Similarity Between Points in Metric Measure Spaces

Evgeny Dantsin and Alexander Wolpert

Roosevelt University, Chicago, IL, USA
{edantsin,awolpert}@roosevelt.edu

Abstract. This paper is about similarity between objects that can be represented as points in metric measure spaces. A metric measure space is a metric space that is also equipped with a measure. For example, a network with distances between its nodes and weights assigned to its nodes is a metric measure space. Given points x and y in different metric measure spaces or in the same space, how similar are they? A well known approach is to consider x and y similar if their neighborhoods are similar. For metric measure spaces, similarity between neighborhoods is well captured by the Gromov-Hausdorff-Prokhorov distance, but it is NP-hard to compute this distance even in quite simple cases. We propose a tractable alternative: the *radial distribution distance* between the neighborhoods of x and y. The similarity measure based on the radial distribution distance is coarser than the similarity based on the Gromov-Hausdorff-Prokhorov distance but much easier to compute.

Keywords: metric measure space · Gromov-Hausdorff-Prokhorov distance · radial distribution.

1 Introduction

A *metric measure space* is a metric space that is also equipped with a measure. Such spaces play an important role in geometry, especially after Gromov's works [8], and they have proven to be useful in other areas of mathematics, for example, in optimization theory [2] and in probability theory [4]. Metric measure spaces are also used to model real-world systems and processes, for example, in image recognition [11], in genetics [15], in machine learning [3], etc.

A natural example of a metric measure space is the set $\mathbb{R}$ of real numbers with the standard distance $|x-y|$ between point x and y and with the Lebesgue measure on $\mathbb{R}$. Another example is given by a connected graph G in which all vertices and edges are labeled with numbers: the number assigned to a vertex is its "weight" and the number assigned to an edge is its "length". The corresponding metric space is formed by the set of all vertices with the shortest path metric in G: the distance between two vertices is the length of a shortest path between them. A measure on this space is defined on all subsets of the vertices: the measure of a subset A is the total weight of all vertices of A. The weights and lengths can be interpreted in various ways. For example, if G is a communication network, then the weight of a vertex can describe the traffic at this vertex. One more example: if G is a propagation network in epidemic models, then the weight can be the number of infected individuals.

Consider objects that can be modeled by points in metric measure spaces: suppose one object is represented by a point x in a space $\mathcal{X}$ and another object is represented by a point y in a space $\mathcal{Y}$. How similar are these objects? How can we measure similarity between them if the only information we have is two pairs $(\mathcal{X}, x)$ and $(\mathcal{Y}, y)$? Such pairs are called *rooted metric measure spaces* or *rooted mm spaces* for short, see Sections 2 and 3 for precise definitions. In this paper we address the question of similarity between objects modeled by rooted mm spaces.

The most obvious type of similarity between $(\mathcal{X}, x)$ and $(\mathcal{Y}, y)$ is an *isomorphism* between them, which means that there is a bijection from $\mathcal{X}$ to $\mathcal{Y}$ that maps x to y and preserves the metric and measure. This is an "all or nothing" measure of similarity: any two rooted mm spaces are either similar or not. Clearly, this measure is not a good solution for applications because real-world objects, like social, biological, or technological networks, are very rarely, if ever, isomorphic to one another.

Can we improve the isomorphism-based approach to make it more flexible? How could we measure to what extent $(\mathcal{X}, x)$ and $(\mathcal{Y}, y)$ look isomorphic? The concept of "approximate isomorphism" between rooted mm spaces can be implemented using the idea proposed by Edwards [6] and Gromov [7]. To compare $(\mathcal{X}, x)$ and $(\mathcal{Y}, y)$, we embed $\mathcal{X}$ and $\mathcal{Y}$ into another metric measure space $\mathcal{Z}$ and compare their images in $\mathcal{Z}$. More exactly, we take embeddings f and g that preserve the metric and measure and compare the images $f(\mathcal{X})$ and $g(\mathcal{Y})$ in $\mathcal{Z}$. We consider $(\mathcal{X}, x)$ and $(\mathcal{Y}, y)$ *similar* if their images are close to each other in the following sense:

- the point $f(x)$ is close to the point $g(y)$ in the space $\mathcal{Z}$;
- the set of points of $f(\mathcal{X})$ is close to the set of points of $g(\mathcal{Y})$ in the space $\mathcal{Z}$;
- the measures induced by f and g in $\mathcal{Z}$ are close to one another.

The second condition is formalized using the *Hausdorff distance* and the third condition is formalized using the *Lévy-Prokhorov distance*, see Section 2. Taking the infimum over all possible spaces $\mathcal{Z}$ and embeddings f and g, we obtain a distance function on rooted mm spaces called the *Gromov-Hausdorff-Prokhorov distance* (the *GHP distance* for short).

Both the isomorphism-based similarity measure and the GHP-based similarity measure have the following disadvantage for applications. Most real-world systems have the distance decay effect, also called the gravity model, which is often expressed as "all things are related, but near things are more related than far things". For example, when comparing points x and y in metric spaces, the role of their local neighborhoods is more important than the role of points that are far away from x and y. However, neither the isomorphism approach nor the GHP distance capture this effect: all points are considered equally important, independently of their distance from x and y.

This disadvantage is eliminated using the distance defined in [1]. Loosely speaking, this distance between two rooted mm spaces combines the GHP distance with an exponential decay: a point is taken into account with a weight that exponentially decreases with increasing its distance from the root. We call it the *neighborhood-based* distance and describe it in Section 3.

Intuitively, the neighborhood-based distance is the best approach to capture similarity between points in metric measure spaces. To put this distance to work in practical applications, we need to compute it efficiently. However, it is NP-hard to compute the neighborhood-based distance, which follows from [12]. Moreover, under standard complexity-theoretic assumptions, it is not possible to approximate it with a reasonable factor in polynomial time [16].

In Section 4, we propose a tractable alternative to the neighborhood-based distance: namely, we define the *radial distribution distance* between rooted mm spaces. This distance can be viewed as a coarser variant of the neighborhood-based distance or, more exactly, as a lower bound on the neighborhood-based distance. The advantage of the radial distribution distance is that it can be computed efficiently: a straightforward algorithm that computes this distance between finite rooted mm spaces takes time quasilinear in the total number of points.

What is the idea of radial distribution distance? First, we view a point x in a metric measure space $\mathcal{X}$ as the center of a ball of radius r around x. This ball has its own measure (its "mass") and we consider how such masses change with increasing r. Second, we consider this change of masses with an exponential distance decay, which means that the "contribution" of points exponentially decreases with increasing their distance from x. The radial distribution distance between $(\mathcal{X}, x)$ and $(\mathcal{Y}, y)$ is basically the distance between two functions that describe the change of masses around x in $\mathcal{X}$ and the change of masses around y in $\mathcal{Y}$.

2 Preliminaries

The purpose of this section is to recall basic notions of metric spaces and measure spaces used in the next sections.

Distance functions. Let X be a set and d be a function from $X \times X$ to $[0, \infty)$. Consider the following properties of d: for all $x, y, z \in X$,

(a) $d(x, y) = d(y, x)$;
(b) $d(x, x) = 0$;
(c) $d(x, y) = 0 \;\Rightarrow\; x = y$;
(d) $d(x, z) \le d(x, y) + d(y, z)$.

We call d a *distance function* if it has properties (a) and (b). We call d a *metric* and call the pair (X, d) a *metric space* if d has all properties (a)–(d). If d is a distance function that has property (d), but not necessarily (c), we say that d is a *pseudometric* on X.

Speaking about metric spaces, it is common to refer to the elements of X as *points*. The *distance* between points $x, y \in X$ is the number $d(x, y)$. We use the following notation for open and closed balls in metric spaces: for every number $r \in [0, \infty)$ and every point $x \in X$,

- $B_r(x) = \{y \in X \mid d(x, y) < r\}$ is an *open ball* of radius r around x;
- $\overline{B}_r(x) = \{y \in X \mid d(x, y) \leq r\}$ is a *closed ball* of radius r around x.

Types of metric spaces. A subset $A \subseteq X$ is *bounded* in (X, d) if there exists $r \in [0, \infty)$ such that $d(a, b) \leq r$ for all $a, b \in A$. The *diameter* of A is the supremum of $d(a, b)$ taken over all $a, b \in A$. We call (X, d) *totally bounded* if for every $r \in (0, \infty)$, there exists a finite set of open balls of radius r such that their union contains X. Note that if (X, d) is totally bounded then X is bounded, but not conversely: for example, an infinite set X with $d(x, y) = 1$ for all $x \neq y$ is bounded, but not totally bounded.

A metric space (X, d) is *complete* if every Cauchy sequence of points in X converges to a point of X. A complete metric space (X, d) is *compact* if it is totally bounded. A subset $A \subseteq X$ is *compact* if the metric space $(A, d|_A)$ is compact. A metric space (X, d) is *separable* if X has a countable subset $A \subseteq X$ such that every open ball contains an element of A.

Isometries. Let (X, d) and (X', d') be metric spaces. A function $f : X \to X'$ is called an *isometry* from (X, d) to (X', d') if it preserves distances: for all $x_1, x_2 \in X$,

$$d'(f(x_1), f(x_2)) = d(x_1, x_2).$$

Note that, according to this definition, any isometry is injective (isometries are sometimes called *isometric embeddings*).

Hausdorff distance. The Hausdorff distance in a metric space (X, d) can be viewed as an extension of d from points to nonempty subsets of X. In this paper, we restrict the Hausdorff distance to bounded subsets, so it takes only finite values.

Let S be a nonempty subset of points in a metric space (X, d). For every number $r > 0$, the *r-neighborhood* of S, denoted by $N_r(S)$, is the union of open balls of radius r around the points of S:

$$N_r(S) = \bigcup_{s \in S} B_r(s) = \{x \in X \mid \inf_{s \in S} d(x, s) < r\}.$$

The *Hausdorff distance function*, denoted d_H, is defined by

$$d_H(S_1, S_2) = \inf\{r \in [0, \infty) \mid S_1 \subseteq N_r(S_2) \text{ and } S_2 \subseteq N_r(S_1)\}$$

where S_1 and S_2 are nonempty bounded subsets of X. The distance function d_H is a pseudometric on the set of all nonempty bounded subsets of X and d_H is a metric on the set of all nonempty compact subsets of X.

Gromov-Hausdorff distance. As said above, the Hausdorff distance d_H in (X, d) extends d from points to subset of points. The Gromov-Hausdorff distance extends d_H from subsets of X to different metric spaces. Speaking informally, metric spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ are close to each other if they can be isometrically embedded into another metric space $\mathcal{X}$ so that the Hausdorff distance between their images in $\mathcal{X}$ is small. This idea is formalized as follows. Let $\mathcal{X}_1 = (X_1, d_1)$ and $\mathcal{X}_2 = (X_2, d_2)$ be compact metric spaces. Consider a metric space $\mathcal{Y}$ and two isometries:

$$\begin{aligned} f_1 &: \mathcal{X}_1 \to \mathcal{Y} \\ f_2 &: \mathcal{X}_2 \to \mathcal{Y} \end{aligned}$$

The *Gromov-Hausdorff distance function*, denoted d_{GH}, is defined by

$$d_{GH}(\mathcal{X}_1, \mathcal{X}_2) = \inf_{\mathcal{Y}, f_1, f_2} \{d_H(f_1(X_1), f_2(X_2))\}$$

where d_H is the Hausdorff distance function in $\mathcal{Y}$ and the infimum is taken over all possible metric spaces $\mathcal{X}$ and isometries f_1 and f_2. An equivalent definition of the Gromov-Hausdorff distance can be given in terms of *correspondence* and *distortion*, see for example [5, section 7.3]

Measures. Let X be a set and $\mathcal{A}$ be a *σ-algebra* on X, i.e., $\mathcal{A}$ is a family of subsets of X such that this family contains X itself and it is closed under complements and countable unions. A function $\mu : \mathcal{A} \to [0, \infty]$ is called a *measure* if $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

for all sequences $\{A_i\}_{i=1}^{\infty}$, where $A_i \in \mathcal{A}$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$. The triplet $(X, \mathcal{A}, \mu)$ is called a *measure space.* Every element of $\mathcal{A}$ is called a *measurable set.* If $\mu(X)$ is finite then the measure μ and the measure space $(X, \mathcal{A}, \mu)$ are called *finite.*

Measurable functions. Consider two sets X and X' equipped with σ-algebras $\mathcal{A}$ and $\mathcal{A}'$ respectively. A function $f : X \to X'$ is called *measurable* if for every measurable subset $S \in \mathcal{A}'$, its preimage $f^{-1}(S)$ is also measurable: $f^{-1}(S) \in \mathcal{A}$. For every measure $\mu : \mathcal{A} \to [0, \infty]$ and for every measurable function $f : X \to X'$, the composition $\mu \circ f^{-1}$ maps every measurable subset $S \in \mathcal{A}'$ to the number $\mu(f^{-1}(S))$. This function is a measure on $(X', \mathcal{A}')$ called the *push-forward measure.*

Borel measures. For a metric space (X, d), let $\mathcal{B}$ be the smallest σ-algebra that contains all open balls (or equivalently all closed balls) of this space. The set $\mathcal{B}$ is called the *Borel* σ-algebra on (X, d). Every function $\mu : \mathcal{B} \to [0, \infty]$ is called a *Borel measure* on (X, d).

Lévy-Prokhorov distance. Let (X, d) be a metric space with the Borel σ-algebra $\mathcal{B}$. Let μ_1 and μ_2 be finite Borel measures on (X, d). The *Lévy-Prokhorov distance* between these measures (sometimes called the *Prokhorov distance*), denoted π, is defined by

$$\begin{aligned}\pi(\mu_1, \mu_2) = \inf\{r \in [0, \infty) \mid\ & \mu_1(S) \le \mu_2(N_r(S)) + r \text{ and} \\ & \mu_2(S) \le \mu_1(N_r(S)) + r \\ & \text{for all } S \in \mathcal{B}\}\end{aligned}$$

Thus, the Lévy-Prokhorov distance can be viewed as a measure-theoretic analogue of the Hausdorff distance.

3 Neighborhood-Based Distance

A *metric measure space* is usually defined as a complete separable metric space with a Borel measure on this metric space. However, in this paper, we deal with only metric measure spaces where the metric space is compact and the measure is finite. Therefore, to simplify terminology, we use the term "metric measure space" to refer to this restricted case.

Definition 1 (mm space). *A* metric measure space *(an* mm space *for short) is a triplet (X, d, μ) where (X, d) is a compact metric space and μ is a Borel measure on (X, d).*

Definition 2 (rooted mm space). *A* rooted mm space *is a pair $(\mathcal{X}, o)$ where $\mathcal{X}$ is an mm space and o is a designated point in $\mathcal{X}$ called the* origin *of $\mathcal{X}$.*

Given two rooted mm spaces, how similar are they? A natural approach to comparing rooted mm spaces is to combine the idea of Gromov-Hausdorff distance with the idea of Lévy-Prokhorov distance. This combination was introduced in [13] and has slight variations. The following definition is the version from [9].

Definition 3 (GHP distance). *Let $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$ be rooted mm spaces with $\mathcal{X}_1 = (X_1, d_1, \mu_1)$ and $\mathcal{X}_2 = (X_2, d_2, \mu_2)$. The* Gromov-Hausdorff-Prokhorov distance *(the* GHP distance *for short), denoted d_{GHP}, is defined by*

$$\begin{aligned}&d_{GHP}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) = \\ &\quad \inf\nolimits_{\mathcal{Y}, f_1, f_2}\{d(o_1, o_2) + d_H(f_1(X_1), f_2(X_2)) + \pi(\mu_1 \circ f_1^{-1}, \mu_2 \circ f_2^{-1})\}\end{aligned}$$

where the infimum is taken over all mm spaces $\mathcal{Y}$ and all measurable isometries $f_1 : \mathcal{X}_1 \to \mathcal{Y}$ and $f_2 : \mathcal{X}_2 \to \mathcal{Y}$. Also, d, d_H, and π denote respectively the distance, the Hausdorff distance, and the Lévy-Prokhorov distance in $\mathcal{Y}$.

The definition above is a straightforward extension of the original "non-rooted" GHP distance to rooted mm spaces. Note that the non-rooted GHP distance is not the only distance function for metric measure spaces. Another distance is the Gromov-Wasserstein distance [11], but this notion is much less amenable for extension to rooted mm spaces.

As noted in section 1, the GHP distance has the disadvantage that d_{GHP} does not capture the distance decay effect occurring in most real-world systems. This disadvantage is eliminated in the distance defined in [1]. We define a simplified version of this distance below and call it *neighborhood-based distance*. Its idea can be informally described as follows. When comparing rooted mm spaces $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$, we consider the restrictions of $\mathcal{X}_1$ and $\mathcal{X}_2$ to closed balls of radius r around o_1 and o_2. For every radius r, we consider the GHP distance between the corresponding restrictions and sum up these distances over all values of r with an exponential decrease when r increases.

Let $\overline{B}$ be a closed ball in an mm space $\mathcal{X} = (X, d, \mu)$. This ball, along with the metric and measure obtained by restricting d and μ to $\overline{B}$, form the mm space called the *restriction* of $\mathcal{X}$ to $\overline{B}$.

Definition 4 (neighborhood-based distance). *Let $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$ be rooted mm spaces. For every number $r \in [0, \infty)$, let δ_r denote the GHP distance between the restriction of $\mathcal{X}_1$ to $\overline{B}_r(o_1)$ and the restriction of $\mathcal{X}_2$ to $\overline{B}_r(o_2)$. The* neighborhood-based distance*, denoted d_{nb}, is defined by*

$$d_{nb}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) = \int_0^\infty e^{-r}\, \delta_r \; dr$$

It follows from the definition that

$$d_{nb}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) = 0$$

for isomorphic $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$ even if they are different. Therefore, d_{nb} is not a metric but, as shown below, d_{nb} is a pseudometric.

Theorem 1. *The distance function d_{nb} is a pseudometric on the set of rooted mm spaces.*

Proof. Proposition 5.3 in [1] is a similar statement about the distance defined there. Though that distance is slightly different from d_{nb}, the proof of Proposition 5.3 is essentially a proof of our theorem. (The distance in [1] and d_{nb} differ in the integral expression: δ_r in our definition is replaced with $\max(1, \delta_r)$ in [1]. However, this difference does not affect the proof.) □

4 Radial Distribution Distance

Can we compute the neighborhood-based distance d_{nb} efficiently? Note that an efficient algorithm for computing d_{nb} could also be used to compute the Gromov-Hausdorff distance efficiently. However, as shown in [10, 12], it is NP-hard to compute the Gromov-Hausdorff distance for finite metric spaces (see Proposition 3.16 in [12]). Moreover, under standard complexity-theoretic assumptions, there is no polynomial-time approximation algorithm with a reasonable factor for computing this distance [16].

In this section, we define another distance between rooted mm spaces called the *radial distribution distance* and denoted by d_{rd}. On the one hand, d_{rd} is coarser than d_{nb}, more exactly, d_{rd} is a lower bound on d_{nb}. On the other hand, d_{rd} can be computed efficiently: computing the radial distribution distance between finite rooted mm spaces takes time quasilinear in the total number of points.

4.1 Definition and Properties

Consider rooted mm spaces $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$ where $\mathcal{X}_1 = (X_1, d_1, \mu_1)$ and $\mathcal{X}_2 = (X_2, d_2, \mu_2)$. To define the *radial distribution distance* between them, we first define the following functions m_1 and m_2 form $[0, \infty)$ to itself: for every number $r \in [0, \infty)$,

$$\begin{aligned} m_1(r) &= \mu_1\left(\overline{B}_r(o_1)\right) = \mu_1\left(\{x \in X_1 \mid d_1(x, o_1) \le r\}\right) \\ m_2(r) &= \mu_2\left(\overline{B}_r(o_2)\right) = \mu_2\left(\{x \in X_2 \mid d_2(x, o_2) \le r\}\right) \end{aligned}$$

That is, $m_1(r)$ is the measure (we could call it “mass” or “weight”) of the ball of radius r around the origin o_1 and, similarly, for $m_2(r)$. The functions are non-decreasing and, since X_1 and X_2 are compact, they are bounded. Also, m_1 and m_2 are càdlàg functions, i.e., they are right continuous with left limits, see Lemma 2.8 in [1]. Therefore, the distance function in the definition below is well defined.

Definition 5 (radial distribution distance). *Let $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$ be rooted mm spaces. The* radial distribution distance*, denoted d_{rd}, is defined by*

$$d_{rd}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) = \int_0^\infty e^{-r} \left|m_1(r) - m_2(r)\right| dr$$

Theorems 2-4 show basic properties of the radial distribution distance.

Theorem 2. *The function d_{rd} is a pseudometric on the set of rooted mm spaces.*

Proof. It is obvious that d_{rd} is a distance function. The triangle inequality for d_{rd} can be seen from the following two facts:

- The functions $e^{-r}m_1(r)$ and $e^{-r}m_2(r)$ are measurable functions on $[0, \infty)$ such that the integrals $\int_0^\infty e^{-r}\, m_1(r)\, dx$ and $\int_0^\infty e^{-r}\, m_2(r)\, dx$ are finite.
- The distance function

  $$d(f_1, f_2) = \int_0^\infty |f_1(x) - f_2(x)|\, dx$$

  is the L^1 metric on the set of measurable functions f such that the integral $\int_0^\infty |f(x)|\, dx$ is finite.

□

Theorem 3. *For all rooted mm spaces* $(\mathcal{X}_1, o_1)$ *and* $(\mathcal{X}_2, o_2)$*,*

$$d_{rd}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) \leq d_{nb}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)). \tag{1}$$

Proof. Let $\overline{B}_r(o_1)$, $\overline{B}_r(o_2)$, and δ_r be the same as in Definition 4. Let μ_1 and μ_2 be the measures in $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively. We prove the following inequality that implies claim (1):

$$|\mu_1(\overline{B}_r(o_1)) - \mu_2(\overline{B}_r(o_2))| \leq \delta_r. \tag{2}$$

By the definition of the GHP distance, we have

$$\delta_r \geq \pi(\mu_1 \circ f_1^{-1}, \mu_2 \circ f_2^{-1}) \tag{3}$$

where f_1 and f_2 are arbitrary measurable isometries of the corresponding restrictions to an arbitrary mm space $\mathcal{Y}$ and π is the Lévy-Prokhorov distance in $\mathcal{Y}$. By the definition of π,

$$\pi(\mu_1 \circ f_1^{-1}, \mu_2 \circ f_2^{-1}) = \gamma \tag{4}$$

where γ is the infimum of all $\epsilon \geq 0$ such that

$$\begin{cases} \mu_1\left(\overline{B}_r(o_1)\right) \leq \mu_2\left(\overline{B}_{r+\epsilon}(o_2)\right) + \epsilon \\ \mu_2\left(\overline{B}_r(o_2)\right) \leq \mu_1\left(\overline{B}_{r+\epsilon}(o_1)\right) + \epsilon \end{cases}$$

Both inequalities above hold if we take

$$\epsilon = |\mu_1(\overline{B}_r(o_1)) - \mu_2(\overline{B}_r(o_2))|$$

and, therefore, we have

$$\gamma \geq |\mu_1(\overline{B}_r(o_1)) - \mu_2(\overline{B}_r(o_2))|. \tag{5}$$

Now, combining (3)–(5), we obtain inequality (2). □

The theorem above shows that d_{rd} is a lower bound on d_{nb}. This bound is strict, which can be seen from the following simple example. Consider a rooted mm space on a set of three points: $\{a, b, c\}$ where a is the origin. The distance between any two points is 1. The measure assigned to each point is 1. Consider another rooted mm space that differs from the first one in only the measure of b and c: in the second space, the measure of b is 0.5 and the measure of c is 1.5. It is easy to see that the radial distribution distance between these spaces is 0, while the neighborhood distance is not zero.

Theorem 4. *Let* $\mathcal{X}_1 = (X_1, d_1, \mu_1)$ *and* $\mathcal{X}_2 = (X_2, d_2, \mu_2)$ *be mm spaces. Let* D_1 *and* D_2 *be the diameters of* $\mathcal{X}_1$ *and* $\mathcal{X}_2$ *respectively. For all origins* $o_1 \in X_1$ *and* $o_2 \in X_2$*,*

$$d_{nb}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) \leq D_1 + D_2 + \mu_1(X_1) + \mu_2(X_2).$$

Proof. Consider a trivial mm space on a one-element set whose measure is zero. Let o be the single element and let $(\{o\}, o)$ denote the corresponding rooted mm space. What is the GHD distance between $(\mathcal{X}_1, o_1)$ and $(\{o\}, o)$? There are two trivial embeddings: the identity map of $(\mathcal{X}_1, o_1)$ to itself and the embedding of $(\{o\}, o)$ to $(\mathcal{X}_1, o_1)$ that preserves the origin. Therefore, by the definition of the GHP distance,

$$d_{GHP}((\mathcal{X}_1, o_1), (\mathcal{X}_1^0, o_1)) \le 0 + D_1 + \mu_1(X_1).$$

Similarly, we have

$$d_{GHP}((\mathcal{X}_2, o_2), (\mathcal{X}_2^0, o_2)) \le 0 + D_2 + \mu_2(X_2).$$

and, by the triangle inequality,

$$d_{GHP}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) \le D_1 + D_2 + \mu_1(X_1) + \mu_2(X_2).$$

Hence,

$$\begin{aligned} d_{nb}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) &\le \textstyle\int_0^\infty e^{-r}\,(D_1 + D_2 + \mu_1(X_1) + \mu_2(X_2))\; dr \\ &= D_1 + D_2 + \mu_1(X_1) + \mu_2(X_2). \end{aligned}$$

□

4.2 Computing the Radial Distribution Distance

We describe a straightforward algorithm that computes the radial distribution distance between finite rooted mm spaces efficiently. In the description, we ignore all problems of numerical analysis and deal with real numbers. The algorithm is illustrated by applying it to the following example of two rooted mm spaces $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$.

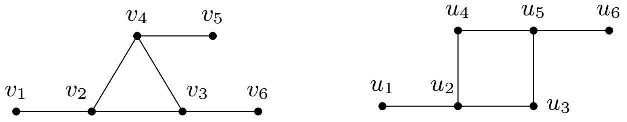

Fig. 1. Graphs for $(\mathcal{X}_1, o_1)$ and $(\mathcal{X}_2, o_2)$

Example. We define $(\mathcal{X}_1, o_1)$, where $\mathcal{X}_1 = (X_1, d_1, \mu_1)$, using the graph on the left in Fig. 1: namely, $X_1 = \{v_1, \ldots, v_6\}$ and $d_1(v_i, v_j)$ is the length of a shortest path between v_i and v_j. The measure μ_1 is defined by assigning numerical values to all one-element subsets of X_1; this assignment is shown in the second column of the table in Fig. 2. The origin o_1 is defined to be v_1.

The rooted mm space $(\mathcal{X}_2, o_2)$, where $\mathcal{X}_2 = (X_2, d_2, \mu_2)$, is defined similarly using the graph on the right in Fig. 1. The underlying set X_2 consists of the vertices $u_1, \ldots, u_6$ with the shortest-path metric. The measure μ_2 is given by the forth column of the table in Fig. 2 and the origin o_2 is defined to be u_1.

Algorithm. The radial distribution distance between finite rooted mm spaces can be computed using only part of information about the input spaces. The algorithm uses the distances between the origins and all other points; it does not use the distances between non-origin points. In terms of matrices, if a distance function in a rooted mm space with n points is represented by an $n \times n$ matrix, then the algorithm uses only one row (or column) corresponding to the origin. Therefore, we assume that the algorithm takes as input only the following:

- the distances between the origins and non-origins;
- the measures of all one-element sets.

When the algorithm is applied to our example, it takes as input the table in Fig. 2.

rooted mm space $(\mathcal{X}_1, o_1)$		rooted mm space $(\mathcal{X}_2, o_2)$	
distances	measures	distances	measures
$d_1(v_1, v_1) = 0$	$\mu_1(\{v_1\}) = 1$	$d_2(u_1, u_1) = 0$	$\mu_2(\{u_1\}) = 1$
$d_1(v_1, v_2) = 1$	$\mu_1(\{v_2\}) = 1$	$d_2(u_1, u_2) = 1$	$\mu_2(\{u_2\}) = 1$
$d_1(v_1, v_3) = 2$	$\mu_1(\{v_3\}) = 2$	$d_2(u_1, u_3) = 2$	$\mu_2(\{u_3\}) = 2$
$d_1(v_1, v_4) = 2$	$\mu_1(\{v_4\}) = 2$	$d_2(u_1, u_4) = 2$	$\mu_2(\{u_4\}) = 2$
$d_1(v_1, v_5) = 3$	$\mu_1(\{v_5\}) = 1$	$d_2(u_1, u_5) = 3$	$\mu_2(\{u_5\}) = 1$
$d_1(v_1, v_6) = 3$	$\mu_1(\{v_6\}) = 1$	$d_2(u_1, u_6) = 4$	$\mu_2(\{u_6\}) = 1$

Fig. 2. Input for the algorithm

Here is a sketch of the algorithm:

1. Compute the set R_1 of all values of radius in $(\mathcal{X}_1, o_1)$:

$$R_1 \;=\; \{r \mid r = d_1(o_1, x_1) \text{ for some point } x_1 \text{ in } \mathcal{X}_1\}$$

   where d_1 is the metric in $\mathcal{X}_1$.
2. Compute the set R_2 of all values of radius in $(\mathcal{X}_2, o_2)$ defined similarly.
3. Compute the union $R = R_1 \cup R_2$. In our example, $R = \{0, 1, 2, 3, 4\}$.
4. Compute the *cumulative radial distribution* of $(\mathcal{X}_1, o_1)$, i.e., the set of pairs $(r, m_1(r))$ where $r \in S$ and $m_1(r) = \mu_1\left(\overline{B}_r(o_1)\right)$. In our example, this set is

$$\{(r, m_1(r)\}_{r \in R} = \{(0, 1), (1, 2), (2, 6), (3, 8), (4, 8)\}.$$

5. Compute the cumulative radial distribution of $(\mathcal{X}_2, o_2)$ defined similarly. In our example, this set is

$$\{(r, m_2(r)\}_{r \in R} = \{(0, 1), (1, 2), (2, 6), (3, 7), (4, 8)\}.$$

6. Compute the radial distribution distance by

$$d_{rd}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) = \sum_{r \in R} e^{-r} |m_1(r) - m_2(r)|.$$

Applying this to our example, we obtain

$$d_{rd}((\mathcal{X}_1, o_1), (\mathcal{X}_2, o_2)) = e^{-3} \cdot 1 \approx 0.05.$$

Note that the algorithm takes $O(n \log n)$ steps where n is the number of points in the input spaces.

Considering the point v_1 in $\mathcal{X}_1$, what points in $\mathcal{X}_2$ are most "similar" to v_1 and what points are most "dissimilar" to it? The calculations show that u_1 and u_6 are most similar to v_1 (with $d_{rd} \approx 0.05$), while the points u_3 and u_4 are most dissimilar to v_1 (with $d_{rd} \approx 2.01$).

4.3 Extension for Feature Vectors

The radial distribution distance d_{rd} can be used to measure similarity between points in a metric space if points are described with a single feature: a value of this feature for a given point is viewed as the point's "weight". However, it is more typical that a point is described by a feature vector rather than a single feature. Each component of the vector corresponds to a measure on the metric space. For example, consider a recommender system for movies that uses item-item collaborative filtering. A metric on movies is based on the similarity between them calculated using people's ratings. In addition to the metric, each movie is described by a feature vector that can include, for example, the number of reviews, budget, box office, etc. [14]

How can we compare points in a metric space if points are described using feature vectors? Suppose a feature vector consists of k components that correspond to measures $\mu_1, \ldots, \mu_k$. Each measure μ_i determines the radial distribution distance $d_{rd}^{(i)}$ and we can consider their sum

$$d_{rd}^{*} = d_{rd}^{(1)} + \ldots + d_{rd}^{(k)}. \qquad (6)$$

The value $d_{rd}^{*}((\mathcal{X}, x), (\mathcal{Y}, y))$ essentially shows the distance between the neighborhoods of x and y if we compare points by their feature vectors. Note that the sum of pseudometrics is also a pseudometric. Also note that instead of the sum in (6), we could take any other norm, for example, the Euclidean norm or the maximum.

5 Summary

In this paper we consider how to define and measure similarity between a point x in a metric measure space $\mathcal{X}$ and a point y in the same space or another metric measure space $\mathcal{Y}$. This problem was studied in the mathematical context, but not in the context of applications where a system under consideration, for example a network, can be modeled by a metric measure space.

Our main approach is to define x and y similar if their neighborhoods are similar, but what is a neighborhood of a point in a metric measure space and what is the similarity between such neighborhoods? We define the *neighborhood-based distance* d_{nb} and, in fact, this definition suggests answers to both questions. The distance d_{nb} combines two equally important ideas:

– The neighborhood of x contains all points of $\mathcal{X}$ but their "contributions" are different: the contribution of a point decreases exponentially with increasing its distance from x. Thus, d_{nb} supports the *distance decay effect*.
– The similarity between metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ is defined using the *Gromov-Hausdorff-Prokhorov distance*. This notion can be viewed as an "approximation" of isomorphism: the GHP distance between $\mathcal{X}$ and $\mathcal{Y}$ shows how far they are from being isomorphic.

The neighborhood-based distance would be perhaps the best approach to capture similarity between x and y, but it has a serious disadvantage: no efficient algorithm for computing d_{nb} is known. It is at least NP-hard to compute d_{nb} and, moreover, it is unlikely that there exists an efficient approximation algorithm for computing d_{nb}.

To overcome this difficulty, we propose a replacement of the intractable part of d_{nb}. We define the *radial distribution distance* d_{rd} that can be computed efficiently. We show that the radial distribution distance is a lower bound on the neighborhood-based distance and we consider d_{rd} as a "good" bound: it captures similarity between $(x, \mathcal{X})$ and $(y, \mathcal{Y})$ well enough.

We use the term "radial distribution distance" keeping in mind radial distribution functions from physics where they are used to measure the probability of finding a particle at distance r from a certain "origin" particle. In the setting of metric measure spaces, a radial distribution function describes how the total "mass" ("weight") of points at distance r from a center x changes when r increases. The radial distribution distance is basically a combination of the L^1 distance between the cumulative versions of the radial distribution functions and an exponential distance decay.

Finally, we show how to use the radial distribution distance for measuring similarity between points in a metric space there every point has an associated feature vector.

References

1. Romain Abraham, Jean-Franois Delmas, and Patrick Hoscheit. A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electronic Journal of Probability*, 18(14):1–21, 2013.
2. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. Birkhäuser, 2005.
3. Martín Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017*, pages 214–223, 2017.
4. Siva Athreya, Wolfgang Löhr, and Anita Winter. Invariance principle for variable speed random walks on trees. *Annals of Probability*, 45(2):625–667, 2017.

5. Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001.
6. David A. Edwards. The structure of superspace. In *Studies in Topology*, pages 121–133. Academic Press, 1975.
7. Misha Gromov. Groups of polynomial growth and expanding maps (with an appendix by jacques tits). *Publications Mathématiques de l'IHÉS*, 53:53–78, 1981.
8. Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser, 1999. Based on the 1981 French original.
9. Tao Lei. Scaling limit of random forests with prescribed degree sequences. *Bernoulli*, 25(4A):2409–2438, 2019.
10. Facundo Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. In *Proceedings of the Symposium on Point Based Graphics, Prague, 2007*, pages 81–90, 2007.
11. Facundo Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011.
12. Facundo Mémoli, Zane Smith, and Zhengchao Wan. Gromov-Hausdorff distances on p-metric spaces and ultrametric spaces. *ArXiv e-prints*, December 2019.
13. Grégory Miermont. Tessellations of random maps of arbitrary genus. *Annales Scientifiques de L'École Normale Supérieure*, 42(5):725–781, 2009.
14. Xia Ning, Christian Desrosiers, and George Karypis. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender Systems Handbook*, pages 37–76. Springer, 2015.
15. Raúl Rabadán and Andrew Blumberg. *Topological data analysis for genomics and evolution: topology in biology*. Cambridge University Press, 2019.
16. Felix Schmiedl. Computational aspects of the Gromov-Hausdorff distance and its application in non-rigid shape matching. *Discrete & Computational Geometry*, 57(4):854–880, 2017.