^n]_\varepsilon^{X_n})$, where $[x_*^n]_\varepsilon^{X_n} = \psi_n^{-1}([x_*]_\varepsilon^X) \in (X_n)_\varepsilon$. This indicates that $\nu_\varepsilon = (\psi_n)_\# (\mu_n)_\varepsilon$ when $n > N$. Therefore, $\mathcal{X}_\varepsilon \cong_{\mathrm{w}} (\mathcal{X}_n)_\varepsilon$ and thus $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{X}_n) \le \varepsilon$.

B.4.2 Proof of Proposition 3.28

Next, we will demonstrate Proposition 3.28. However, before we come to this we recall some facts about p-metric and p-geodesic spaces.

Lemma B.15 ([64, Prop. 7.30]) *Given $p \in [1, \infty)$, if X is a p-metric space, then X is not q-geodesic for all $1 \le q < p$.*

Lemma B.16 ([64, Prop. 7.27]) *Let X be a geodesic metric space. Then, for any $p \ge 1$, $S_{1/p}(X)$ is p-geodesic, where S_α denotes the snowflake transform for $\alpha > 0$ (cf. Sect. 3.3).*

For $p = 1$, the proof is based on the following property of the 1-Wasserstein space.

Lemma B.17 ([9, Thm. 5.1]) *Let X be a compact metric space. Then, the space $W_1(X) := (\mathcal{P}(X), d_{\mathrm{W},1}^X)$ is a geodesic space.*

Based on the above results and Corollary B.2, the proof of Proposition 3.28 is straightforward.

Proof of Proposition 3.28 Let $\mathcal{X}$ and $\mathcal{Y}$ be two compact ultrametric measure spaces. First, we consider the case $p = 1$. By Corollary B.2, there exist a compact ultrametric space Z and isometric embeddings $\phi\colon X \hookrightarrow Z$ and $\psi\colon Y \hookrightarrow Z$ such that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = d_{\mathrm{W},p}^Z(\phi_\# \mu_X, \psi_\# \mu_Y).$$

The space $W_1(Z)$ is geodesic (cf. Lemma B.17). Therefore, there exists a Wasserstein geodesic $\widetilde{\gamma}\colon [0,1] \to W_1(Z)$ connecting $\phi_\# \mu_X$ and $\psi_\# \mu_Y$. This induces a curve $\gamma\colon [0,1] \to \mathcal{U}^{\mathrm{w}}$ where for each $t \in [0,1]$,

$$\gamma(t) := \big(\mathrm{supp}\,(\widetilde{\gamma}(t)), u|_{\mathrm{supp}(\widetilde{\gamma}(t)) \times \mathrm{supp}(\widetilde{\gamma}(t))}, \widetilde{\gamma}(t)\big).$$

Note that $\gamma(0) \cong_{\mathrm{w}} \mathcal{X}$ and $\gamma(1) \cong_{\mathrm{w}} \mathcal{Y}$ and hence we simply replace $\gamma(0)$ and $\gamma(1)$ with $\mathcal{X}$ and $\mathcal{Y}$, respectively. Now, for each $s, t \in [0,1]$, we have that

$$\begin{aligned} d_{\mathrm{GW},1}^{\mathrm{sturm}}(\gamma(s), \gamma(t)) &\le d_{\mathrm{W},1}^Z(\widetilde{\gamma}(s), \widetilde{\gamma}(t)) \\ &= |s-t|\, d_{\mathrm{W},1}^Z(\widetilde{\gamma}(0), \widetilde{\gamma}(1)) = |s-t|\, d_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}). \end{aligned}$$

Therefore, γ is a geodesic connecting $\mathcal{X}$ and $\mathcal{Y}$ and thus $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$ is geodesic.

For the case $p > 1$, by Corollary B.13, $S_p(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}}) \cong (\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}})$. This implies that $S_{1/p}(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},1}^{\mathrm{sturm}}) \cong (\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$. Hence, by Lemma B.16, $(\mathcal{U}^{\mathrm{w}}, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ is p-geodesic. □

B.5 Technical Details from Sect. 3

In this section, we address various technical issues from Sect. 3.

B.5.1 The Wasserstein Pseudometric

Given a set X, a pseudometric is a symmetric function $d_X : X \times X \to \mathbb{R}_{\geq 0}$ satisfying the triangle inequality and $d_X(x, x) = 0$ for all $x \in X$. Note that if moreover $d_X(x, y) = 0$ implies $x = y$, then d_X is a metric. There is a canonical identification on pseudometric spaces (X, d_X): $x \sim x'$ if $d_X(x, x') = 0$. Then, $\sim$ is in fact an equivalence relation and we define the quotient space $\widetilde{X} = X/\sim$. Define a function $\widetilde{d}_X : \widetilde{X} \times \widetilde{X} \to \mathbb{R}_{\geq 0}$ as follows:

$$\widetilde{d}_X([x], [x']) := \begin{cases} d_X(x, x') & \text{if } d_X(x, x') \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

$\widetilde{d}_X$ turns out to be a metric on $\widetilde{X}$. In the sequel, the metric space $(\widetilde{X}, \widetilde{d}_X)$ is referred to as the *metric space induced by the pseudometric space* (X, d_X). Note that $\widetilde{d}_X$ preserves the induced topology (see e.g. [41]) and thus the quotient map $\Psi : X \to \widetilde{X}$ is continuous.

Analogously to the Wasserstein distance, which is defined for probability measures on metric spaces, we define the *Wasserstein pseudometric* for measures on compact pseudometric spaces as done in [85]. Let $\alpha, \beta \in \mathcal{P}(X)$. Then, we define for $p \in [1, \infty)$ the Wasserstein pseudometric of order p as

$$d_{\mathrm{W},p}^{(X,d_X)}(\alpha, \beta) := \left(\inf_{\mu \in \mathcal{C}(\alpha,\beta)} \int_{X \times X} (d_X(x, y))^p \, \mu(dx \times dy) \right)^{1/p} \tag{28}$$

and for $p = \infty$ as

$$d_{\mathrm{W},\infty}^{(X,d_X)}(\alpha, \beta) := \inf_{\mu \in \mathcal{C}(\alpha,\beta)} \sup_{(x,y) \in \mathrm{supp}(\mu)} u(x, y). \tag{29}$$

It is easy to see that the Wasserstein pseudometric is closely related to the Wasserstein distance on the induced metric space. More precisely, one can show the following.

Lemma B.18 *Let (X, d_X) denote a compact pseudometric space, let $\alpha, \beta \in \mathcal{P}(X)$. Then, it follows for $p \in [1, \infty]$ that*

$$d_{\mathrm{W},p}^{(X,d_X)}(\alpha, \beta) = d_{\mathrm{W},p}^{(\widetilde{X},\widetilde{d}_X)}(\Psi_\# \alpha, \Psi_\# \beta) \tag{30}$$

and that the infimum in (28) (resp. in (29) if $p = \infty$) is attained for some $\mu \in \mathcal{C}(\alpha, \beta)$.

Proof In the course of this proof we focus on the case $p < \infty$ and remark that the case $p = \infty$ follows by similar arguments. The quotient map allows us to define the map $\theta\colon \mathcal{C}(\alpha,\beta) \to \mathcal{C}(\Psi_\#\alpha, \Psi_\#\beta)$ via $\mu \mapsto (\Psi\times\Psi)_\#\mu$. It is easy to see that θ is well defined and surjective. Furthermore, it holds by construction that

$$\int_{X\times X}(d_X(x,y))^p\,\mu(dx\times dy) = \int_{\widetilde{X}\times\widetilde{X}}(\widetilde{d}_X(x,y))^p\,\theta(\mu)(dx\times dy)$$

for all $\mu\in\mathcal{C}(\alpha,\beta)$. Hence, (30) follows.

We come to the second part of the claim. By [91, Sect. 4] there exists an optimal coupling $\widetilde{\mu}^*\in\mathcal{C}(\Psi_\#\alpha,\Psi_\#\beta)$ such that

$$d_{\mathrm{W},p}^{(\widetilde{X},\widetilde{d}_X)}(\Psi_\#\alpha,\Psi_\#\beta) = \left(\int_{\widetilde{X}\times\widetilde{X}}(\widetilde{d}_X(x,y))^p\,\widetilde{\mu}^*(dx\times dy)\right)^{1/p}.$$

In consequence, we find using our previous results that for any $\mu^*\in\theta^{-1}(\widetilde{\mu}^*)$ it holds

$$\begin{aligned}d_{\mathrm{W},p}^{(\widetilde{X},\widetilde{d}_X)}(\Psi_\#\alpha,\Psi_\#\beta) &= \left(\int_{\widetilde{X}\times\widetilde{X}}(\widetilde{d}_X(x,y))^p\,\widetilde{\mu}^*(dx\times dy)\right)^{1/p}\\ &= \left(\int_{X\times X}(d_X(x,y))^p\,\mu^*(dx\times dy)\right)^{1/p} = d_{\mathrm{W},p}^{(X,d_X)}(\alpha,\beta).\end{aligned}$$

This yields the claim. □

B.5.2 Regularity of the Cost Functionals of $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

In the remainder of this section, we collect various technical results required to demonstrate the existence of optimizers in the definitions of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see (8)) and $u_{\mathrm{GW},p}$ (see (11)).

Lemma B.19 *Let $\mathcal{X}=(X,u_X,\mu_X)$ and $\mathcal{Y}=(Y,u_Y,\mu_Y)$ be compact ultrametric measure spaces. Then, $\mu\in\mathcal{C}(\mu_X,\mu_Y)\subseteq\mathcal{P}(X\times Y,\max(u_X,u_Y))$ is compact w.r.t. weak convergence.*

Proof The proof follows directly from [21, Lem. 2.2]. □

Lemma B.20 *Let $\mathcal{X},\mathcal{Y}\in\mathcal{U}^{\mathrm{w}}$. Let $D_1\subseteq\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ be a non-empty subset satisfying the following: there exist $(x_0,y_0)\in X\times Y$ and $C>0$ such that $u(x_0,y_0)\le C$ for all $u\in D_1$. Then, D_1 is pre-compact with respect to uniform convergence.*

Proof Let $\{u_n\}_{n\in\mathbb{N}}\subseteq D_1$ be a sequence. Note that $X\times Y\subseteq X\sqcup Y\times X\sqcup Y$. Let $v_n := u_n|_{X\times Y}$. For any $n\in\mathbb{N}$ and any $(x,y),(x',y')\in X\times Y$, we have that

$$\begin{aligned}|u_n(x,y)-u_n(x',y')| &\le u_X(x,x')+u_Y(y,y')\\ &\le 2\max(u_X,u_Y)((x,y),(x',y')).\end{aligned}$$

This means that $\{v_n\}_{n\in\mathbb{N}}$ is equicontinuous with respect to the ultrametric $\max\{u_X,u_Y\}$ on $X\times Y$. Now, since $u_n(x_0,y_0)\le C$, we have that for any $(x,y)\in X\times Y$,

$$u_n(x,y)\le 2\max(u_X,u_Y)((x,y),(x_0,y_0))+u_n(x_0,y_0)$$

$$\leq 2\max(\mathrm{diam}(X), \mathrm{diam}(Y)) + C.$$

Consequently, $\{v_n\}_{n\in\mathbb{N}}$ is uniformly bounded. By the Arzéla–Ascoli theorem ([47, Thm. 7 on p. 61]), each subsequence of $\{v_n\}_{n\in\mathbb{N}}$ has a uniformly convergent subsequence. Hence, we assume without loss of generality that $\{v_n\}_{n\in\mathbb{N}}$ converges to $v\colon X\times Y\to\mathbb{R}_{\geq 0}$.

Now, we define a symmetric function $u\colon X\sqcup Y\times X\sqcup Y\to\mathbb{R}_{\geq 0}$ as follows:

(i) $u|_{X\times X} := u_X$ and $u|_{Y\times Y} := u_Y$;
(ii) $u|_{X\times Y} := v$; for $(y,x)\in Y\times X$, we let $u(y,x) := u(x,y)$.

It is easy to verify that $u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and that u is a cluster point of the sequence $\{u_n\}_{n\in\mathbb{N}}$. Therefore, D_1 is pre-compact. □

Lemma B.21 *Let $\mathcal{X}=(X,u_X,\mu_X)$ and $\mathcal{Y}=(Y,u_Y,\mu_Y)$ be compact ultrametric measure spaces. Let $\{\mu_n\}_{n\in\mathbb{N}}\subseteq\mathcal{C}(\mu_X,\mu_Y)$ be a sequence weakly converging to $\mu\in\mathcal{C}(\mu_X,\mu_Y)$. Let $\{u_n\}_{n\in\mathbb{N}}\subseteq\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$. Suppose that there exist a non-decreasing sequence $\{p_n\}_{n\in\mathbb{N}}\subseteq[1,\infty)$ and $C>0$ such that for all $n\in\mathbb{N}$,*

$$\left(\int_{X\times Y}(u_n(x,y))^{p_n}\,\mu_n(dx\times dy)\right)^{1/p_n}\leq C.$$

Then, $\{u_n\}_{n\in\mathbb{N}}$ uniformly converges to some $u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ (up to taking a subsequence).

Proof The following argument adapts the proof of [83, Lem. 3.3] to the current setting. For any $(x_0,y_0)\in\mathrm{supp}(\mu)$, there exist $\varepsilon,\delta>0$ and $N\in\mathbb{N}$ such that for all $n\geq N$

$$\begin{aligned}
C &\geq \left(\int_{X\times Y}(u_n(x,y))^{p_n}\,\mu_n(dx\times dy)\right)^{1/p_n}\\
&\geq \int_{X\times Y} u_n(x,y)\,\mu_n\,(dx\times dy)\\
&\geq \int_{B_\varepsilon^X(x_0)\times B_\varepsilon^Y(y_0)} u_n(x,y)\,\mu_n(dx\times dy)\\
&\geq \int_{B_\varepsilon^X(x_0)\times B_\varepsilon^Y(y_0)} (u_n(x_0,y_0)-2\varepsilon)\,\mu_n(dx\times dy)\\
&\geq (u_n(x_0,y_0)-2\varepsilon)\big(\mu(B_\varepsilon^X(x_0)\times B_\varepsilon^Y(y_0))-\delta\big).
\end{aligned}$$

Therefore, $\{u_n(x_0,y_0)\}_{n\geq N}$ is uniformly bounded. By Lemma B.20, we have that $\{u_n\}_{n\in\mathbb{N}}$ has a uniformly convergent subsequence. □

Lemma B.22 *Let X,Y be ultrametric spaces, then*

$$\Lambda_\infty(u_X,u_Y)\colon X\times Y\times X\times Y\to\mathbb{R}_{\geq 0}$$

is continuous with respect to the product topology (induced by $\max(u_X,u_Y,u_X,u_Y)$*).*

Proof Fix $(x, y, x', y') \in X \times Y \times X \times Y$ and $\varepsilon > 0$. Choose $0 < \delta < \varepsilon$ such that $\delta < u_X(x, x')$ if $x \neq x'$ and $\delta < u_Y(y, y')$ if $y \neq y'$. Then, consider any point $(x_1, y_1, x_1', y_1') \in X \times Y \times X \times Y$ such that

$$u_X(x, x_1), u_Y(y, y_1), u_X(x', x_1'), u_Y(y', y_1') \leq \delta.$$

For $u_X(x_1, x_1')$, we have the following two situations:

(i) $x = x'$: $u_X(x_1, x_1') \leq \max(u_X(x_1, x), u_X(x, x_1')) \leq \delta < \varepsilon$;
(ii) $x \neq x'$: $u_X(x_1, x_1') \leq \max(u_X(x_1, x), u_X(x, x'), u_X(x', x_1')) = u_X(x, x')$. Similarly, $u_X(x, x') \leq u_X(x_1, x_1')$ and thus $u_X(x, x') = u_X(x_1, x_1')$.

Similar result holds for $u_Y(y_1, y_1')$.

This leads to four cases for $\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1'))$:

(i) $x = x', y = y'$: In this case we have $u_X(x_1, x_1'), u_Y(y_1, y_1') < \varepsilon$. Then,

$$\begin{aligned}\left|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right| \\ = \Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) \leq \varepsilon;\end{aligned}$$

(ii) $x = x', y \neq y'$: Now $u_X(x_1, x_1') < \varepsilon$ and $u_Y(y_1, y_1') = u_Y(y, y')$. If $u_Y(y, y') \geq \varepsilon > u_X(x_1, x_1')$, then

$$\begin{aligned}\left|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right| \\ = |u_Y(y, y') - u_Y(y, y')| = 0.\end{aligned}$$

Otherwise $u_Y(y, y') < \varepsilon$, which implies that $\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) \leq \varepsilon$ and $\Lambda_\infty(u_X(x, x'), u_Y(y, y')) = u_Y(y, y') \leq \varepsilon$. Therefore,

$$\left|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right| \leq \varepsilon;$$

(iii) $x \neq x', y = y'$: Similarly with (ii) we have

$$\left|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right| \leq \varepsilon;$$

(iv) $x \neq x', y \neq y'$: Now $u_X(x_1, x_1') = u_X(x, x')$ and $u_Y(y_1, y_1') = u_Y(y, y')$. Therefore,

$$\left|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right| = 0.$$

In conclusion, whenever $u_X(x, x_1), u_Y(y, y_1), u_X(x', x_1'), u_Y(y', y_1') \leq \delta$ we have that

$$\left|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right| \leq \varepsilon.$$

Therefore, $\Lambda_\infty(u_X, u_Y)$ is continuous with respect to the metric $\max(u_X, u_Y, u_X, u_Y)$. $\square$

B.5.3 $u_{\mathrm{GW},p}$ and the One Point Space

Below, we prove that $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, between an arbitrary $\mathcal{X} \in \mathcal{U}^{\mathrm{w}}$ and the one point ultrametric measure space $*$ agrees with the p*-diameter* of $\mathcal{X}$ (see e.g., [60]): for $1 \leq p \leq \infty$ as $\mathrm{diam}_p(\mathcal{X}) := \|d_X\|_{L^p(\mu_X \otimes \mu_X)}$.

Proposition B.23 *Let* $* \in \mathcal{U}^{\mathrm{w}}$ *be the one-point space. Then, it holds for any* $1 \leq p \leq \infty$ *that* $u_{\mathrm{GW},p}(\mathcal{X}, *) = \mathrm{diam}_p(\mathcal{X})$.

Proof Note that in this case, for every $x, x' \in X$ $\Lambda_\infty(u_X(x,x'), u_*(*,*)) = \Lambda_\infty(u_X(x,x'), 0) = u_X(x,x')$. Therefore, thanks to this observation, and the fact that $\mu := \mu_X \otimes \delta_*$ is the unique coupling between μ_X and δ_*, (10) leads to the claim. □

C Technical Details from Sect. 4

C.1 Proofs from Sect. 4

In this section, we state the full proofs of the results from Sect. 4.

C.1.1 Proof of Theorem 4.1

Part 1. We observe that for any point x in an ultrametric space X, there always exists $x' \in X$ such that $u_X(x,x') = \mathrm{diam}(X)$ (see [27]). Since by assumption μ_X is fully supported, $s_{X,\infty} \equiv \mathrm{diam}(X)$ is a constant function. Therefore, $\Lambda_\infty(s_{X,\infty}(x), s_{Y,\infty}(y)) \equiv \Lambda_\infty(\mathrm{diam}(X), \mathrm{diam}(Y))$ for all $x \in X$ and $y \in Y$. This implies that $\mathbf{FLB}_\infty^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) = \Lambda_\infty(\mathrm{diam}(X), \mathrm{diam}(Y))$. By [64, Cor. 5.3] and Corollary 3.16, we have that

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) \geq u_{\mathrm{GH}}(X, Y) \geq \Lambda_\infty(\mathrm{diam}(X), \mathrm{diam}(Y)) = \mathbf{FLB}_\infty^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}).$$

Part 2. The proof for $d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{TLB}_p(\mathcal{X}, \mathcal{Y})$ in [60, Sect. 6] can be used essentially without any change for showing $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$. Hence, it remains to show that $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$:

Proposition C.1 *Let* $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^{\mathrm{w}}$ *and let* $p \in [1, \infty]$. *Then,* $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$.

In order to prove Proposition C.1, we need the following technical lemma.

Lemma C.2 *Let* $\mathcal{X} = (X, d_X, \mu_X) \in \mathcal{U}^{\mathrm{w}}$. *Then,* $\mathrm{spec}(X) := \{u_X(x,x') \,|\, x, x' \in \mathcal{X}\}$ *is a compact subset of* $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$.

Proof By Lemma A.7, we have that for each $t > 0$, X_t is a finite set. Let $\{t_n\}_{n=1}^\infty$ be a positive sequence decreasing to 0. Then, it is easy to see that $\mathrm{spec}(X) = \bigcup_{n=1}^\infty \mathrm{spec}(X_{t_n})$. Since each $\mathrm{spec}(X_{t_n})$ is a finite set, $\mathrm{spec}(X)$ is a countable set.

Now, pick any $0 \neq t \in \mathrm{spec}(X)$. Suppose t is a cluster point in $\mathrm{spec}(X)$. Then, there exists infinitely many $s \in \mathrm{spec}(X)$ greater than $t/2$. However, this will result in

$X_{t/2}$ being an infinite set, which contradicts the fact that $X_{t/2}$ is finite. Therefore, 0 is the only possible cluster point of $\mathrm{spec}(X)$. By Lemma A.2, we have that $\mathrm{spec}(X)$ is compact. $\square$

Next we demonstrate Proposition C.1 and hence finish the proof of Theorem 4.1.

Proof of Proposition C.1 We first prove the case when $p < \infty$. Let $dh_{\mathcal{X}}(x) := u_X(x,\cdot)_\# \mu_X$ and let $dh_{\mathcal{Y}}(y) := u_Y(y,\cdot)_\# \mu_Y$. Further, define

$$dH_{\mathcal{X}} := (u_X)_\# (\mu_X \otimes \mu_X), \quad dH_{\mathcal{Y}} := (u_Y)_\# (\mu_Y \otimes \mu_Y).$$

Lemma C.2 implies that the set $S := \mathrm{spec}(X) \cup \mathrm{spec}(Y)$ is a compact subset of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$. It is easy to see that $\mathrm{supp}(dh_{\mathcal{X}}), \mathrm{supp}(dh_{\mathcal{Y}}), \mathrm{supp}(dH_{\mathcal{X}}), \mathrm{supp}(dH_{\mathcal{Y}}) \subseteq S \subseteq \mathbb{R}_{\geq 0}$. Now, recall by Proposition 4.4 that $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) = d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dH_{\mathcal{X}}, dH_{\mathcal{Y}})$ and

$$\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) = \left(\inf_{\pi \in \mathcal{C}(\mu_X,\mu_Y)} \int_{X\times Y} \left(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y))\right)^p \mu(dx\times dy)\right)^{1/p}.$$

Further, we observe for any $x \in X$ and $y \in Y$ that

$$d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y)) = \inf_{\pi_{xy} \in \mathcal{C}(dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y))} \left(\int_{S\times S} \Lambda_\infty^p(s,t)\, \pi_{xy}(ds\times dt)\right)^{1/p}.$$

For the remainder of this proof, the metric on $S \subseteq \mathbb{R}_{\geq 0}$ is always given by Λ_∞. Additionally, $\mathcal{P}(S)$ denotes the set of probability measures on S and we equip $\mathcal{P}(S)$ with the Borel σ-field with respect to the topology induced by weak convergence.

Claim 1 There is a measurable choice $(x,y) \mapsto \pi_{xy}^*$ such that for each $(x,y) \in X \times Y$, $\pi_{x,y}^*$ is an optimal transport plan between $dh_{\mathcal{X}}(x)$ and $dh_{\mathcal{Y}}(y)$.

Proof of Claim 1 Since both Λ_1 and Λ_∞ induce the same topology on S, and thus the same Borel sets on S, $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_1)}$ and $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ metrize the same weak topology on $\mathcal{P}(S)$. By [61, Rem. 2.5], the following two maps are continuous with respect to the weak topology and thus measurable:

$$\Phi_1 \colon X \to \mathcal{P}(S),\ x \mapsto dh_{\mathcal{X}}(x) \ \text{ and } \ \Phi_2 \colon Y \to \mathcal{P}(S), \quad y \mapsto dh_{\mathcal{Y}}(y).$$

Since S is compact, the space $(\mathcal{P}(S), d_{\mathrm{W},p}^{(S,\Lambda_\infty)})$ is separable [91, Thm. 6.18]. This yields that $\mathscr{B}(\mathcal{P}(S)\times\mathcal{P}(S)) = \mathscr{B}(\mathcal{P}(S)) \otimes \mathscr{B}(\mathcal{P}(S))$ [33, Prop. 1.5]. Hence, the product $\Phi \colon X \times Y \to \mathcal{P}(S)\times\mathcal{P}(S)$ of Φ_1 and Φ_2, defined by $(x,y) \mapsto (dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y))$ is measurable [33, Prop. 2.4]. Then, a direct application of [91, Cor. 5.22] gives the claim. $\square$

Now, we have that for every $\mu \in \mathcal{C}(\mu_X,\mu_Y)$ that

$$\int_{X\times Y} \left(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y))\right)^p \mu(dx\times dy)$$

$$= \int_{X\times Y}\int_{S\times S} \Lambda_\infty^p(s,t)\,\pi_{xy}^*(ds\times dt)\,\mu(dx\times dy)$$
$$= \int_{S\times S} \Lambda_\infty^p(s,t)\,\overline{\mu}(ds\times dt),$$

by Fubini's Theorem, where $\overline{\mu} \in \mathcal{P}(S\times S)$ is defined as

$$\overline{\mu}(A) := \int_{X\times Y} \pi_{xy}^*(A)\,\mu(dx\times dy)$$

for any measurable $A \subseteq S\times S$. We remark that by Claim 1 the measure $\overline{\mu}$ is well defined. Next, we verify that $\overline{\mu} \in \mathcal{C}(dH_{\mathcal{X}}, dH_{\mathcal{Y}})$. For any measurable $A \subseteq (S, \Lambda_\infty)$ we have

$$\begin{aligned}
\overline{\mu}(A\times S) &= \int_{X\times Y} \pi_{x,y}^*(A\times S)\,\mu(dx\times dy)\\
&= \int_{X\times Y} dh_{\mathcal{X}}(x)(A)\,\mu(dx\times dy)\\
&= \int_{X} dh_{\mathcal{X}}(x)(A)\,\mu_X(dx)\\
&\overset{\text{(i)}}{=} \int_X\int_X \mathbb{1}_{\{d_X(x,x')\in A\}}\,\mu_X(dx')\,\mu_X(dx) = dH_{\mathcal{X}}(A),
\end{aligned}$$

where we have applied the marginal constraints for π_{xy} and μ. Further, (i) follows by the change-of-variables formula. The analogous arguments give that $\overline{\mu}(S\times B) = dH_{\mathcal{Y}}(B)$ for any measurable $B \subseteq S$. Thus, we conclude that for every $\mu \in \mathcal{C}(\mu_X, \mu_Y)$

$$\begin{aligned}
\int_{X\times Y} \big(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y))\big)^p\,\mu(dx\times dy) &= \int_{S\times S} \Lambda_\infty^p(s,t)\,\overline{\mu}(ds\times dt)\\
&\geq \inf_{\pi\in\mathcal{C}(dH_{\mathcal{X}},dH_{\mathcal{Y}})} \int_{S\times S} \Lambda_\infty(s,t)\,\pi(ds\times dt) = \big(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dH_{\mathcal{X}}, dH_{\mathcal{Y}})\big)^p.
\end{aligned}$$

This gives the claim for $p < \infty$.

Next, we prove the assertion for the case $p = \infty$. Note that for any $p < \infty$

$$\begin{aligned}
\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) &= \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \big\|d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(\cdot), dh_{\mathcal{Y}}(\cdot))\big\|_{L^p(\mu)}\\
&\leq \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \big\|d_{\mathrm{W},\infty}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(\cdot), dh_{\mathcal{Y}}(\cdot))\big\|_{L^\infty(\mu)} = \mathbf{TLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}),
\end{aligned}$$

where the inequality holds since $d_{\mathrm{W},p}^{(S,\Lambda_\infty)} \leq d_{\mathrm{W},\infty}^{(S,\Lambda_\infty)}$ and $\|\cdot\|_{L^p(\mu)} \leq \|\cdot\|_{L^\infty(\mu)}$.

By [35, Prop. 3] we have that

$$\begin{aligned}
\mathbf{SLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) &= d_{\mathrm{W},\infty}^{(S,\Lambda_\infty)}(dH_{\mathcal{X}}, dH_{\mathcal{Y}})\\
&= \lim_{p\to\infty} d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dH_{\mathcal{X}}, dH_{\mathcal{Y}}) = \lim_{p\to\infty} \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}).
\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbf{SLB}_{\infty}^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) &= \lim_{p\to\infty}\mathbf{SLB}_{p}^{\mathrm{ult}}(\mathcal{X},\mathcal{Y})\\ &\le \limsup_{p\to\infty}\mathbf{TLB}_{p}^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) \le \mathbf{TLB}_{\infty}^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}).\end{aligned}$$

□

C.1.2 Proof of Proposition 4.4

We only prove the first statement for $p \in [1,\infty)$. The case $p=\infty$ as well as the second statement can be proven in a similar manner.

By directly using the change-of-variables formula, we have the following:

$$\begin{aligned}\mathbf{SLB}_{p}^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) &= \inf_{\gamma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)}\|\Lambda_\infty(u_X,u_Y)\|_{L^p(\gamma)}^p\\ &= \inf_{\gamma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)}\|\Lambda_\infty\|_{L^p((u_X\times u_Y)_\#\gamma)}^p,\end{aligned}$$

where

$$u_X\times u_Y\colon X\times X\times Y\times Y\to\mathbb{R}_{\ge0}\times\mathbb{R}_{\ge0}$$

maps (x,x',y,y') to $(u_X(x,x'),u_Y(y,y'))$. By Lemma A.5,

$$(u_X\times u_Y)_\#\,\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)=\mathcal{C}((u_X)_\#(\mu_X\otimes\mu_X),(u_Y)_\#(\mu_Y\otimes\mu_Y)).$$

Therefore,

$$\begin{aligned}\mathbf{SLB}_{p}^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) &= \inf_{\gamma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)}\int_{\mathbb{R}_{\ge0}\times\mathbb{R}_{\ge0}}(\Lambda_\infty(s,t))^p\,(u_X\times u_Y)_\#\gamma(ds\times dt)\\ &= \inf_{\tilde\gamma\in\mathcal{C}((u_X)_\#(\mu_X\otimes\mu_X),(u_Y)_\#(\mu_Y\otimes\mu_Y))}\int_{\mathbb{R}_{\ge0}\times\mathbb{R}_{\ge0}}(\Lambda_\infty(s,t))^p\,\tilde\gamma(ds\times dt)\\ &= d_{\mathrm{W},p}^{(\mathbb{R}_{\ge0},\Lambda_\infty)}\big((u_X)_\#(\mu_X\otimes\mu_X),(u_Y)_\#(\mu_Y\otimes\mu_Y)\big).\end{aligned}$$

C.1.3 An Example: $\mathbf{SLB}^{\mathrm{ult}}$ vs. $\mathbf{TLB}^{\mathrm{ult}}$

We will demonstrate that there are ultrametric measure spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ such that $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}_1,\mathcal{X}_2)=0$, while it holds $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}_1,\mathcal{X}_2)>0$.

Consider the three point space $\Delta_3(1)=(\{x_1,x_2,x_3\},u)$ where $u(x_i,x_j)=1$ whenever $i\neq j$. Construct two probability measures $\mu_1:=\frac{2}{3}\delta_{x_1}+\frac{1}{6}\delta_{x_2}+\frac{1}{6}\delta_{x_3}$ and $\mu_2:=\frac{1}{3}\delta_{x_1}+\big(\frac{1}{3}-\frac{1}{2\sqrt{3}}\big)\delta_{x_2}+\big(\frac{1}{3}+\frac{1}{2\sqrt{3}}\big)\delta_{x_3}$. We then let $\mathcal{X}_1:=(\Delta_3(1),\mu_1)$ and $\mathcal{X}_2:=(\Delta_3(1),\mu_2)$. Obviously, $u_\#(\mu_1\otimes\mu_1)=u_\#(\mu_2\otimes\mu_2)=\delta_0/2+\delta_1/2$. Then, by Proposition 4.4 we immediately have that $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}_1,\mathcal{X}_2)=0$ for any $p\in[1,\infty]$.

Now, note that $u(x_1,\cdot)_\# \mu_1 = 2\delta_0/3 + \delta_1/3$, which is different from $u(x_i,\cdot)_\# \mu_2$ for each $i = 1, 2, 3$. This implies (by Proposition 4.4) that $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}_1,\mathcal{X}_2) > 0$ for any $p \in [1,\infty]$.

Note that this example works as well for showing that $\mathbf{TLB}_p(\mathcal{X}_1,\mathcal{X}_2) > \mathbf{SLB}_p(\mathcal{X}_1,\mathcal{X}_2) = 0$.

D Technical Details from Sect. 5

D.1 Technical Details from Sect. 5.2

Here, we list the precise results for the comparisons of the spaces $\mathcal{X}_i$, $1 \le i \le 4$, illustrated in Fig. 7. They are gathered in Tables 1 and 2.

Table 1 Comparison of different ultrametric measure spaces I:

	$u_{\mathrm{GW},1}$				$u_{\mathrm{GW},\infty}$			
	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$
$\mathcal{X}_1$	0.0000	0.9333	0.2444	0.2511	0.0000	2.1000	1.1000	2.000
$\mathcal{X}_2$	0.9333	0.0000	1.1778	1.0956	2.1000	0.0000	2.1000	2.1000
$\mathcal{X}_3$	0.2444	1.1778	0.0000	0.4493	1.1000	2.1000	0.0000	2.0000
$\mathcal{X}_4$	0.2511	1.0956	0.4493	0.0000	2.0000	2.1000	2.0000	0.0000

The values of $u_{\mathrm{GW},1}(\mathcal{X}_i,\mathcal{X}_j)$ (approximated by Algorithm 1) and $u_{\mathrm{GW},\infty}(\mathcal{X}_i,\mathcal{X}_j)$, $1 \le i \le j \le 4$, where $\mathcal{X}_i$, $1 \le i \le 4$, denote the ultrametric measure spaces displayed in Fig. 7

Table 2 Comparison of different ultrametric measure spaces II:

	$\mathbf{SLB}_1^{\mathrm{ult}}$			
	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$
$\mathcal{X}_1$	0.0000	0.9333	0.2444	0.0778
$\mathcal{X}_2$	0.9333	0.0000	1.1778	1.0111
$\mathcal{X}_3$	0.2444	1.1778	0.0000	0.2764
$\mathcal{X}_4$	0.0778	1.0111	0.2764	0.0000

The values of $\mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{X}_i,\mathcal{X}_j)$, $1 \le i \le j \le 4$, where $\mathcal{X}_i$, $1 \le i \le 4$, denote the ultrametric measure spaces displayed in Fig. 7

D.2 Technical Details from Sect. 5.3

Here, we state more results for the comparison of the ultrametric measure spaces illustrated in Fig. 7 and give the precise construction of the ultrametric spaces $Z_{k,t}^i$, $2 \le k \le 5$, $t = 0, 0.2, 0.4, 0.4$, $1 \le i \le 15$.

The ultrametric measure spaces from Fig. 7 See Table 3 for the results of comparing the ultrametric dissimilarity spaces in Fig. 7 based on $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$.

Table 3 Comparison of different ultrametric measure spaces III:

	$d_{\mathrm{GW},1}$				$\mathbf{SLB}_1$			
	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$
$\mathcal{X}_1$	0.0000	0.0444	0.0222	0.2111	0.0000	0.0444	0.0222	0.0422
$\mathcal{X}_2$	0.0444	0.0000	0.0667	0.2556	0.0444	0.0000	0.0667	0.0867
$\mathcal{X}_3$	0.0222	0.0667	0.0000	0.2253	0.0222	0.0667	0.0000	0.0573
$\mathcal{X}_4$	0.2111	0.2556	0.2253	0.0000	0.0422	0.0867	0.0573	0.0000

The values of $d_{\mathrm{GW},1}(\mathcal{X}_i, \mathcal{X}_j)$ (approximated by Algorithm 1) and $\mathbf{SLB}_1(\mathcal{X}_i, \mathcal{X}_j)$, $1 \leq i \leq j \leq 4$, where $(X_i, d_{X_i}, \mu_{X_i})$, $1 \leq i \leq 4$, denote the ultrametric measure spaces displayed in Fig. 7

Construction of Z_k For each $k = 2, 3, 4, 5$ we first draw a sample with $100 \times k$ points from the distribution $\sum_{i=0}^{k} U[1.5(k-1), 1.5(k-1)+1]/k$, where $U[a, b]$ denotes the uniform distribution on $[a, b]$. For each sample, we employ the single linkage algorithm to create a dendrogram, which then induces an ultrametric on the given sample. We further draw a 30-point subspace from each ultrametric space and denote it by Z_k. These four spaces have similar diameter values between 0.5 and 0.6. Each space Z_k is equipped with the uniform probability measure and the resulting ultrametric measure space is denoted by $\mathcal{Z}_k = (Z_k, u_{Z_k}, \mu_{Z_k})$, $k = 2, 3, 4, 5$. We remark that k can be regarded as the number of blocks in the dendrogram representation of the obtained ultrametric measure spaces (see the top row of Fig. 8 for a visualization of three 3-block spaces).

Perturbations at level t. Given a perturbation level $t \geq 0$ and an ultrametric space X, we consider the quotient space X_t. Each equivalence class $[x]_t \subseteq X$ is an ultrametric subspace of X. If $|[x]_t| > 1$, we let $m := |\mathrm{spec}([x]_t)| - 1$ and write $\mathrm{spec}([x]_t) = \{0 < s_1 < \cdots < s_m\}$. Let $\delta := \mathrm{diam}([x]_t)$. We generate m uniformly distributed numbers from $[0, t - \delta]$ and sort them according to ascending order to obtain $a_1 \leq \cdots \leq a_m$. We then perturb $u_X|_{[x]_t \times [x]_t}$ by replacing s_i with $s_i + a_i$ for each $i = 1, \ldots, m$. We do the same for all equivalence classes $[x]_t$ and thus obtain a new ultrametric on X.

References

1. Adelson-Welsky, G.M., Kronrode, A.S.: Sur les lignes de niveau des fonctions continues possédant des dérivées partielles. C. R. (Doklady) Acad. Sci. URSS (N.S.) **49**, 235–237 (1945)
2. Agarwal, P.K., Fox, K., Nath, A., Sidiropoulos, A., Wang, Y.: Computing the Gromov-Hausdorff distance for metric trees. ACM Trans. Algorithms **14**(2), 24 (2018)
3. Aho, A.V., Hopcroft, J.E., Ullman, J.D.: The Design and Analysis of Computer Algorithms. Pearson Education, London (1974)
4. Alvarez-Melis, D., Jaakkola, T.: Gromov–Wasserstein alignment of word embedding spaces. In: 2018 Conference on Empirical Methods in Natural Language Processing (Brussels 2018), pp. 1881–1890. Association for Computational Linguistics (2018)
5. Bartal, Y.: Probabilistic approximation of metric spaces and its algorithmic applications. In: 37th Annual Symposium on Foundations of Computer Science (Burlington 1996), pp. 184–193. IEEE, Los Alamitos (1996)
6. Billera, L.J., Holmes, S.P., Vogtmann, K.: Geometry of the space of phylogenetic trees. Adv. Appl. Math. **27**(4), 733–767 (2001)

7. Billingsley, P.: Convergence of Probability Measures. Probability and Statistics. Wiley, New York (2013)
8. Bonneel, N., Rabin, J., Peyré, G., Pfister, H.: Sliced and Radon Wasserstein barycenters of measures. J. Math. Imaging Vision **51**(1), 22–45 (2015)
9. Bottou, L., Arjovsky, M., Lopez-Paz, D., Oquab, M.: Geometrical insights for implicit generative modeling. In: Rozonoer, L., et al. (eds.) Braverman Readings in Machine Learning. Lecture Notes in Computer Science, vol. 11100, pp. 229–268. Springer, Cham (2018)
10. Brinkman, D., Olver, P.J.: Invariant histograms. Am. Math. Mon. **119**(1), 4–24 (2012)
11. Bronstein, A.M., Bronstein, M.M., Bruckstein, A.M., Kimmel, R.: Partial similarity of objects, or how to compare a centaur to a horse. Int. J. Comput. Vis. **84**(2), 163–183 (2009)
12. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Efficient computation of isometry-invariant distances between surfaces. SIAM J. Sci. Comput. **28**(5), 1812–1836 (2006)
13. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. Proc. Natl. Acad. Sci. USA **103**(5), 1168–1172 (2006)
14. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Topology-invariant similarity of nonrigid shapes. Int. J. Comput. Vis. **81**(3), 281–301 (2009)
15. Bronstein, A.M., Bronstein, M.M., Kimmel, R., Mahmoudi, M., Sapiro, G.: A Gromov–Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. Int. J. Comput. Vis. **89**(2–3), 266–286 (2010)
16. Brown, P., Pullan, W., Yang, Y., Zhou, Y.: Fast and accurate non-sequential protein structure alignment using a new asymmetric linear sum assignment heuristic. Bioinformatics **32**(3), 370–377 (2016)
17. Bunne, C., Alvarez-Melis, D., Krause, A., Jegelka, S.: Learning generative models across incomparable spaces. In: 36th International Conference on Machine Learning (Long Beach 2019), pp. 851–861. PMLR (2019)
18. Carlsson, G., Mémoli, F.: Characterization, stability and convergence of hierarchical clustering methods. J. Mach. Learn. Res. **11**, 1425–1470 (2010)
19. Chazal, F., Cohen-Steiner, D., Guibas, L.J., Mémoli, F., Oudot, S.Y.: Gromov–Hausdorff stable signatures for shapes using persistence. In: 7th Symposium on Geometry Processing (Berlin 2009), pp. 1393–1403. ACM, New York (2009)
20. Chen, J., Safro, I.: Algebraic distance on graphs. SIAM J. Sci. Comput. **33**(6), 3468–3490 (2011)
21. Chowdhury, S., Mémoli, F.: The Gromov-Wasserstein distance between networks and stable network invariants. Inf. Inference **8**(4), 757–787 (2019)
22. Chowdhury, S., Needham, T.: Generalized spectral clustering via Gromov–Wasserstein learning. In: 24th International Conference on Artificial Intelligence and Statistics (San Diego 2021), pp. 712–720. PMLR (2021)
23. Colijn, C., Plazzotta, G.: A metric on phylogenetic tree shapes. Syst. Biol. **67**(1), 113–126 (2018)
24. David, G., Semmes, S.W.: Fractured Fractals and Broken Dreams: Self-Similar Geometry Through Metric and Measure. Oxford Lecture Series in Mathematics and its Applications, vol. 7. Oxford University Press, New York (1997)
25. Do Ba, K., Nguyen, H.L., Nguyen, H.N., Rubinfeld, R.: Sublinear time algorithms for Earth mover's distance. Theory Comput. Syst. **48**(2), 428–442 (2011)
26. Dong, Y., Sawin, W.: COPT: Coordinated optimal transport on graphs. In: Advances in Neural Information Processing Systems, vol. 33, 19, 327–19, 338. Curran Associates, Red Hook (2020)
27. Dordovskyi, D., Dovgoshey, O., Petrov, E.: Diameter and diametrical pairs of points in ultrametric spaces. p-Adic Numbers Ultrametric Anal. Appl. **3**(4), 253–262 (2011)
28. Dudley, R.M.: Real Analysis and Probability. CRC Press, Boca Raton (2017)
29. Edwards, D.A.: The structure of superspace. In: Studies in Topology (Charlotte 1974), pp. 121–133. Academic Press, New York (1975)
30. Evans, S.N.: Probability and Real Trees. Lectures from the 35th Summer School on Probability Theory (Saint-Flour 2005). Springer, Berlin (2008)
31. Evans, S.N., Matsen, F.A.: The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. J. R. Stat. Soc. Ser. B. Stat. Methodol. **74**(3), 569–592 (2012)
32. Fakcharoenphol, J., Rao, S., Talwar, K.: A tight bound on approximating arbitrary metrics by tree metrics. J. Comput. Syst. Sci. **69**(3), 485–497 (2004)
33. Folland, G.B.: Real Analysis: Modern Techniques and their Applications. 2nd edn. Pure and Applied Mathematics (New York). Wiley, New York (1999)

34. Gellert, M., Hossain, M.F., Berens, F.J.F., Bruhn, L.W., Urbainsky, C., Liebscher, V., Lillig, C.H.: Substrate specificity of thioredoxins and glutaredoxins—towards a functional classification. Heliyon **5**(12), e02943 (2019)
35. Givens, C.R., Shortt, R.M.: A class of Wasserstein metrics for probability distributions. Mich. Math. J. **31**(2), 231–240 (1984)
36. Greven, A., Pfaffelhuber, P., Winter, A.: Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). Probab. Theory Relat. Fields **145**(1–2), 285–322 (2009)
37. Grindstaff, G., Owen, M.: Representations of partial leaf sets in phylogenetic tree space. SIAM J. Appl. Algebra Geom. **3**(4), 691–720 (2019)
38. Gromov, M.: Groups of polynomial growth and expanding maps (with an appendix by Jacques Tits). Inst. Hautes Études Sci. Publ. Math. **53**, 53–78 (1981)
39. Hein, J.: Reconstructing evolution of sequences subject to recombination using parsimony. Math. Biosci. **98**(2), 185–200 (1990)
40. Holm, L., Sander, C.: Protein structure comparison by alignment of distance matrices. J. Mol. Biol. **233**(1), 123–138 (1993)
41. Howes, N.R.: Modern Analysis and Topology. Springer, Berlin (2012)
42. Jain, A.K., Dorai, C.: 3D object recognition: representation and matching. Stat. Comput. **10**(2), 167–182 (2000)
43. Jardine, N., Sibson, R.: Mathematical Taxonomy. Wiley Series in Probability and Mathematical Statistics, Wiley, London (1971)
44. Kantorovich, L.: On the translocation of masses. C. R. (Doklady) Acad. Sci. URSS (N.S.) **37**, 199–201 (1942)
45. Kantorovich, L.V., Rubinstein, G.S.: On a space of completely additive functions. Vestnik Leningrad. Univ. **13**(7), 52–59 (1958) (in Russian)
46. Kloeckner, B.R.: A geometric study of Wasserstein spaces: ultrametrics. Mathematika **61**(1), 162–178 (2015)
47. Kolmogorov, A.N., Fomin, S.V.: Elements of the Theory of Functions and Functional Analysis, vol. 1. Graylock Press, Rochester (1957)
48. Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R., Rohde, G.: Generalized sliced Wasserstein distances. In: Advances in Neural Information Processing Systems, vol. 32, pp. 261–272. Curran Associates, Red Hook (2019)
49. Kufareva, I., Abagyan, R.: Methods of protein structure comparison. Methods Mol. Biol. **857**, 231–257 (2012)
50. Kuo, H.-Y., Su, H.-R., Lai, S.-H., Wu, C.-C.: 3D object detection and pose estimation from depth image for robotic bin picking. In: 2014 IEEE International Conference on Automation Science and Engineering (New Taipei 2014), pp. 1264–1269. IEEE (2014)
51. Lafond, M., El-Mabrouk, N., Huber, K.T., Moulton, V.: The complexity of comparing multiply-labelled trees by extending phylogenetic-tree metrics. Theoret. Comput. Sci. **760**, 15–34 (2019)
52. Lambert, A., Uribe Bravo, G.: The comb representation of compact ultrametric spaces. p-Adic Numbers Ultrametric Anal. Appl. **9**(1), 22–38 (2017)
53. Le, T., Ho, N., Yamada, M.: Computationally Efficient Tree Variants of Gromov–Wasserstein (2019). arXiv:1910.04462
54. Le, T., Yamada, M., Fukumizu, K., Cuturi, M.: Tree-sliced variants of Wasserstein distances. In: 33rd Conference on Neural Information Processing Systems (Vancouver 2019), pp. 12304–12315. Curran Associates, Red Hook (2019)
55. Liebscher, V.: New Gromov-inspired metrics on phylogenetic tree space. Bull. Math. Biol. **80**(3), 493–518 (2018)
56. Lowe, D.G.: Local feature view clustering for 3D object recognition. In: 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Kauai 2001), pp. I–I. IEEE (2001)
57. Mallows, C.L.: A note on asymptotic joint normality. Ann. Math. Stat. **43**, 508–515 (1972)
58. McGregor, A., Stubbs, D.: Sketching Earth-Mover distance on graph metrics. In: Approximation, Randomization, and Combinatorial Optimization (Berkeley 2013). Lecture Notes in Computer Science, vol. 8096, pp. 274–286. Springer, Heidelberg (2013)
59. Mémoli, F.: On the use of Gromov–Hausdorff distances for shape comparison. In: Eurographics Symposium on Point-Based Graphics (Prague 2007). The Eurographics Association (2007). https://doi.org/10.2312/SPBG/SPBG07/081-090

60. Mémoli, F.: Gromov–Wasserstein distances and the metric approach to object matching. Found. Comput. Math. **11**(4), 417–487 (2011)
61. Mémoli, F., Needham, T.: Distance distributions and inverse problems for metric measure spaces. Stud. Appl. Math. **149**(4), 943–1001 (2022)
62. Mémoli, F., Sapiro, G.: Comparing point clouds. In: 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing (Nice 2004), pp. 32–40. ACM, New York (2004)
63. Mémoli, F., Smith, Z., Wan, Z.: The Gromov–Hausdorff distance between ultrametric spaces: its structure and computation. J. Comput. Geom. (to appear). arXiv:2110.03136
64. Mémoli, F., Wan, Z.: On p-metric spaces and the p-Gromov–Hausdorff distance. p-Adic Numbers Ultrametric Anal. Appl. **14**(3), 173–223 (2022)
65. Mézard, M., Parisi, G., Virasoro, M.A.: Spin Glass Theory and Beyond. World Scientific Lecture Notes in Physics, vol. 9. World Scientific, Teaneck (1987)
66. Morozov, D., Beketayev, K., Weber, G.H.: Interleaving distance between merge trees. TopoInVis'13. https://www.mrzv.org/publications/interleaving-distance-merge-trees/manuscript/
67. Nies, T.G., Staudt, T., Munk, A.: Transport dependency: Optimal transport based dependency measures (2021). arXiv:2105.02073 (2021)
68. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Shape distributions. ACM Trans. Graph. **21**(4), 807–832 (2002)
69. Owen, M., Provan, J.S.: A fast algorithm for computing geodesic distances in tree space. IEEE/ACM Trans. Comput. Biol. Bioinform. **8**(1), 2–13 (2011)
70. Papazov, C., Haddadin, S., Parusel, S., Krieger, K., Burschka, D.: Rigid 3D geometry matching for grasping of known objects in cluttered scenes. Intern. J. Robotics Res. **31**(4), 538–553 (2012)
71. Pardalos, P.M., Vavasis, S.A.: Quadratic programming with one negative eigenvalue is NP-hard. J. Global Optim. **1**(1), 15–22 (1991)
72. Peyré, G., Cuturi, M., Solomon, J.: Gromov–Wasserstein averaging of kernel and distance matrices. In: 33rd International Conference on Machine Learning (New York 2016), pp. 2664–2672. JMLR (2016)
73. Qiu, D.: Geometry of non-Archimedean Gromov–Hausdorff distance. p-Adic Numbers Ultrametric Anal. Appl. **1**(4), 317–337 (2009)
74. Rammal, R., Toulouse, G., Virasoro, M.A.: Ultrametricity for physicists. Rev. Mod. Phys. **58**(3), 765–788 (1986)
75. Reeb, G.: Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. C. R. Acad. Sci. Paris **222**, 847–849 (1946)
76. Robinson, D.F.: Comparison of labeled trees with valency three. J. Comb. Theory Ser. B **11**(2), 105–119 (1971)
77. Robinson, D.F., Foulds, L.R.: Comparison of phylogenetic trees. Math. Biosci. **53**(1–2), 131–147 (1981)
78. Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Mover's distance as a metric for image retrieval. Int. J. Comput. Vis. **40**(2), 99–121 (2000)
79. Scetbon, M., Peyré, G., Cuturi, M.: Linear-time Gromov–Wasserstein distances using low rank couplings and costs. In: 39th International Conference on Machine Learning (Baltimore 2022), pp. 19,347–19,365. PMLR (2022)
80. Schmiedl, F.: Computational aspects of the Gromov–Hausdorff distance and its application in non-rigid shape matching. Discrete Comput. Geom. **57**(4), 854–880 (2017)
81. Semmes, S.: An introduction to the geometry of ultrametric spaces (2007). arXiv:0711.0709
82. Semple, C., Steel, M.: Phylogenetics. Oxford Lecture Series in Mathematics and Its Applications, vol. 24. Oxford University Press, New York (2003)
83. Sturm, K.-T.: On the geometry of metric measure spaces. I. Acta Math. **196**(1), 65–131 (2006)
84. Sturm, K.T.: The space of spaces: Curvature bounds and gradient flows on the space of metric measure spaces (2012). arXiv:1208.0434
85. Thorsley, D., Klavins, E.: Model reduction of stochastic processes using Wasserstein pseudometrics. In: 2008 American Control Conference (Seattle 2008), pp. 1374–1381. IEEE (2008)
86. Titouan, V., Courty, N., Tavenard, R., Flamary, R.: Optimal transport for structured data with application on graphs. In: 36th International Conference on Machine Learning (Long Beach 2019), pp. 6275–6284. PMLR (2019)
87. Touli, E.F., Wang, Y.: FPT-algorithms for computing Gromov–Hausdorff and interleaving distances between trees. In: 27th Annual European Symposium on Algorithms (Munich 2019). Leibniz Int. Proc. Inform., vol. 144, # 83. Leibniz-Zent. Inform., Wadern (2019)

88. Vallender, S.S.: Calculation of the Wasserstein distance between probability distributions on the line. Theory Probab. Appl. **18**(4), 784–786 (1974)
89. Vayer, T., Flamary, R., Tavenard, R., Chapel, L., Courty, N.: Sliced Gromov–Wasserstein (2019). arXiv:1905.10124
90. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
91. Villani, C.: Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften, vol. 338. Springer, Berlin (2008)
92. Wan, Z.: A novel construction of Urysohn universal ultrametric space via the Gromov–Hausdorff ultrametric. Topology Appl. **300**, # 107759 (2021)
93. Zarichnyi, I.: Gromov–Hausdorff ultrametric (2005). arXiv:math/0511437