Ollivier Curvature of Random Geometric Graphs Converges to Ricci Curvature of Their Riemannian Manifolds

Pim van der Hoorn[1,2,3,4] ⬤ · **Gabor Lippner**[2] · **Carlo Trugenberger**[3,6] ·
Dmitri Krioukov[2,4,5]

Abstract

Curvature is a fundamental geometric characteristic of smooth spaces. In recent years different notions of curvature have been developed for combinatorial discrete objects such as graphs. However, the connections between such discrete notions of curvature and their smooth counterparts remain lurking and moot. In particular, it is not rigorously known if any notion of graph curvature converges to any traditional notion of curvature of smooth space. Here we prove that in proper settings the Ollivier–Ricci curvature of random geometric graphs in Riemannian manifolds converges to the Ricci curvature of the manifold. This is the first rigorous result linking curvature of random graphs to curvature of smooth spaces. Our results hold for different notions of graph distances, including the rescaled shortest path distance, and for different graph densi-

Editor in Charge: Kenneth Clarkson

Pim van der Hoorn
w.l.f.v.d.hoorn@tue.nl

Gabor Lippner
g.lippner@northeastern.edu

Carlo Trugenberger
ca.trugenberger@bluewin.ch

Dmitri Krioukov
dima@northeastern.edu

[1] Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

[2] Department of Mathematics, Northeastern University, Boston, MA, USA

[3] Department of Physics, Northeastern University, Boston, MA, USA

[4] Northeastern University, Network Science Institute, Boston, MA, USA

[5] Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA

[6] SwissScientific Technologies, Genève, Switzerland

ties. Here the scaling of the average degree, as a function of the graph size, can range from nearly logarithmic to nearly linear.

Keywords Graph curvature · Ollivier–Ricci curvature · Random geometric graphs · Riemannian manifolds

Mathematics Subject Classification 60D05 · 05C80

1 Introduction

Curvature is a fundamental concept in the study of geometric spaces. It is a local parameter whose behavior often controls global phenomena on the manifold. In particular, bounds on the Ricci curvature are known to imply an array of properties, including diameter bounds, control of the spectrum, and sub-Gaussian decay of the heat kernel. If the curvature of some space is upper-bounded by a negative value, then such space has a boundary at infinity and some other universal characteristics of (coarsely) hyperbolic spaces. Unfortunately, most notions of curvature are applicable only to smooth continuous spaces, such as Riemannian and pseudo-Riemannian manifolds. While there exist some *combinatorial* notions of curvature [6, 11], none has the same power as their smooth counterparts. We refer to [25] for a general overview of discrete curvatures. The focus of this paper is graph curvature.

In [27–29], Yann Ollivier introduced a definition of curvature for general metric spaces as a discretization of the well-known Ricci curvature. Since this definition is applicable to any metric space, it is applicable to graphs in particular. Even though relatively recent, it has already proven to be quite influential and fruitful. In analysis of networks, Ollivier–Ricci curvature has been used, for example, to identify communities [36], analyze cancer cells [33], asses the fragility of financial networks [34] and robustness of brain networks [10], and to embed networks for machine learning applications [12]. Ollivier–Ricci curvature has also been analyzed for several types of (random) graphs including Erdős–Rényi random graphs [23]. Some general bounds for this curvature have also been established based on different graph properties [4, 16, 23]. These and other applications of Ollivier–Ricci curvature have also stimulated general interest in graph curvature, leading to the introduction and studies of many other notions of graph curvature [8, 17, 24, 37].

An interesting aspect of Ollivier–Ricci curvature (or any other notion of discrete curvature) is that it creates a bridge between geometry and discrete structures. For example, discrete curvatures play an important role in the field of manifold learning where the discrete objects are data points lying on some manifold, and the task is to learn from the data the properties of the manifold [1].

A related task is that of graph embedding: given a graph, find its embedding in a smooth space such that graph distances between nodes are approximated by distances in the space. Curvature has proven to be important for finding the right space to embed the graph into [12].

In addition to these classical applications, geometry has also proven to be an important and powerful concept for designing latent-space models of random graphs whose

properties—such as degree distributions, clustering, distance distributions—closely resemble those of real-world networks [5, 14, 19, 20]. These relations between geometry and network properties inevitably lead to the question whether characteristics of latent geometries of networks can be inferred from discrete properties of graphs that represent these networks. Since curvature is a fundamental characteristic of geometry, it is a natural first candidate for uncovering latent geometry in networks. Hence, a proper notion of graph curvature is needed, a notion that would be known to converge to the true curvature of the geometric space underlying the graph, if it exists.

Quantum gravity is yet another area where convergence of graph curvature is of interest. Here one wants to find a discrete geometry that converges in the continuum limit to the geometry of physical spacetime. To this end, Ollivier–Ricci curvature and its variations have been extensively investigated recently [7, 18, 40].

Despite the interest in Ollivier–Ricci and related curvatures of discrete and combinatorial spaces, the fundamental question of convergence remains largely open. That is, does there exist a discrete notion of curvature that converges in some limit to a traditional notion of curvature of smooth spaces.

There are some positive results in this direction. One is for the convergence of an angle-defect-based notion of curvature of smooth triangulations of Riemannian manifolds [6]. Another one is a manifold learning method designed for consistent estimation of Ricci curvature of a submanifold in Euclidean space based on a point cloud sprinkled uniformly onto the submanifold [1]. Perhaps the closest result to ours is the one in [2, 3] where a discrete version of the d'Alembertian operator is defined for causal sets in 2- and 4-dimensional Lorentzian manifolds. This discrete d'Alembertian is then shown to converge to the traditional d'Alembertian in the continuum limit. To the best of our knowledge, there currently exist no general convergence results for truly combinatorial objects in general and random graphs in Riemannian manifolds in particular.

In this paper we study the question of convergence of Ollivier–Ricci curvature of graphs. We consider random geometric graphs whose nodes are a Poisson process in a Riemannian manifold and whose edges are formed only between nodes that lie within a given distance threshold from each other in the manifold. We show that as the size of such graphs tends to infinity, their Ollivier–Ricci curvature recovers the Ricci curvature of the underlying manifold. To the best of our knowledge, this is the first result that relates a discrete notion of curvature of graphs to the continuum version of curvature of their underlying geometry.

The remainder of the paper is structured as follows. In the next Sect. 2 we introduce the basic notations and definitions needed to present our main results. We present these results in Sect. 3. That section ends with some general comments and outlook. We then provide a general overview of the proof strategy in the first half of Sect. 4. The second half of that section contains the proofs of the main results. The final Sect. 5 contains all the remaining details and proofs of intermediate results that are skipped in Sect. 4.

2 Notations and Definitions

2.1 Geometric Graphs

Given a metric space $(\mathcal{X}, d)$, a countable *node set* $X \subseteq \mathcal{X}$, and *connection radius* $\varepsilon > 0$, we define $G(X, \varepsilon)$ as the graph whose nodes are all the elements in X. An edge between $x, y \in X$ exists if and only if $d(x, y) \leq \varepsilon$. Since the nodes of G are points in the metric space, we will refer to them using x and y, instead of indices i and j, and write $x \in G$ if x is a node of G.

We will also use G_{xy} to denote the indicator of an edge between x and y in G and define $\mathcal{N}_x$ to be the neighborhood of node x, i.e.,

$$\mathcal{N}_x = \{y \in G : G_{xy} = 1\}.$$

Note that $\mathcal{N}_x = X \cap \mathcal{B}(x; \varepsilon)$, where $\mathcal{B}(x; \varepsilon)$ denotes the closed ball around $x \in \mathcal{X}$ of radius ε with respect to the distance d, but excluding x.

2.2 Random Geometric Graphs

In this paper we consider graphs that are constructed by randomly placing points in the metric space $\mathcal{X}$, according to a Poisson process. In order to analyze a notion of curvature on these graphs we need to impose some additional structure on $\mathcal{X}$. More precisely, we will consider Riemannian manifolds as the spaces on which graphs are constructed. We briefly recall some notions of Riemannian geometry needed for the setup and refer the reader to [15, 30] for more details on the topic.

Formally, a Riemannian manifold is a pair $(\mathcal{M}, g)$ where $\mathcal{M}$ is a smooth manifold and for each $x \in \mathcal{M}$, g_x is a smooth inner product on the tangent space $T_x \mathcal{M}$ at x. This inner product induces a metric $d_{\mathcal{M}}$, called the Riemannian metric. Since we are mainly interested in metric spaces, we denote a Riemannian manifold by the pair $(\mathcal{M}, d_{\mathcal{M}})$.

Throughout the remainder of this paper we work with Riemannian manifolds that are *smooth*, *connected*, and *compact*. This allows us to integrate over points in de manifold and ensures that for any two points $x, y \in \mathcal{M}$ there exists a shortest path (geodesic) in $\mathcal{M}$ connecting x and y, whose length is $d_{\mathcal{M}}(x, y)$. For any $U \subseteq \mathcal{M}$ we will write $\text{vol}_{\mathcal{M}}(U) = \int_U \text{dvol}_{\mathcal{M}}$ to denote the volume of U, where $\text{vol}_{\mathcal{M}}$ is the Riemannian measure on $\mathcal{M}$. With this setup we can define a random geometric graph on a Riemannian manifold in an analogous way to classic random geometric graph in Euclidean space.

Definition 2.1 Let $(\mathcal{M}, d_{\mathcal{M}})$ be a smooth, connected, and compact N-dimensional Riemannian manifold. Fix $\varepsilon > 0$ and consider a Poisson process $\mathcal{P}_n$ on $\mathcal{M}$ with intensity measure $(n/\text{vol}_{\mathcal{M}}(\mathcal{M})) \, \text{dvol}_{\mathcal{M}}$. Then we define the *random geometric graph* $\mathbb{G}_n(\varepsilon) := G(\mathcal{P}_n, \varepsilon)$.

Remark 1 *(conditions on the manifold)* From a technical perspective, we only need the manifold to be smooth. This is because we will be working on shrinking neighborhoods of some fixed point $x^* \in \mathcal{M}$. For a sufficiently small neighborhood U, we

can always construct a volume form that is well defined on U and ensure that every two points in U are connected by a geodesic path. We could then fix a sufficiently small and compact neighborhood $\mathcal{C}$ of x^* and then consider a Poisson process on $\mathcal{C}$ with intensity measure $(n/\mathrm{vol}_{\mathcal{M}}(\mathcal{C}))\,\mathrm{dvol}_{\mathcal{M}}$.

The only difference with the global setup is that we would need to frame everything in terms of sufficiently small neighborhoods and deal with possible boundary issues in our proofs. In the end, since curvature is a local property, these issues would vanish. Still, framing all results in this *local* setting would add additional technical layers to the proofs. For convenience, we therefore choose to present everything in terms of global and nice requirements on the manifold.

We shall next introduce a notion of curvature on random geometric graphs. Since curvature is inherently a local property, it makes sense to define curvature on graphs as a property of an edge. For our analysis we will take a more general approach and consider curvature between two fixed nodes in the graph that are connected by a path. We then analyze its behavior as the size of the graph tends to infinity.

For any $x \in \mathcal{M}$, we write $\mathbb{G}_n(x, \varepsilon) := G(X_n, \varepsilon)$, where $X_n = \{x\} \cup \mathcal{P}_n$. That is, $\mathbb{G}_n(x, \varepsilon)$ is a random geometric graph with x added to the node set. Similarly, for any pair of points $(x, y) \in \mathcal{M}$ we write $\mathbb{G}_n(x, y, \varepsilon) := G(X_n', \varepsilon)$, with $X_n' = \{x, y\} \cup \mathcal{P}_n$. We refer to both $\mathbb{G}_n(x, \varepsilon)$ and $\mathbb{G}_n(x, y, \varepsilon)$ as *rooted random graphs*.

2.3 Ollivier–Ricci Curvature on Graphs

The definition of Ollivier–Ricci curvature uses the Wasserstein metric (transportation distance), which we shall introduce next. Recall that a coupling between two probability measures μ_1 and μ_2 is a joint probability measure μ whose marginals are μ_1 and μ_2.

Definition 2.2 Let μ_1 and μ_2 be probability measures on a metric space $(\mathcal{X}, d)$ and let $\Gamma(\mu_1, \mu_2)$ denote the set of all couplings μ between μ_1 and μ_2. Then the Wasserstein metric (Kantorovich–Rubinstein distance of order one) is given by

$$W_1(\mu_1, \mu_2) = \inf_{\mu \in \Gamma(\mu_1, \mu_2)} \int_{\mathcal{X} \times \mathcal{X}} d(x, y)\,d\mu(x, y) \tag{1}$$

Remark 2 The following property of the Wasserstein distance will prove useful for us. Suppose two probability measures have support on $U \subset \mathcal{X}$ and there exists a metric $\tilde{d}$ on $\mathcal{X}$ that coincides with d on U. Then the Wasserstein metric $\tilde{W}_1(\mu_1, \mu_2)$ associated with metric $\tilde{d}$ equals the original Wasserstein metric associated with d.

Let G be a graph. The definition of Ollivier–Ricci curvature on graphs relies on two ingredients, a metric on G and a family of probability measures, indexed by the vertices.

Definition 2.3 An Ollivier-triple $\mathcal{G}$ is a triple $(G, d_G, \boldsymbol{m})$, where G is a graph, d_G a metric on G and $\boldsymbol{m} = \{m_x\}_{x \in G}$ a family of probability measures on G for each node $x \in G$.

Given an Ollivier-triple $\mathcal{G} = (G, d_G, \boldsymbol{m})$, we write $W_1^{\mathcal{G}}$ for the Wasserstein metric with respect to the metric space (G, d_G). We then define for any pair of nodes $x, y \in G$ the associated Ollivier curvature as

$$\kappa(x, y; \mathcal{G}) = \begin{cases} 1 - \dfrac{W_1^{\mathcal{G}}(m_x, m_y)}{d_G(x, y)} & \text{if } d_G(x, y) < \infty, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Remark 3 1. The concept of Ollivier–Ricci curvature is not restricted to graphs and can be defined on any metric space where we have a sequence of probability measures. A specific example of these are Riemannian manifolds $(\mathcal{M}, d_{\mathcal{M}})$.

2. Note that a sequence $\{m_x\}_{x \in G}$ of probability measures on G gives rise to a random walk on the graph. The transition probabilities are given by $\mathbb{P}(x_{t+1} \in A \mid x_t = x) = m_x(A)$. So an Ollivier-triple consists of a graph, a metric and a random walk on the graph. However, since we will only use concepts related to the measures m_x we refrain from using any random walk terminology.

3. When d_G is the shortest path metric on G and $\boldsymbol{m}$ corresponds to the uniform probability measures on the neighborhoods $\mathcal{N}_x$, i.e., $m_x(y) = G_{xy}/|\mathcal{N}_x|$, we are in the classic setting for Ollivier–Ricci curvature on graphs [16, 26, 31]. In this work, however, we shall use different combinations of metrics on graphs and probability measures to obtain our results. This is why we define Ollivier–Ricci curvature on graphs in a more general way.

4. The reason why we set $\kappa(x, y; \mathcal{G}) = 0$ if the nodes are not in the same connected component is because we work with random graphs and this way we ensure that $\kappa(x, y; \mathcal{G})$ is a real-valued random variable.

2.4 Curvature in Riemannian Manifolds

Our main results relate the standard Ricci curvature of a manifold to the Ollivier–Ricci curvature of the random geometric graph constructed on this manifold. For this, we briefly recall the definition of the Ricci curvature, see [15, 30].

In general, the curvature of a geometric space is intended as a local measure for how "different" a region of the space is from that of the flat Euclidean space. Notions of curvature in Riemannian geometry are governed by the *Riemannian curvature tensor R*. Given an N-dimensional Riemannian Manifold $(\mathcal{M}, d_{\mathcal{M}})$, a point $x \in \mathcal{M}$ and two vectors $\mathbf{v}, \mathbf{w} \in T_x\mathcal{M}$ (the tangent space of x), the Riemannian curvature tensor with respect to $\mathbf{v}, \mathbf{w}$ is a linear map $R(\mathbf{v}, \mathbf{w}) \colon T_x\mathcal{M} \to T_x\mathcal{M}$, written as $\mathbf{u} \mapsto R(\mathbf{v}, \mathbf{w})\mathbf{u}$ and defined in terms of the Levi-Civita connection on the tangent bundle. It quantifies to what extent the manifold $\mathcal{M}$ is not isometric to flat Euclidean space.

In this paper we will use the notion of curvature called *Ricci curvature*. For two vectors $\mathbf{v}$ and $\mathbf{w}$, the Ricci curvature $\mathrm{Ric}(\mathbf{v}, \mathbf{w})$ is defined, in terms of the Riemannian tensor, as the trace of the linear map

$$\mathbf{u} \mapsto R(\mathbf{u}, \mathbf{v})\mathbf{w}, \quad \mathbf{u} \in T_x\mathcal{M}.$$

Given a point $x \in \mathcal{M}$ and a unit vector $\mathbf{v} \in T_x \mathcal{M}$, we often refer to $\mathrm{Ric}(\mathbf{v}, \mathbf{v})$ as the *Ricci curvature* of x with respect to v.

This Ricci curvature is related to another notion of curvature, called *sectional curvature*, which is defined as

$$K(\mathbf{v}, \mathbf{w}) = \frac{\langle R(\mathbf{v}, \mathbf{w})\mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle^2},$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on the tangent space. One can show that $\mathrm{Ric}(\mathbf{v}, \mathbf{v})$ is obtained by averaging the sectional curvature $K(\mathbf{v}, \mathbf{w})$ over all unit vectors $\mathbf{w} \in T_x \mathcal{M}$.

In the remainder of this paper we will work with the Ricci curvature of a point x, with respect to some tangent vector $\mathbf{v}$. We note that it is not needed to understand the fine details behind curvature of Riemannian manifolds to understand all the details of the results or proofs.

3 Main Results

Here we state our results regarding the convergence of Ollivier–Ricci curvature of random geometric graphs on Riemannian manifolds. We note that if the manifold dimension is $N = 1$, then there is nothing to prove, so that we always assume that $N \geq 2$.

We mainly consider two different distances on the graphs, leading to two different but related results. Although we consider different distances on graphs, we shall always consider uniform measures on balls of a certain radius. We shall clearly distinguish between the connection radius of the graph and the radius used for the uniform measures: The former is the connectivity distance threshold: if the distance between

connection radius:	ε_n
measure radius:	δ_n

a pair of nodes in the manifold is below this threshold, then these nodes are connected by an edge in the graph. The latter radius is the radius of the ball (either in the graph or in the manifold) over which the uniform probability measure is distributed.

Let $G_n = \mathbb{G}_n(\varepsilon_n)$ be a random geometric graph on $\mathcal{M}$ and d_G a distance on G_n. Then, for a node $x \in G_n$, we define the graph ball of radius λ around x as

$$\mathcal{B}_G(x; \lambda) := \{y \in G_n \setminus \{x\} : d_G(x, y) \leq \lambda\}.$$

Note that $\mathcal{B}_G(x; \lambda)$ depends on the definition of the graph distance d_G. For our results we consider Ollivier-triples $\mathcal{G}_n = (G_n, d_G, \boldsymbol{m}^G)$, where $\boldsymbol{m}^G$ are the uniform measures

on $\mathcal{B}_G(x; \delta_n)$, i.e.,

$$
m_x^G(y) = \begin{cases} \dfrac{1}{|\mathcal{B}_G(x; \delta_n)|} & \text{if } y \in \mathcal{B}_G(x; \delta_n), \\ 0 & \text{else.} \end{cases} \tag{3}
$$

We reiterate that if $\varepsilon_n = \delta_n$ and the graph metric d_G is the shortest path distance, then we are in the classical setting of Ollivier–Ricci curvature on graphs as considered in the past literature [16, 26, 31].

3.1 Graphs with Manifold Weighted Distance

Let $G_n = \mathbb{G}_n(x^*, \varepsilon_n)$ be a random rooted graph on $\mathcal{M}$. Then we define the manifold weighted graph distance d_G^w as the weighted shortest-path distance on G_n where each edge (u, v) is assigned weight $d_{\mathcal{M}}(u, v)$, corresponding to the distance between the nodes on the manifold. Similarly to $\mathcal{B}_G(x; \lambda)$, we denote by $\mathcal{B}_G^w(x; \lambda)$ the graph ball of radius λ with respect to d_G^w and let $\boldsymbol{m}^{G,w} = (m_x^{G,w})_{x \in G}$ denote the uniformly measures on the balls $\mathcal{B}_G^w(x; \delta_n)$. Finally, given a point $x \in \mathcal{M}$ and a vector $\mathbf{v} \in T_x \mathcal{M}$, we say that another point $y \in \mathcal{M}$ is *at distance δ in the direction of* $\mathbf{v}$, if $d_{\mathcal{M}}(x, y) = \delta$ and y lies on the geodesic starting at x in the direction of $\mathbf{v}$.

Our first result shows that for certain combinations of connection radius ε_n and measure radius δ_n, the Ollivier–Ricci curvature on G_n converges to the Ricci curvature.

Theorem 1 *Let $N \geq 2$, $(\mathcal{M}, d_{\mathcal{M}})$ be a smooth, connected, and compact N-dimensional Riemannian manifold, $x^* \in \mathcal{M}$, and $\mathbf{v}$ a unit tangent vector at x^*. Furthermore, let $\varepsilon_n = \Theta((\log n)^a n^{-\alpha})$, $\delta_n = \Theta((\log n)^b n^{-\beta})$ (as $n \to \infty$) where the constants satisfy*

$$
0 < \beta \leq \alpha, \quad \alpha + 2\beta \leq \frac{1}{N},
$$

and $a \leq b$ if $\alpha = \beta$ and $\min\{a, a + 2b\} > 2/N$ if $\alpha + 2\beta = 1/N$. Let $y_n^* \in \mathcal{M}$ be at distance δ_n in the direction of $\mathbf{v}$ and $G_n = \mathbb{G}_n(x^*, y_n^*, \varepsilon_n)$ be rooted random graphs on $\mathcal{M}$. Then for the Ollivier-triple $\mathcal{G}_n^w = (G_n, d_{G_n}^w, \boldsymbol{m}^{G,w})$, it holds

$$
\lim_{n \to \infty} \mathbb{E}\left[\left| \frac{2(N + 2)\kappa(x^*, y_n^*; \mathcal{G}_n^w)}{\delta_n^2} - \mathrm{Ric}(\mathbf{v}, \mathbf{v}) \right| \right] = 0.
$$

Theorem 1 relates two different quantities. The first is the Ollivier–Ricci curvature in the graph between the node x^* and another node y_n^* that is at distance δ_n from x^* in the direction of vector $\mathbf{v}$. The second is the Ricci curvature of the manifold at x^* in the $\mathbf{v}$-direction. The theorem says that if we properly rescale the former, it converges in expectation to the latter.

Remark 4 1. Note that Theorem 1 states that $\delta_n^{-2} 2(N + 2)\kappa(x^*, y_n^*; \mathcal{G}_n^w)$ converges in the L^1 sense to $\text{Ric}(\mathbf{v}, \mathbf{v})$. In particular, this implies the concentration result

$$\lim_{n \to \infty} \mathbb{P}\left(\left|\frac{2(N + 2)\kappa(x^*, y_n^*; \mathcal{G}_n^w)}{\delta_n^2} - \text{Ric}(\mathbf{v}, \mathbf{v})\right| \geq \eta\right) = 0, \quad \text{for all } \eta > 0.$$

2. Since $\varepsilon_n, \delta_n \to 0$, both the connectivity and measure neighborhoods of x^* become smaller as n grows. Indeed, curvature is a local property, so that measuring it more accurately requires smaller regions.

3. While the connectivity neighborhood of x^* is shrinking, the expected number of x^*'s neighbors lying in it is growing with n. To see this, note that for large enough n the volume of the ball $\mathcal{B}_{\mathcal{M}}(x; \varepsilon_n)$ around $x \in \mathcal{M}$ can be approximated by that of the N-dimensional Euclidean ball. Hence, for any $x \in \mathbb{G}_n(x^*, y_n^*, \varepsilon_n)$, as $n \to \infty$,

$$\mathbb{E}[|\mathcal{N}_x|] = n \, \text{vol}_{\mathcal{M}}(\mathcal{B}_{\mathcal{M}}(x; \varepsilon_n)) = \Theta(n\varepsilon_n^N) = \Theta\big((\log n)^{aN} n^{1-\alpha N}\big).$$

The conditions of the theorem imply that $\alpha \leq \alpha + 2\beta \leq 1/N$, so that $1 - \alpha N \geq 0$. This means that the average degree diverges faster than logarithmically if $\alpha N < 1$. More generally, the conditions of Theorem 1 imply that the average degree always diverges faster than $(\log n)^2$.

If we consider the classic setting where the connection and measure radii are the same, $\varepsilon_n = \delta_n$, then the following result is a direct consequence of Theorem 1.

Corollary 1 *Let $N \geq 2$, $(\mathcal{M}, d_{\mathcal{M}})$ be a smooth, connected, and compact N-dimensional Riemannian manifold, $x^* \in \mathcal{M}$, and $\mathbf{v}$ a unit tangent vector at x^*. Furthermore, let $\delta_n = \Theta((\log n)^b n^{-\beta})$, with $\beta \leq 1/(3N)$ and $b > 2/N$ whenever $\beta = 1/(3N)$. Let $y_n^* \in \mathcal{M}$ be at distance δ_n in the direction of $\mathbf{v}$ and $G_n = \mathbb{G}_n(x^*, y_n^*, \delta_n)$ be rooted random graphs on $\mathcal{M}$. Then for the Ollivier-triple $\mathcal{G}_n^w = (G_n, d_{G_n}^w, \mathbf{m}^{G,w})$, it holds*

$$\lim_{n \to \infty} \mathbb{E}\left[\left|\frac{2(N + 2)\kappa(x^*, y_n^*; \mathcal{G}_n^w)}{\delta_n^2} - \text{Ric}(\mathbf{v}, \mathbf{v})\right|\right] = 0.$$

While the conditions in this corollary imply that the average degree in $\mathbb{G}_n(x^*, y_n^*, \delta_n)$ diverges faster than $n^{2/3}$, Theorem 1 works for graphs where the average degree can be almost as small as $(\log n)^2$. The crucial component for establishing the curvature convergence in graphs with so much smaller average degree is to consider different connection and measure radii and let the connection radius decrease at a faster rate than the measure radius, i.e., $\varepsilon_n \ll \delta_n$.

Remark 5 *(extreme cases for convergence of curvature)* Corollary 1 covers one set of extreme cases for the combination a, b, α and β from Theorem 1, were we take β to be as big as possible. This means that we compute the curvature using uniform probability measures on a set of nodes that is *as small as possible*. For the true extreme case, let $\eta > 0$ be arbitrarily small and define $\beta = (1 - \eta)/(3N)$ and $b = (2 + \eta)/N$.

Then, to calculate the curvature, we need to compute the Wasserstein metric between uniform probability measures on neighborhoods that contain

$$\Theta(n\varepsilon_n^N) = \Theta(n\delta_n^N) = \Theta\big((\log n)^{2+\eta} n^{(2+\eta)/3}\big),$$

number of nodes. The consequence, however, is that our graphs have average degree diverging at the same rate: $(\log n)^{2+\eta} n^{(2+\eta)/3}$.

In order to get graphs whose average degree diverges as slow as possible, we need to consider an other extreme case. Again let $\eta > 0$ be arbitrary small. Now we define

$$a = \frac{2+\eta}{N}, \quad \alpha = \frac{1-\eta}{N}, \quad b = a, \quad \beta = \frac{\eta}{2N}.$$

For these choices we have that $\alpha + 2\beta = 1/N$ and $\min\{a, a+2b\} = a > 2/N$ so that the result from Theorem 1 holds. In this case, the average degree scales as

$$\Theta(n\varepsilon_n^N) = \Theta((\log n)^{Na} n^{1-N\alpha}) = \Theta((\log n)^{2+\eta} n^{\eta}),$$

which is almost logarithmic. However, we now need to compute the Wasserstein metric with respect to the uniform measure on a number of nodes that scales as

$$\Theta(n\delta_n^N) = \Theta((\log n)^{2+\eta} n^{1-\eta/2}).$$

That is, in order to compute curvature on graphs with almost logarithmic average degree, we need to consider the uniform probability measure on almost the entire graph.

3.2 Graphs with Hop Count Distance

In the previous section we considered Ollivier–Ricci curvature of graphs on Riemannian manifolds, with graph edges weighted by manifold distances. These weights encode a lot of information about the manifold metric structure, so that one may feel not terribly surprised that we can recover manifold curvature from graph curvature using this information. The natural question is then if it is possible to prove convergence of Ollivier–Ricci curvature based on shortest path distances d_G^s in *unweighted* graphs. It turns out that this can be done under some slightly more restrictive conditions on the connection and measure radii.

For this we define, for any random geometric graph $G_n = \mathbb{G}(\varepsilon_n)$, the rescaled shortest path distance $d_G^*(x, y) = \varepsilon_n d_G^s(x, y)$. Similar to the previous setting we let $\mathcal{B}_G^*(x; \delta_n)$ denote the balls of radius δ_n around in $x \in G_n$ with respect to the metric d_G^* and define the random walk measures

$$m_x^{G,*}(y) = \begin{cases} \dfrac{1}{|\mathcal{B}_G^*(x; \delta_n)|} & \text{if } y \in \mathcal{B}_G^*(x; \delta_n), \\ 0 & \text{else.} \end{cases}$$

Theorem 2 *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a* ***smooth, connected, and compact*** *2-dimensional Riemannian manifold, $x^* \in \mathcal{M}$, and* **v** *a unit tangent vector at x^*. Furthermore, let $\varepsilon_n = \Theta((\log n)^a n^{-\alpha})$, $\delta_n = \Theta((\log n)^b n^{-\beta})$ where the constants satisfy*

$$0 < \beta \leq \frac{1}{9} \quad and \quad 3\beta \leq \alpha \leq \frac{1 - 3\beta}{2},$$

and $a < 3b$ if $\alpha = 3\beta$ and $2a + 3b > 1$ if $\alpha = (1 - 3\beta)/2$. Let $y_n^* \in \mathcal{M}$ be at distance δ_n in the direction of **v** *and $G_n = \mathbb{G}_n(x^*, y_n^*, \varepsilon_n)$ be rooted random graphs on $\mathcal{M}$. Then for the Ollivier-triple $\mathcal{G}_n^* = (G_n, d_{G_n}^*, \mathbf{m}^{G,*})$, it holds*

$$\lim_{n \to \infty} \mathbb{E}\left[\left| \frac{2(N + 2)\kappa(x^*, y_n^*; \mathcal{G}_n^*)}{\delta_n^2} - \mathrm{Ric}(\mathbf{v}, \mathbf{v}) \right| \right] = 0.$$

Remark 6 1. Note that unlike Theorem 1, here we do not include any information on the distances between nodes on the manifold. This is because the distance $\delta_{G_n}^*$ is simply the shortest path distance on the graph G_n rescaled by the connection radius ε_n.

2. Observe that the theorem allows to select an α that is arbitrary close to $1/2$. In particular,

$$\mathbb{E}[|\mathcal{N}_x|] = \Theta(n\varepsilon_n^2) = \Theta((\log n)^{2a} n^{1-2\alpha}) \leq \Theta((\log n)^{2a} n^{6\beta}).$$

Hence by selecting a small β we have a discrete notion of curvature that converges on graphs with almost logarithmic average degree, without using any information on the manifold.

3. Theorem 2 currently only works in 2-dimensional manifolds. This is because the proof relies on results for the stretch (the fraction $d_G/d_{\mathcal{M}}$) for random geometric graphs in 2-dimensional Euclidean space [9]. Our proof techniques, however, immediately allow the results to be extended to higher dimensions, once similar types of stretch results for these spaces are obtained.

3.3 Summary, Comments, Caveats, and Outlook

In summary, we have proven that upon proper rescaling, the Ollivier–Ricci curvature of random geometric graphs on a Riemannian manifold converges to the Ricci curvature of the underlying manifold.

Our first result, Theorem 1, establishes convergence of Ollivier–Ricci curvature for a wide range of connectivity and measure radii. In particular, it contains as a corollary the classical setting where both radii are the same, Corollary 1. The theorem does, however, require knowledge of pairwise distances between connected nodes in the manifold.

Our second result, Theorem 2, relaxes this requirement and establishes the same convergence without any knowledge of distances in the manifold. This does come at the price of slightly more restrictive conditions on the possible connection and measure

radii. Still, as for the first result, the convergence holds all the way up to graphs whose average degree grows very slowly (almost logarithmically).

To the best of our knowledge, these are the first rigorous results on the convergence of a discrete notion of curvature of random combinatorial objects to a traditional continuum notion of curvature of smooth space.

While the classical setting for Ollivier–Ricci graph curvature uses probability measures (random walks) on balls of the same radius as the graph connection radius, in this paper we allow the radii to be different. This is an important generalization. In particular, we find that in order for the curvature to converge on graphs with almost logarithmic average degree, we need the probability measure radius to be much larger than the connection radius. This is intuitively expected because in order to "feel" any curvature in graphs with such a low density, we really need to consider large "mesoscopic" neighborhoods in them since otherwise all we could see is local "microscopic" Euclidean flatness. It would be interesting to see how this more general approach would generalize known results for the classical setting of Ollivier–Ricci curvature of graph families that have been investigated in the past, such as trees or Erdős–Rényi random graphs [4, 16].

In our recent numeric experiments [13], we have seen that in manifold-distance-weighted random geometric graphs, the Ollivier–Ricci curvature convergence holds even for graphs with constant average degree. Unfortunately, the proof techniques presented in this paper do not allow for a direct generalization to this setting. Therefore, other techniques are needed to (dis)confirm the convergence of Ollivier–Ricci curvature of graphs with constant average degree. We note that one definitely cannot expect Ollivier–Ricci curvature to converge in all possible graph sparsity settings. For example, we definitely need the giant component to exist to talk about any curvature convergence.

For the task of learning latent geometry in networks, our results can still be improved, particularly by removing the requirement to know the connection radius. When presented just with a truly unweighted realization of a random geometric graph, this radius needs first to be learnt, estimated. It would thus be interesting to see if convergence would still hold if we replace the true value of the connection radius with its consistent estimation, e.g. based on the average degree. Here we expect the speed of curvature convergence (if any) to depend on the speed of estimator convergence in a possibly nontrivial way.

Finally, now that we have seen that Ollivier–Ricci curvature of random combinatorial discretizations of smooth spaces converges to their Ricci curvature, it would be interesting to investigate whether such convergence also holds for other popular notions of discrete curvature. Forman–Ricci curvature [37] appears to be a good next candidate for such investigation.

4 Proof Overview

Our main results in Theorems 1 and 2 follow from our more general result on the Ollivier–Ricci curvature convergence in graphs whose edges are always weighted by some weights. That is, we assume that all edges in our graphs always have some

weights, assigned according to some scheme. For our general result it is not important what these weights or their assignment scheme are. What is important is that the graph distance d_G between node pairs is a good approximation of the manifold distance $d_{\mathcal{M}}$ between the corresponding pair of points. Here by graph distance we mean any metric on the vertex set of the graph. To quantify how good this approximation is, we introduce the following definition.

Definition 4.1 Let $(\mathcal{M}, d_{\mathcal{M}})$ be an N-dimensional Riemannian manifold and $G_n = \mathbb{G}_n(x^*, \varepsilon_n)$ a rooted random graph on $\mathcal{M}$. A graph distance d_G on G_n is said to be a δ_n-*good approximation of* $d_{\mathcal{M}}$ if $d_{\mathcal{M}} \leq d_G$ and the following holds (as $n \to \infty$): there exists a $Q > 3$ and $\xi_n = o(\delta_n)$ such that with probability $1 - o(\delta_n^3)$,

$$|d_{\mathcal{M}}(u, v) - d_G(u, v)| \leq d_{\mathcal{M}}(u, v)\xi_n^2 + \xi_n^3, \tag{4}$$

holds for all $u, v \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n) \cap G_n$.

Remark 7 *(asymptotic expressions)*   Most of our results will deal with asymptotic relations, e.g. $\xi_n = o(\delta_n)$. Unless stated otherwise, these asymptotic relations will always be understood as $n \to \infty$.

There are several examples of shortest weighted path distance that are δ_n-good approximation of the manifold distance. In this paper we consider two cases. In one, each edge (u, v) has weight equal to $d_{\mathcal{M}}(u, v)$, while in the other case the weight is simply the connection radius ε_n. An explicit example of the latter case is when the manifold is 2-dimensional and the connection and measure radii are given by $\varepsilon_n = n^{-1/3}$ and $\delta_n = n^{-1/9} \log n$, respectively. See Propositions 4 and 5 for more details.

Recall that $\mathcal{B}_G(x; \delta)$ denotes the set of nodes in the graph that are at graph distance at most δ from x,

$$m_x^G(y) = \begin{cases} \dfrac{1}{|\mathcal{B}_G(x; \delta_n)|} & \text{if } y \in \mathcal{B}_G(x; \delta_n), \\ 0 & \text{else,} \end{cases}$$

and define

$$\lambda_n = (\log n)^{2/N} n^{-1/N}. \tag{5}$$

This λ_n will play the role of an additional radius, for extending the graph distance d_G to the manifold. In short, to define a distance between $u, v \in \mathcal{M}$, we will connect u and v to all points of the graph within radius λ_n and then use the graph distance. The radius λ_n has been selected such that the expected number of nodes inside any ball $\mathcal{B}_{\mathcal{M}}(x; \lambda_n)$ is of the order $\Theta((\log n)^2)$. Hence, the probability of observing no node of the graph inside any such ball is $O(e^{-(\log n)^2}) = o(n^{-1})$, which is sufficiently small. More details on the use of λ_n can be found in Sect. 5.1. Our general result is then as follows.

Theorem 3 *Let $N \geq 2$, $(\mathcal{M}, d_{\mathcal{M}})$ be a smooth, connected, and compact N-dimensional Riemannian manifold, $x^* \in \mathcal{M}$, and $\mathbf{v}$ a unit tangent vector at x^*. Furthermore, let $\varepsilon_n \leq \delta_n = o(1)$ be such that $\lambda_n = o(\varepsilon_n)$ and $\lambda_n = o(\delta_n^3)$. Let $y_n^* \in \mathcal{M}$ be at distance δ_n in the direction of $\mathbf{v}$, $G_n = \mathbb{G}_n(x^*, y_n^*, \varepsilon_n)$ be rooted random graphs on $\mathcal{M}$, and d_G a δ_n-good approximation of $d_{\mathcal{M}}$. Then, if we consider the Ollivier-triple $\mathcal{G}_n = (G_n, d_G, \boldsymbol{m}^G)$,*

$$\lim_{n \to \infty} \mathbb{E}\left[\left|\frac{2(N+2)\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \mathrm{Ric}(\mathbf{v}, \mathbf{v})\right|\right] = 0.$$

Once we have established this general result, our main results in Theorems 1 and 2 follow if we can show that the considered graph distances are δ_n-good approximations.

A key ingredient in the proof of Theorem 3 is the convergence result for Ollivier–Ricci curvature for uniform measures on Riemannian manifolds, proved in the seminal paper on the topic [28]. In a high-level overview, our proof approximates Ollivier–Ricci curvature of probability measures on the graph with those on the manifold. Having obtained such an approximation with a required accuracy, we then apply the convergence result from [28].

Since Ollivier–Ricci curvature is defined by the Wasserstein metric on probability measures, our analysis focuses on approximating the Wasserstein metric of discrete probability measures on the graph by the Wasserstein metric of uniform probability measures on the manifold. This is done in three steps: 1) extend the graph distance d_G to a distance $\widetilde{d}_{\mathcal{M}}$ on the manifold such that the Wasserstein metric $\widetilde{W}_1$ with respect to this new distance is a good approximation of the Wasserstein metric W_1 on the manifold, 2) show that the Wasserstein metric between the probability measure m_x^G on the graph and the discrete probability measure $m_x^{\mathcal{M}}$ on the nodes within the ball $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$ is sufficiently small, and 3) show that the Wasserstein metric between the uniform measure on $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$ and the discrete probability measure $m_x^{\mathcal{M}}$ is sufficiently small.

Remark 8 In all cases, sufficiently small means that the error terms are of smaller order than δ_n^3. This is because the Wasserstein metric is first divided by δ_n to obtain the curvature, which is then divided by δ_n^2 to make it converge to the Ricci curvature.

We proceed with explaining all ingredients and the three steps in more detail. We reiterate that unless stated otherwise, we will assume that $\varepsilon_n \leq \delta_n$ are two sequences converging to zero such that $\lambda_n = o(\varepsilon_n)$ and $\lambda_n = o(\delta_n^3)$.

4.1 Ollivier Curvature on Riemannian Manifolds

Let $(\mathcal{M}, d_{\mathcal{M}})$ be a smooth, orientable, connected and compact N-dimensional Riemannian manifold. For $x \in \mathcal{M}$ and $\delta > 0$, we write $\mathcal{B}_{\mathcal{M}}(x; \delta) \subseteq \mathcal{M}$ to denote the closed ball of radius δ around x, i.e., $\mathcal{B}_{\mathcal{M}}(x; \delta) = \{y \in \mathcal{M} : d_{\mathcal{M}}(x, y) \leq \delta\}$. Recall that

$$\mathrm{vol}_{\mathcal{M}}(\mathcal{B}_{\mathcal{M}}(x; \delta)) := \int_{\mathcal{B}_{\mathcal{M}}(x;\delta)} \mathrm{dvol}_{\mathcal{M}}(y),$$

denotes the volume of the ball $\mathcal{B}_{\mathcal{M}}(x; \delta)$. Now fix $\delta > 0$ and consider the uniform measure on balls of radius δ. That is, for $x \in \mathcal{M}$ we take the probability measure μ_x^δ given by

$$d\mu_x^\delta(y) = \begin{cases} \dfrac{1}{\mathrm{vol}_{\mathcal{M}}(\mathcal{B}_{\mathcal{M}}(x; \delta))} \, d\mathrm{vol}_{\mathcal{M}}(y) & \text{if } y \in \mathcal{B}_{\mathcal{M}}(x; \delta), \\ 0 & \text{else.} \end{cases} \tag{6}$$

We will refer to μ_x^δ as the *uniform δ-measure*. The following result from [28] shows that for a uniform δ-measure on a Riemannian manifold, the Ollivier curvature (properly rescaled) converges to the Ricci curvature as $\delta \to 0$.

Theorem 4 [28, Exam. 7] *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a smooth complete N-dimensional Riemannian manifold $x \in \mathcal{M}$ and $\mathbf{v}$ a unit tangent vector at x. Let $\delta > 0$ and y_δ be the point at distance δ in the direction of $\mathbf{v}$. Then if we consider the Ollivier–Ricci curvature κ for the uniform δ-measures given by* (6),

$$\lim_{\delta \to 0} \frac{2(N+2)}{\delta^2} \kappa(x, y_\delta) = \mathrm{Ric}(\mathbf{v}, \mathbf{v}).$$

Remark 9 The result in Theorem 4 clearly exhibits the local nature of curvature as it holds in the limit where the distance $d_{\mathcal{M}}(x, y) = \delta$ between the two points goes to zero.

Taking $\delta = \delta_n$, $x = x^*$, and $y = y_n^*$ in the above theorem, we have that the rescaled Ollivier–Ricci curvature associated to the uniform δ_n-measures converges to the Ricci curvature as $n \to \infty$. The main strategy for proving Theorem 3 is to compare this "uniform" version of the curvature κ on the manifold to the discrete version on the graph. More precisely, we need to prove that

$$\mathbb{E}\big[\big|W_1^G\big(m_{x^*}^G, m_{y_n^*}^G\big) - W_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big)\big|\big] = o(\delta_n^3). \tag{7}$$

There are two complicating factors here. First, we have to deal with two Wasserstein metrics defined on two different spaces. Second, we have to compare discrete probability measures with continuous ones. We deal with the different Wasserstein metrics in the next section and with comparing the different measures in Sects. 4.3 and 4.4.

4.2 Extending the Graph Distance to the Manifold

In order to compare the two different Wasserstein metrics in (7) we extend the graph distance d_G to a distance $\widetilde{d}_{\mathcal{M}}$ defined on a sufficiently large part of $\mathcal{M}$. In particular, we will consider the ball $\mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$, with $Q > 3$ from Definition 4.1. The extension is such that for any two nodes $x, y \in G_n$, $d_G(x, y) = \widetilde{d}_{\mathcal{M}}(x, y)$, so that $W_1^G(m_{x^*}^G, m_{y_n^*}^G)$ can be replaced by the Wasserstein metric associated with $\widetilde{d}_{\mathcal{M}}$.

Recall the definition of λ_n from (5), $\lambda_n = (\log n)^{2/N} n^{-1/N}$. Denote $G_n = \mathbb{G}_n(x^*, y_n^*, \delta_n)$ and let $U \subset \mathcal{M}$ be a countable set of points. Then we define the
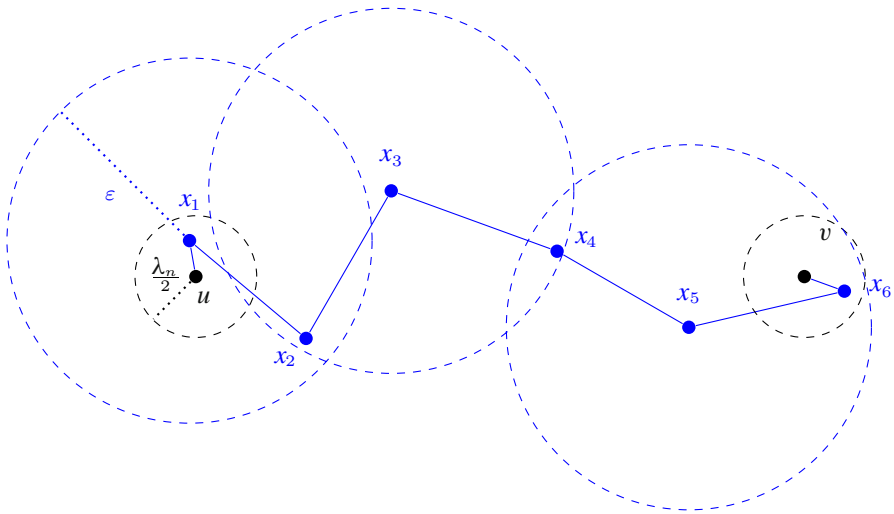
Fig. 1 Illustration of the extended graph distance $\widetilde{d}_{\mathcal{M}}$. Here u is connected to node x_1 and v to x_6 and the shortest geodesic-weighted path between x_1 and x_6 goes over five edges

graph $G_n(U)$ obtained from G_n by adding the points of U to the vertex set and connecting each $u \in U$ to any other node $x \in G_n \setminus U$ for which $d_{\mathcal{M}}(x, u) \leq \lambda_n/2$. After this, we assign to each new edge (u, x) the weight $d_{\mathcal{M}}(x, u)(1 + \xi_n^2) + \xi_n^3$, with ξ_n from Definition 4.1. We can then extend the graph distance to the manifold by defining $\widetilde{d}_{\mathcal{M}}(u, v)$ to be the graph distance $d_G(u, v)$ computed in the extended graph $G_n(\{u, v\})$ with the added weights. That is, $\widetilde{d}_{\mathcal{M}}(u, v)$ is the shortest weighted path distance in the extended graph $G_n(\{u, v\})$, where the weights follow the same scheme as for the original graph.

Observe that if $x, y \in G_n$ then $\widetilde{d}_{\mathcal{M}}(x, y) = d_G(x, y)$ so that the distance on nodes of G_n does not change and hence $\widetilde{d}_{\mathcal{M}}$ is a true extension of d_G. In addition, by definition of the graph distance it immediately follows that $\widetilde{d}_{\mathcal{M}}(u, v) = 0$ if and only if $u = v$. Figure 1 shows an illustration of the extended distance.

It is important to note that this extended distance depends on the random graph G_n. Therefore, it could happen that two added points $u, v \in U$ are not connected in $G_n(U)$, i.e., there does not exist a path from u to v in the extended graph. This happens if there are no nodes in $\mathcal{B}_{\mathcal{M}}(u; \lambda_n/2)$ or in $\mathcal{B}_{\mathcal{M}}(v; \lambda_n/2)$ or if none of the node pairs $(x, y) \in \mathcal{B}_{\mathcal{M}}(u; \lambda_n/2) \times \mathcal{B}_{\mathcal{M}}(v; \lambda_n/2)$ are connected by a path in G_n. Therefore, to justify the definition of the extended manifold distance we need to make sure that, with sufficiently high probability, theses situations do not occur.

Lemma 1 *Let $G_n = \mathbb{G}_n(x^*, y_n^*, \delta_n)$ and $Q > 3$ be the constant from Definition 4.1. Then, there exists an event Ω_n satisfying $\mathbb{P}(\Omega_n) \geq 1 - o(\delta_n^3)$ such that on this event the following holds:*

($\Omega 1$) *$(\mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n), \widetilde{d}_{\mathcal{M}})$ is a metric space and*
($\Omega 2$) *$\widetilde{d}_{\mathcal{M}}(u, v) = d_{\mathcal{M}}(u, v) + o(\delta_n^3)$.*

The first property ensures that our extended distance is an actual distance. Moreover, by the second property, this extended distance is a good approximation of the true distance on the manifold. Finally, we also note that the first property makes sure that $d_G(x^*, y_n^*) = \widetilde{d}_{\mathcal{M}}(x^*, y_n^*) < \infty$, so that the curvature κ between x^* and y_n^* is well defined and not forced to be zero. The precise definition of Ω_n is not needed to understand the high level arguments as well as the proof of the main results. For now, let us refer to Ω_n as the *good event*. Details on this event can be found in Sect. 5.1.

Let $\widetilde{W}_1$ denote the Wasserstein metric with respect to $\widetilde{d}_{\mathcal{M}}$, which is only well defined on the *good event* Ω_n. Since the distance is determined by the graph $G_n = \mathbb{G}_n(x^*, y_n^*, \delta_n)$, the Wasserstein metric is also a random object. The following proposition states that, on the event Ω_n, the difference between the Wasserstein metrics $\widetilde{W}_1$ and W_1 is small. The proof is given in Sect. 5.1.

Proposition 1 *Let $G_n = \mathbb{G}_n(x^*, \varepsilon_n)$ and μ_1, μ_2 be two probability measures on $\mathcal{M}$ with support contained in $\mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$. Then*

$$\mathbb{E}\big[|\widetilde{W}_1(\mu_1, \mu_2) - W_1(\mu_1, \mu_2)| \,|\, \Omega_n\big] = o(\delta_n^3).$$

Recall that $\widetilde{d}_{\mathcal{M}}(x, y) = d_G(x, y)$ if $x, y \in G_n$, and therefore $W_1^G(m_{x^*}^G, m_{y_n^*}^G) = \widetilde{W}_1(m_{x^*}^G, m_{y_n^*}^G)$. Hence, since the uniform δ_n-measures $\mu_{x^*}^{\delta_n}$ and $\mu_{y_n^*}^{\delta_n}$ are probability measures on $\mathcal{M}$ with support contained in $\mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$, Proposition 1 implies that on the good event,

$$\big|W_1^G(m_{x^*}^G, m_{y_n^*}^G) - W_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})\big| = \big|\widetilde{W}_1(m_{x^*}^G, m_{y_n^*}^G) - \widetilde{W}_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})\big| + o(\delta_n^3)$$

holds in expectation. This is helpful because both Wasserstein metrics in the expression on the right hand side are now defined on the same space. Therefore, since $\widetilde{W}_1$ is a distance, the reverse triangle inequality implies

$$\big|\widetilde{W}_1(m_{x^*}^G, m_{y_n^*}^G) - \widetilde{W}_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})\big| \leq \widetilde{W}_1(m_{x^*}^G, \mu_{x^*}^{\delta_n}) + \widetilde{W}_1(m_{y_n^*}^G, \mu_{y_n^*}^{\delta_n}).$$

Applying Proposition 1 again we get that

$$\big|\widetilde{W}_1(m_{x^*}^G, m_{y_n^*}^G) - \widetilde{W}_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})\big| \leq W_1(m_{x^*}^G, \mu_{x^*}^{\delta_n}) + W_1(m_{y_n^*}^G, \mu_{y_n^*}^{\delta_n}) + o(\delta_n^3)$$

holds in expectation, conditioned on the good event. However, the right hand side no longer involves the extended distance. Hence, it now suffices to show that for any $x \in \mathcal{B}_{\mathcal{M}}(x^*; \delta_n)$,

$$\mathbb{E}\big[W_1(m_x^G, \mu_x^{\delta_n}) \,|\, \Omega_n\big] = o(\delta_n^3). \tag{8}$$

4.3 Approximating Probability Measures on Graph Balls

Recall that $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$ denotes the closed ball around $x \in \mathcal{M}$ with radius δ_n according to the manifold distance $d_{\mathcal{M}}$. The first step in establishing (8) is to move from uniform

measures on the graph balls $\mathcal{B}_G(x; \delta_n)$ to uniform measures on the nodes of the graph that lie in the manifold balls $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$. The reason for this is that $y \in \mathcal{B}_G(x; \delta_n)$ does not necessarily imply that $y \in \mathcal{B}_{\mathcal{M}}(x; \delta_n)$, nor vice versa. This creates difficulties when comparing the measures m_x^G and $\mu_x^{\delta_n}$.

Let $G_n = \mathbb{G}_n(x^*, \varepsilon_n)$ be rooted random graphs on $\mathcal{M}$. Then we define the probability measures $\boldsymbol{m}^{\mathcal{M}}$ on the nodes of G_n as

$$m_x^{\mathcal{M}}(y) = \begin{cases} \dfrac{1}{|\mathcal{B}_{\mathcal{M}}(x; \delta_n) \cap G_n|} & \text{if } y \in \mathcal{B}_{\mathcal{M}}(x; \delta_n) \cap G_n, \\ 0 & \text{else.} \end{cases} \tag{9}$$

Although the uniform measures $m_{x^*}^G$ and $m_{x^*}^{\mathcal{M}}$ are not the necessarily equal, the Wasserstein metric between them is sufficiently small.

Proposition 2 *Let $G_n = \mathbb{G}_n(x^*, \varepsilon_n)$ be rooted random graphs on $\mathcal{M}$ with graph distance d_G that is a δ_n-good approximation of $d_{\mathcal{M}}$. Let $x \in \mathcal{B}_{\mathcal{M}}(x^*; \delta_n)$ and denote by m_x^G the uniform measure on $\mathcal{B}_G(x; \delta_n)$ and by $m_x^{\mathcal{M}}$ the uniform measure on $\mathcal{B}_{\mathcal{M}}(x; \delta_n) \cap G_n$. Then*

$$\mathbb{E}\big[W_1\big(m_x^G, m_x^{\mathcal{M}}\big) \mid \Omega_n\big] = o(\delta_n^3).$$

The proof of this result is based on some simple computations regarding Poisson random variables and can be found in Sect. 5.2. Proposition 2 allows us to replace (8) with

$$\mathbb{E}\big[W_1\big(m_x^{\mathcal{M}}, \mu_x^{\delta_n}\big)\big] = o(\delta_n^3). \tag{10}$$

Note that the only dependence on the graph is now in the amount of nodes placed inside the ball $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$, which is completely determined by the Poisson process. All dependencies on the actual structure of the graph have been removed. This allows us to compute the Wasserstein metric between $m_x^{\mathcal{M}}$ and $\mu_x^{\delta_n}$.

4.4 Coupling Continuous and Discrete Probability Measures on $\mathcal{M}$

Recall that the Wasserstein metric $W_1(\mu_1, \mu_2)$ takes an infimum over all possible joint distributions (couplings) between the measures μ_1 and μ_2. Hence, to show that (10) holds, we need to design an optimal coupling (transport plan) between $m_x^{\mathcal{M}}$ and $\mu_x^{\delta_n}$. The main idea here is to view $m_x^{\mathcal{M}}$ as a discrete version of $\mu_x^{\delta_n}$.

For now, let us assume that we are working in the N-dimensional Euclidean cube $\mathcal{M} = [0, 1]^N$. Given a realization of the Poisson process, a transport plan between $m_x^{\mathcal{M}}$ and $\mu_x^{\delta_n}$ should assign to each measurable set $A \subseteq \mathcal{B}_{\mathcal{M}}(x; \delta_n)$ how much of the associated mass $\mu_x^{\delta_n}(A)$ is transported to each point of the Poisson process. To make it optimal, we should distribute the mass over those points that are closest to A. This problem is actually related to that of finding a minimal matching between points of a Poisson process and points of a grid on $[0, 1]^N$, see [22, 35, 39]. Here, minimal means

that the largest distance between a point of the Poisson process and its matched grid point is minimized. The idea for the transport plan is as follows:

- Place a grid on $[0, 1]^N$.
- Find a minimal matching between the Poisson process and the grid.
- Given a $A \subseteq \mathcal{B}_{\mathcal{M}}(x; \delta_n)$, we take all points of the Poisson process that are matched to grid points inside A and distribute the mass $\mu_x^{\delta_n}(A)$ equally over those points.

Using known results for minimal matchings, it can then be shown that, under suitable conditions, the Wasserstein metric between $m_x^{\mathcal{M}}$ and $\mu_x^{\delta_n}$ is $o(\delta_n^3)$.

Finally, we need to extend these results in flat Euclidean space to the ball $\mathcal{B}_{\mathcal{M}}(x; Q\delta_n)$ in general $\mathcal{M}$. For this we use that $\delta_n \to 0$ and that small neighborhoods of $x \in \mathcal{M}$ can be mapped diffeomorphically to the flat N-dimensional tangent space by the exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$. We then apply the matching results there and map back. Here we need to tread carefully, since the exponential map does not preserve distances. We thus fix a sufficiently small neighborhood U around the origin of the tangent space at x. Then, for some fixed $0 < \xi < 1$ and large enough n we have

$$\mathcal{B}_N\left(0; \frac{\delta_n}{1+\xi}\right) \subseteq \exp^{-1} \mathcal{B}_{\mathcal{M}}(x; \delta_n) \subseteq \mathcal{B}_N\left(0; \frac{\delta_n}{1-\xi}\right),$$

where $\mathcal{B}_N(0; \delta)$ is the Euclidean ball of radius δ. This then yields matching upper and lower bounds on the Wasserstein metric on $\mathcal{M}$ in terms of the Wasserstein metric on the Euclidean space. All details of this approach are provided in Sect. 5.3. In the end we obtain the following result.

Proposition 3 *For any point $x \in \mathcal{M}$,*

$$\mathbb{E}\big[W_1\big(m_x^{\mathcal{M}}, \mu_x^{\delta_n}\big)\big] = o(\delta_n^3).$$

4.5 Proof of the Main Results

We now have all ingredients to prove the main results. We start with Theorem 3, where we bound the expression inside the expectation as a sum of several terms and use the above results and the fact that d_G is a δ_n-good approximation to show that each individual term goes to zero.

Proof of Theorem 3 First, we bound the term inside the expectation as follows:

$$\left|\frac{2(N+2)\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \text{Ric}(\mathbf{v}, \mathbf{v})\right|$$

$$\leq \left|\frac{2(N+2)\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \frac{2(N+2)\kappa(x^*, y_n^*)}{\delta_n^2}\right|$$

$$+ \left|\frac{2(N+2)\kappa(x^*, y_n^*)}{\delta_n^2} - \text{Ric}(\mathbf{v}, \mathbf{v})\right|.$$

The last term is deterministic and goes to zero by Theorem 4. For the first term we note that the absolute value of each curvature term can be bounded from above by 2. Now let C_n denote the event that x^* and y_n^* are connected. Since this is implied by good event Ω_n, see Lemma 1, it follows that $C_n^c \subseteq \Omega_n^c$, where the superscript c denotes the complement of the event. Moreover, on the event C_n^c, $\kappa(x^*, y_n^*; \mathcal{G}_n) = 0$ by definition. Finally, since d_G is a δ_n-good approximation it follows that $\delta_n^2 = d_{\mathcal{M}}(x^*, y_n^*)^2 \leq d_G(x^*, y_n^*)^2$. Therefore, we have

$$\mathbb{E}\left[\left|\frac{2(N+2)\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \frac{2(N+2)\kappa(x^*, y_n^*)}{\delta_n^2}\right| \;\middle|\; C_n^c\right]$$
$$\leq \frac{4(N+2)}{\delta_n^2}(1 - \mathbb{P}(\Omega_n))$$

and

$$\mathbb{E}\left[\left|\frac{2(N+2)\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \frac{2(N+2)\kappa(x^*, y_n^*)}{\delta_n^2}\right| \;\middle|\; C_n \cap \Omega_n^c\right]$$
$$\leq \frac{8(N+2)}{\delta_n^2}(1 - \mathbb{P}(\Omega_n)).$$

It then follows that

$$\mathbb{E}\left[\left|\frac{2(N+2)\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \frac{2(N+2)\kappa(x^*, y_n^*)}{\delta_n^2}\right|\right]$$
$$\leq 2(N+2)\,\mathbb{E}\left[\left|\frac{\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \frac{\kappa(x^*, y_n^*)}{\delta_n^2}\right| \;\middle|\; \Omega_n\right] + (1 - \mathbb{P}(\Omega_n))\frac{12(2+N)}{\delta_n^2}.$$

By construction of the good event we have $1 - \mathbb{P}(\Omega_n) = o(\delta_n^3)$ and thus, the last term in the above bound goes to zero. For the other term we recall that

$$\kappa(x^*, y_n^*; \mathcal{G}_n) = 1 - \frac{W_1^G(m_{x^*}^G, m_{y_n^*}^G)}{d_G(x^*, y_n^*)} \quad \text{and} \quad \kappa(x^*, y_n^*) = 1 - \frac{W_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})}{\delta_n}.$$

Then the expression inside the conditional expectation can be bounded as follows:

$$\left|\frac{\kappa(x^*, y_n^*; \mathcal{G}_n)}{d_G(x^*, y_n^*)^2} - \frac{\kappa(x^*, y_n^*)}{\delta_n^2}\right|$$
$$\leq \left|\kappa(x^*, y_n^*; \mathcal{G}_n)\left(\frac{1}{d_G(x^*, y_n^*)^2} - \frac{1}{\delta_n^2}\right)\right| + \frac{|\kappa(x^*, y_n^*; \mathcal{G}_n) - \kappa(x^*, y_n^*)|}{\delta_n^2}$$
$$\leq 2\frac{|\delta_n^2 - d_G(x^*, y_n^*)^2|}{\delta_n^4} + \frac{1}{\delta_n^2}\left|\frac{W_1^G(m_{x^*}^G, m_{y_n^*}^G)}{d_G(x^*, y_n^*)} - \frac{W_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})}{\delta_n}\right|$$

$$\leq \frac{|\delta_n^2 - d_G(x^*, y_n^*)^2|}{\delta_n^4} + \frac{W_1^G\big(m_{x^*}^G, m_{y_n^*}^G\big)|\delta_n - d_G(x^*, y_n^*)|}{\delta_n^4} \tag{11}$$

$$+ \frac{\big|W_1^G\big(m_{x^*}^G, m_{y_n^*}^G\big) - W_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big)\big|}{\delta_n^3}. \tag{12}$$

Next, since d_G is a δ_n-good approximation, we can apply (4)

$$|\delta_n - d_G(x^*, y_n^*)| = |d_{\mathcal{M}}(x^*, y_n^*) - d_G(x^*, y_n^*)| \leq \delta_n \xi_n^2 + \xi_n^3 = o(\delta_n^3).$$

Since $W_1^G(m_{x^*}^G, m_{y_n^*}^G) \leq \delta_n$ it then follows that the second term in (11) goes to zero. For the first term we have

$$|\delta_n^2 - d_G(x^*, y_n^*)^2| \leq |\delta_n - d_G(x^*, y_n^*)|(\delta_n + d_G(x^*, y_n^*))$$
$$\leq (\delta_n \xi_n^2 + \xi_n^3)(\delta_n + \delta_n(1 + \xi_n^2) + \xi_n^3) = o(\delta_n^4),$$

which implies that this term also goes to zero. We are thus left with (12), for which we have to show that

$$\mathbb{E}\big[\big|W_1^G\big(m_{x^*}^G, m_{y_n^*}^G\big) - W_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big)\big| \,\big|\, \Omega_n\big] = o(\delta_n^3).$$

We first replace $W_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})$ with $\widetilde{W}_1(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n})$ by invoking Proposition 1:

$$\mathbb{E}\big[\big|\widetilde{W}_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big) - W_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big)\big| \,\big|\, \Omega_n\big] = o(\delta_n^3).$$

This then implies

$$\mathbb{E}\big[\big|W_1^G\big(m_{x^*}^G, m_{y_n^*}^G\big) - W_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big)\big| \,\big|\, \Omega_n\big]$$
$$\leq \mathbb{E}\big[\big|\widetilde{W}_1\big(m_{x^*}^G, m_{y_n^*}^G\big) - \widetilde{W}_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big)\big| \,\big|\, \Omega_n\big] + o(\delta_n^3).$$

To show that the first term in the upper bound is also $o(\delta_n^3)$ we apply the reverse triangle inequality twice to obtain

$$\big|\widetilde{W}_1\big(m_{x^*}^G, m_{y_n^*}^G\big) - \widetilde{W}_1\big(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\big)\big| \leq \widetilde{W}_1\big(m_{x^*}^G, \mu_{x^*}^{\delta_n}\big) + \widetilde{W}_1\big(m_{y_n^*}^G, \mu_{y_n^*}^{\delta_n}\big).$$

We proceed to show that $\widetilde{W}_1(m_{x^*}^G, \mu_{x^*}^{\delta_n}) = o(\delta_n^3)$ holds in expectation on the event Ω_n. The proof for $\widetilde{W}_1(m_{y_n^*}^G, \mu_{y_n^*}^{\delta_n})$ is similar. Applying Proposition 1 again we get

$$\mathbb{E}\big[\widetilde{W}_1\big(m_{x^*}^G, \mu_{x^*}^{\delta_n}\big) \,\big|\, \Omega_n\big] \leq \mathbb{E}\big[W_1\big(m_{x^*}^G, \mu_{x^*}^{\delta_n}\big) \,\big|\, \Omega_n\big] + o(\delta_n^3)$$
$$\leq \mathbb{E}\big[W_1\big(m_{x^*}^G, m_{x^*}^{\mathcal{M}}\big) \,\big|\, \Omega_n\big] + \mathbb{E}\big[W_1\big(m_{x^*}^{\mathcal{M}}, \mu_{x^*}^{\delta_n}\big)\big] + o(\delta_n^3).$$

Since both expectations are $o(\delta_n^3)$ by, respectively, Propositions 2 and 3, we conclude that

$$\mathbb{E}\left[\left|W_1^G\left(m_{x^*}^G, m_{y_n^*}^G\right) - W_1\left(\mu_{x^*}^{\delta_n}, \mu_{y_n^*}^{\delta_n}\right)\right| \,\Big|\, \Omega_n\right] = o(\delta_n^3),$$

which finishes the proof.   □

Now that we have the general result, Theorems 1 and 2 directly follow from Theorem 3 if we can show that the graph distances that are considered there are δ_n-good approximations.

Throughout the remainder of this section we will assume that

$$\varepsilon_n = \Theta((\log n)^a n^{-\alpha}), \quad \delta_n = \Theta((\log n)^b n^{-\beta}),$$

for some $a, b \in \mathbb{R}$ and $0 \le \alpha, \beta \le 1$. We shall also assume that $\varepsilon_n \le \delta_n$. The following results show that for appropriate choices of the constants a, b and α, β both the weighted manifold and the rescaled hopcount distance are δ_n-good approximations. The proofs are given in Sects. 5.4 and 5.5, respectively.

Proposition 4 *Suppose the constants in ε_n and δ_n satisfy*

$$0 < \beta \le \alpha, \quad \alpha + 2\beta \le \frac{1}{N}$$

with $a \le b$ if $\alpha = \beta$ and $a + 2b > 2/N$ if $\alpha + 2\beta = 1/N$. Let $y_n^* \in \mathcal{M}$ be at distance δ_n in the direction of $\mathbf{v}$ and $G_n = \mathbb{G}_n(x^*, y_n^*, \varepsilon_n)$ be rooted random graphs on $\mathcal{M}$. Then the manifold-weighted graph distance d_G^w on G_n is a δ_n-good approximation of $d_{\mathcal{M}}$.

Proposition 5 *Suppose the constants in ε_n and δ_n satisfy*

$$0 < \beta \le \frac{1}{9} \quad and \quad 3\beta \le \alpha \le \frac{1 - 3\beta}{2},$$

and $a < 3b$ if $\alpha = 3\beta$ and $2a + 3b > 1$ if $\alpha = (1 - 3\beta)/2$. Let $y_n^* \in \mathcal{M}$ be at distance δ_n in the direction of $\mathbf{v}$. Let $G_n = \mathbb{G}_n(x^*, y_n^*, \varepsilon_n)$ be rooted random graphs on a 2-dimensional Riemannian manifold $\mathcal{M}$ and denote by d_G^s the shortest path distance. Then the ε_n-weighted graph distance $d_G^* := \varepsilon_n d_G^s$ on G_n is a δ_n-good approximation of $d_{\mathcal{M}}$.

Observe that the conditions of the constants in Propositions 4 and 5 are exactly the same as in Theorems 1 and 2, respectively. Moreover, these conditions imply that $\lambda_n = o(\varepsilon_n)$ and $\lambda_n = o(\delta_n^3)$, with λ_n as defined in (5), as we will now demonstrate.

In Proposition 4 we have $\beta > 0$ and $\alpha + 2\beta \le 1/N$. It then follows that $\alpha < 1/N$ which implies $\lambda_n = o(\varepsilon_n)$. When the inequality $3\beta \le \alpha + 2\beta \le 1/N$ is strict we have that $\lambda_n = o(\delta_n^3)$. When $3\beta = 1/N$ it must be that $\alpha + 2\beta = 1/N$ and hence the conditions of Proposition 4 imply that $3b \ge a + 2b > 2/N$. From this we deduce that $\lambda_n/\delta_n^3 = \Theta((\log n)^{2/N-a-2b}) = o(1)$.

In Proposition 5, since $N = 2$, the conditions $\lambda_n = o(\varepsilon_n)$ and $\lambda_n = o(\delta_n^3)$ follow if $\alpha < 1/2$ and $3\beta < 1/2$. The first inequality holds since $\beta > 0$ and $\alpha \le (1 - 3\beta)/2$, while the second is due to the fact that $3\beta \le 3/9 = 1/3$.

We thus conclude that under the conditions in both propositions, the radii satisfy the conditions of Theorem 3. Hence, Theorems 1 and 2 follow from it.

5 Proofs

Here we prove all the intermediate results that we used to prove our main results in the previous section. We start with the proof of Lemma 1 and Proposition 1 in the next Sect. 5.1. In Sect. 5.2 we provide the details for Proposition 2, while the proof of Proposition 3 is given in Sect. 5.3. We end with Sects. 5.4 and 5.5 where we prove Propositions 4 and 5, respectively, leading to the main results of this paper.

Recall that

$$\lambda_n = (\log n)^{2/N} n^{-1/N},$$

and $\varepsilon_n \le \delta_n \to 0$ are such that $\lambda_n = o(\varepsilon_n)$ and $\lambda_n = o(\delta_n^3)$.

5.1 Extended Graph Distance

Our first goal is to proof Lemma 1. We start by showing that there exists a radius $r_n \to 0$ such that for any finite set of points $u \in \mathcal{M}$, the balls $\mathcal{B}_{\mathcal{M}}(u; r_n)$ will still each contain at least one node from the rooted graphs $G_n = \mathbb{G}_n(x^*, y_n^*, \varepsilon_n)$. The reason why we need r_n to decrease is because the connection radius ε_n also decreases and we want the ball $\mathcal{B}_{\mathcal{M}}(u; r_n)$ to be contained inside the connection area of the point u.

Lemma 2 *Let $U \subset \mathcal{M}$ be a finite set of points in $\mathcal{M}$ such that $|U| = O(n^c)$, for some $c > 0$, and let $r_n = \Theta(\lambda_n)$. Then, for $G_n = \mathbb{G}_n(\varepsilon_n)$,*

$$\mathbb{P}\left(\bigcup_{u \in U}\{|\mathcal{B}_{\mathcal{M}}(u; r_n) \cap G_n| = 0\}\right) = o(\delta_n^3),$$

as $n \to \infty$.

Proof First note that for r_n small enough the ball $\mathcal{B}_{\mathcal{M}}(u; r_n)$ can be mapped diffeomorphically onto the tangent space $T_u\mathcal{M}$ at u. In particular, for small enough r_n we have that, as $n \to \infty$, $\mathrm{vol}_{\mathcal{M}}(\mathcal{B}_{\mathcal{M}}(u; r_n)) = \Theta(r_n^N) = \Theta(\lambda_n^N)$. Next, since the nodes in G_n are placed according to a Poisson process with intensity $n/\mathrm{vol}_{\mathcal{M}}(\mathcal{M})$ it follows that

$$\mathbb{P}\left(|\mathcal{B}_{\mathcal{M}}(u; r_n) \cap G_n| = 0\right) = \exp\left(-\frac{n\mathrm{vol}_{\mathcal{M}}(\mathcal{B}_{\mathcal{M}}(u; \lambda_n))}{\mathrm{vol}_{\mathcal{M}}(\mathcal{M})}\right) = e^{-n\Theta(\lambda_n^N)}$$

$$= e^{-\Theta((\log n)^2)}.$$

Therefore, by applying the union bound we get

$$\mathbb{P}\left(\bigcup_{u \in U} \{|\mathcal{B}_{\mathcal{M}}(u; \varepsilon_n) \cap G| = 0\} \right) \leq |U| \min_{u \in U} \mathbb{P}\left(|\mathcal{B}_{\mathcal{M}}(u; \varepsilon_n) \cap G| = 0 \right)$$

$$= e^{-\Theta((\log n)^2) + \log |U|} \leq e^{-\Theta((\log n)^2) + c \log n}.$$

To finish the proof we note that $e^{-\Theta((\log n)^2) + c \log n} = o(\lambda_n)$ which by assumption is $o(\delta_n^3)$. □

With this lemma we obtain the following corollary.

Corollary 2 *There exists a collection* $\{B_1, \ldots, B_m\}$ *of* $m = \Theta(\lambda_n^{-N})$ *balls of radius* $\lambda_n/4$ *that cover* $\mathcal{M}$*, such that if we denote by* $c_1, \ldots, c_m$ *their centers and define the event*

$$C_n = \bigcap_{t=1}^{m} \{|\mathcal{B}_{\mathcal{M}}(c_t; \lambda_n/4) \cap G_n| \neq 0\}. \tag{13}$$

Then $\mathbb{P}(C_n) = 1 - o(\delta_n^3)$.

Proof The collection is constructed using the standard trick of taking a maximal set of *disjoint* balls of radius $\lambda_n/8$ in $\mathcal{M}$. Denote their centers by $c_1, \ldots, c_m$. Simple volume comparison, and the compactness of $\mathcal{M}$, gives $m = O(\lambda_n^{-N})$. By construction, the balls $B_i = \mathcal{B}_{\mathcal{M}}(c_i; \lambda_n/4)$ then cover $\mathcal{M}$, and hence $m = \Theta(\lambda_n^{-N}) = \Theta((\log n)^{-2} n) = O(n)$. The result then follows from Lemma 2. □

The event C_n will play a crucial part in defining the *good event* Ω_n. Let D_n denote the event on which (4) holds. Then we define the *good event* as

$$\Omega_n = C_n \cap D_n. \tag{14}$$

On this event, with sufficiently high probability, $(\mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n), \widetilde{d}_{\mathcal{M}})$ is a metric space for any constant $Q > 0$ and the extended distance $\widetilde{d}_{\mathcal{M}}$ is a good approximation of the original distance $d_{\mathcal{M}}$. Note that we do not need to consider the whole manifold since curvature is a local property.

Lemma 3 *Let* Ω_n *be the event defined in* (14) *and* $Q > 3$ *the constant from Definition 4.1. Then on the event* Ω_n,

- *each pair of points* $u, v \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$ *is connected by a path in the extended graph* $G_n(u, v)$, *and*
- $(\mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n), \widetilde{d}_{\mathcal{M}})$ *is a metric space.*
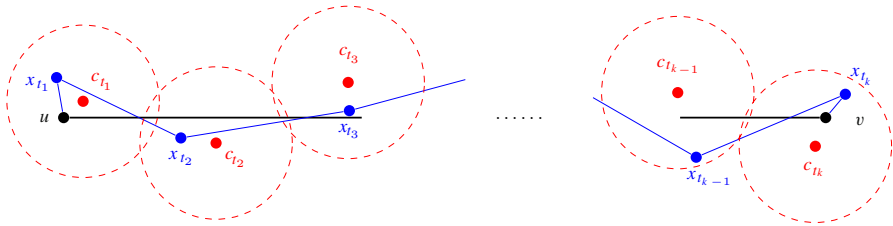
Fig. 2 Depiction of the covering of the geodesic between u and v by the balls B_{t_i}

Proof We first prove the first statement. For this, take any $u, v \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$ and let $\gamma(u, v)$ denote the geodesic between u and v. This geodesic will be covered by a subsequence $B_{t_1}, \ldots, B_{t_k}$ of the cover of $\mathcal{M}$, which we rank in order of appearance moving from u to v. Let $c_{t_1}, \ldots, c_{t_k}$ denote the corresponding centers of these balls, see Fig. 2. On the event C_n each ball contains a vertex $x_{t_i} \in G_n$ and since

$$d_{\mathcal{M}}(u, x_{t_1}), d_{\mathcal{M}}(v, x_{t_k}) \leq 2\frac{\lambda_n}{4} = \frac{\lambda_n}{2},$$

the edges (u, x_{t_1}) and (v, x_{t_k}) are present in $G_n(u, v)$. Moreover, since $d_{\mathcal{M}}(x_{t_i}, x_{t_{i+1}})$ is bounded by four times the radius of the balls, it follows that for large enough n, $d_{\mathcal{M}}(x_{t_i}, x_{t_{i+1}}) \leq \lambda_n = o(\varepsilon_n)$ and thus, for n large enough, $\{x_{t_1}, \ldots, x_{t_k}\}$ is a path in G_n. We thus conclude that u and v are connected in $G_n(u, v)$. Note that because of this property, on the event Ω_n, the extended manifold distance between $\tilde{d}_{\mathcal{M}}$ is well defined on $\mathcal{M}$.

We are left to show that on the event Ω_n, the extended manifold distance is a true distance. Note that the only non-trivial part is the triangle inequality. Let $u, v, z \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$ and consider the graphs $G^{(1)} = G_n(u, v)$ and $G^{(2)} = G_n(u, v, z)$. Now observe that the triangle inequality can only be violated if z creates a short-cut, i.e., if the shortest weighted path between u and v in $G^{(1)}$ is longer than in $G^{(2)}$. Suppose that this is true, and let $\pi_1 = \{u, \ldots, y_1, z, y_2, \ldots, v\}$ denote this new weighted shortest path in $G^{(2)}$. Since y_1 and y_2 are connected to z in $G^{(2)}$ it follows that $d_{\mathcal{M}}(z, y_i) \leq \lambda_n/2$. However, by the triangle inequality for $d_{\mathcal{M}}$, this implies that $d_{\mathcal{M}}(y_1, y_2) \leq \lambda_n = o(\varepsilon_n)$ and hence, for sufficiently large n, the edge (y_1, y_2) is present in G_n and thus also in $G^{(1)}$ and $G^{(2)}$.

Let $\hat{\pi} = \{y_1 := x_0, x_1, \ldots, x_{m-1}, y_2 := x_m\}$ denote the shortest weighted path in G_n between y_1 and y_2, i.e., $d_G(y_1, y_2) = \sum_{t=1}^m w_{x_{t-1}x_t}$, and take $\pi_2 = \{u, \ldots, y_1, x_1, \ldots, x_{m-1}, y_2, \ldots, v\}$. Then π_2 is a path between u and v that excludes z. See also Fig. 3. We will show that the total weight of this path is at most that of π_1.

For simplicity lets us denote by $\|\pi\|$ the total weight of a path π. Since d_G is a δ_n-good approximation,

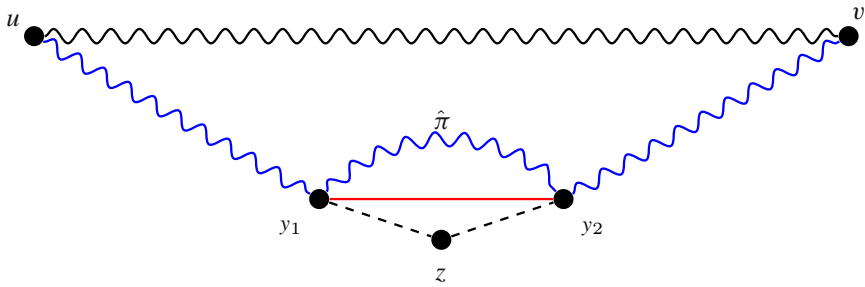$$\|\hat{\pi}\| := \sum_{t=1}^m w_{x_{t-1}x_t} = d_G(y_1, y_2) \leq d_{\mathcal{M}}(y_1, y_2)(1 + \xi_n^2) + \xi_n^3$$

Fig. 3 Abstract depiction of the weighted shortest path between u and v created by adding z and the path π_2, given in blue

holds on the event Ω_n. Applying the triangle inequality for $d_{\mathcal{M}}$ we get

$$
\begin{aligned}
\|\hat{\pi}\| &\leq d_{\mathcal{M}}(y_1, z)(1 + \xi_n^2) + d_{\mathcal{M}}(z, y_2)(1 + \xi_n^2) + \xi_n^3 \\
&\leq d_{\mathcal{M}}(y_1, z)(1 + \xi_n^2) + d_{\mathcal{M}}(z, y_2)(1 + \xi_n^2) + 2\xi_n^3 = w_{y_1 z} + w_{y_2 z}.
\end{aligned}
$$

This implies that the total weight of the path π_2 is at most that of π_1 from which we conclude that z cannot create a short-cut and hence $\widetilde{d}_{\mathcal{M}}$ satisfies the triangle inequality. $\qquad\square$

We are now ready to prove Lemma 1.

Proof of Lemma 1 Note that for any two nodes $u, v \in G_n$ with $u, v \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$, Lemma 3 implies that u and v are connected by a path in G_n. Hence the only part of Lemma 1 to prove is property $(\Omega 1)$ there.

Take any $u, v \in \mathcal{B}_{\mathcal{M}}(x^*; 3\delta_n)$. Then on the event Ω_n, by definition of the extended distance $\widetilde{d}_{\mathcal{M}}$, there exists $x_u, x_v \in G_n$ such that $d_{\mathcal{M}}(u, x_u) \leq \lambda_n/2$, $d_{\mathcal{M}}(v, x_v) \leq \lambda_n/2$, and

$$
\begin{aligned}
\widetilde{d}_{\mathcal{M}}(u, v) &= d_{\mathcal{M}}(u, x_u)(1 + \xi_n^2) + d_{\mathcal{M}}(v, x_v)(1 + \xi_n^2) + 2\xi_n^3 + d_G(x_u, x_v) \\
&\leq \lambda_n(1 + \xi_n^2) + 2\xi_n^3 + d_G(x_u, x_v).
\end{aligned}
\tag{15}
$$

Moreover, since $Q > 3$ and $\lambda_n = o(\delta_n^3)$ we can assume that $x_u, x_v \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$, for sufficiently large n. Since the approximation (4) holds on the event Ω_n, we have

$$
\begin{aligned}
|d_G(x_u, x_v) - d_{\mathcal{M}}(u, v)| &\leq |d_G(x_u, x_v) - d_{\mathcal{M}}(x_u, x_v)| + |d_{\mathcal{M}}(x_u, x_v) - d_{\mathcal{M}}(u, v)| \\
&\leq |d_G(x_u, x_v) - d_{\mathcal{M}}(x_u, x_v)| + d_{\mathcal{M}}(x_u, u) + d_{\mathcal{M}}(x_v, v) \\
&\leq d_{\mathcal{M}}(x_u, x_v)\xi_n^2 + \xi_n^3 + d_{\mathcal{M}}(x_u, u) + d_{\mathcal{M}}(x_v, v) \\
&\leq d_{\mathcal{M}}(x_u, x_v)\xi_n^2 + \xi_n^3 + \lambda_n.
\end{aligned}
\tag{16}
$$

Combining (15) and (16) we get

$$
|\widetilde{d}_{\mathcal{M}}(u, v) - d_{\mathcal{M}}(u, v)| \leq |\widetilde{d}_{\mathcal{M}}(u, v) - d_G(x_u, x_v)| + |d_G(x_u, x_v) - d_{\mathcal{M}}(u, v)|
$$

$$\leq \lambda_n(1 + \xi_n^2) + 2\xi_n^3 + |d_G(x_u, x_v) - d_{\mathcal{M}}(u, v)|$$
$$\leq \lambda_n(1 + \xi_n^2) + d_{\mathcal{M}}(x_u, x_v)\xi_n^2 + 3\xi_n^3 + \lambda_n.$$

Applying the triangle inequality to the last distance,

$$d_{\mathcal{M}}(x_u, x_v) \leq d_{\mathcal{M}}(u, v) + d_{\mathcal{M}}(u, x_u) + d_{\mathcal{M}}(v, x_v) \leq d_{\mathcal{M}}(u, v) + \lambda_n,$$

we get

$$|\widetilde{d}_{\mathcal{M}}(u, v) - d_{\mathcal{M}}(u, v)| \leq d_{\mathcal{M}}(u, v)\xi_n^2 + 2\lambda_n(1 + \xi_n^2) + 3\xi_n^3 = o(\delta_n^3). \qquad \square$$

Finally, we need to prove Proposition 1. Since, on the event Ω_n,

$$|\widetilde{d}_{\mathcal{M}}(u, v) - d_{\mathcal{M}}(u, v)| \leq o(\delta_n^3),$$

the proof follows immediately from the following elementary result on Wasserstein metrics.

Lemma 4 *Let $(\mathcal{X}, d)$ and $(\mathcal{X}, \widetilde{d})$ be two metric spaces such that*

$$|d(x, y) - \widetilde{d}(x, y)| \leq K$$

holds for all $x, y \in \mathcal{X}$ and some $K > 0$. Denote by W_1 and $\widetilde{W}_1$ the Wasserstein metric associated with d and $\widetilde{d}$, respectively. Then for any two probability measures μ_1 and μ_2 on $\mathcal{X}$,

$$|\widetilde{W}_1(\mu_1, \mu_2) - W_1(\mu_1, \mu_2)| \leq K.$$

Proof For any coupling μ between μ_1 and μ_2,

$$\int \widetilde{d}(x, y) \, d\mu(x, y) \leq \int (d(x, y) + K) \, d\mu(u, v) \leq \int d(x, y) \, d\mu(x, y) + K$$

and similarly

$$\int d(x, y) \, d\mu(x, y) \geq \int \widetilde{d}(x, y) \, d\mu(x, y) + K$$

Then it follows that

$$\widetilde{W}_1(\mu_1, \mu_2)$$
$$= \inf_\mu \int \widetilde{d}(x, y) \, d\mu(x, y) \leq \inf_\mu \int d(x, y) \, d\mu(x, y) + K = W_1(\mu_1, \mu_2) + K$$

and similarly

$$W_1(\mu_1, \mu_2) \leq \widetilde{W}_1(\mu_1, \mu_2) + K,$$

from which the result follows.   □

5.2 Probability Measures on Graphs

In this section we give the proof of Proposition 2. Recall that m_x^G and $m_x^{\mathcal{M}}$ denote the uniform probability measures on the set of nodes in $\mathcal{B}_G(x; \delta_n)$ and $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$, respectively. The goal is then to show that

$$\mathbb{E}\big[W_1(m_x^G, m_x^{\mathcal{M}})\big] = o(\delta_n^3).$$

As we mentioned, these two sets are not necessarily contained in each other. Hence, to bound the Wasserstein metric we will work with slightly smaller and larger balls B^- and B^+ such that

$$B^- \cap G_n \subseteq \mathcal{B}_G(x; \delta_n), \mathcal{B}_{\mathcal{M}}(x; \delta_n) \cap G_n \subseteq B^+ \cap G_n.$$

We can then obtain an upper bound by comparing the Wasserstein metric between m_x^G, $m_x^{\mathcal{M}}$, and the uniform probability measure on $B^+ \cap G_n$. This bound can be made $o(\delta_n^3)$, by carefully selecting the radii of B^- and B^+.

Before we give the details, we need the following general result concerning Poisson random variables.

Lemma 5 *Let $\alpha_n, \beta_n \to \infty$ and X_n, Y_n be two independent Poisson random variables with means α_n and β_n, respectively. Then*

$$\mathbb{E}\left[\frac{X_n}{X_n + Y_n} \,\Big|\, X_n + Y_n \geq 1\right] = O\left(\frac{\alpha_n}{\alpha_n + \beta_n}\right).$$

Proof First, let $C > \sqrt{2}$ be some large fixed constant. Then we have that (c.f. [32, Lemma 2.1])

$$\mathbb{P}\big(|X_n - \alpha_n| > C\sqrt{\alpha_n \log \alpha_n}\big) = O\big(\alpha_n^{-C^2/2}\big).$$

In particular, if we define $\alpha_n^{\pm} = \alpha_n \pm C\sqrt{\alpha_n \log \alpha_n}$, then

$$\max\{\mathbb{P}(X_n < \alpha_n^-), \mathbb{P}(X_n > \alpha_n^+)\} = O\big(\alpha_n^{-C^2/2}\big).$$

Similar results hold for Y_n with $\beta_n^{\pm}$ defined similarly. We start by conditioning on X_n:

$$\mathbb{E}\left[\frac{X_n}{X_n + Y_n} \,\Big|\, X_n + Y_n \geq 1\right] = \sum_{k=0} \mathbb{E}\left[\frac{k}{k + Y_n} \,\Big|\, Y_n \geq 1\right] \cdot \mathbb{P}(X_n = k)$$

$$= \sum_{k < \alpha_n^-} \mathbb{E}\left[\frac{k}{k + Y_n} \,\Big|\, Y_n \geq 1\right] \cdot \mathbb{P}(X_n = k)$$

$$+ \sum_{k \geq \alpha_n^-} \mathbb{E}\left[\frac{k}{k + Y_n} \,\Big|\, Y_n \geq 1\right] \cdot \mathbb{P}\left(X_n = k\right) := I_n^{(1)} + I_n^{(2)}.$$

We will bound each term separately.

First we bound the expectation inside each summation by further conditioning on Y_n:

$$\mathbb{E}\left[\frac{k}{k + Y_n} \,\Big|\, Y_n \geq 1\right] \leq \frac{k}{k+1}\mathbb{P}(1 \leq Y_n < \beta_n^-)$$

$$+ \sum_{\beta_n^- \leq y \leq \beta_n^+} \frac{k}{k + y}\mathbb{P}(Y_n = y) + \frac{k}{k + \beta_n^+}\mathbb{P}(Y_n > \beta_n^+)$$

$$\leq k\left(\frac{1}{k + \beta_n^-} + \frac{\mathbb{P}\big(|Y_n - \beta_n| > C\sqrt{\beta_n \log \beta_n}\big)}{k + 1}\right)$$

$$\leq \frac{k}{k + \beta_n^-}\big(1 + O\big(\beta_n^{1 - C^2/2}\big)\big) = \frac{k}{k + \beta_n^-}(1 + o(1)),$$

because $C > \sqrt{2}$. We can now bound $I_n^{(1)}$ as follows:

$$I_n^{(1)} \leq \frac{\alpha_n^-}{\beta^-}\mathbb{P}(X_n < \alpha_n^-) = O\big((\beta_n^-)^{-1}\alpha_n^{1 - C^2/2}\big) = O\big(\beta_n^{-1}\alpha_n^{1 - C^2/2}\big),$$

where we used that $\beta_n^- \sim \beta_n$, i.e., $\beta_n^- / \beta_n \to 1$. For $I_n^{(2)}$ we have, using that $\alpha_n^- \sim \alpha_n$,

$$I_n^{(2)} \leq (1 + o(1)) \sum_{k \geq \alpha_n^-} \frac{k}{k + \beta_n^-}\mathbb{P}\left(X_n = k\right) \leq O\left(\frac{\mathbb{E}[X_n]}{\alpha_n^- + \beta_n^-}\right) = O\left(\frac{\alpha_n}{\alpha_n + \beta_n}\right),$$

and thus the result follows since we are free to select $C > \sqrt{2}$ large enough so that $I_n^{(1)}$ is of smaller order. □

We are now ready to prove Proposition 2.

Proof of Proposition 2 Let $\delta_n^\pm = (\delta_n \pm \xi_n^3)/(1 \mp \xi_n^2)$ and let D_n be the event on which approximation (4) of Definition 4.1 holds and recall that $\Omega_n \subset D_n$. Therefore, since $\mathbb{P}\left(\Omega_n\right) \to 1$,

$$\mathbb{E}\big[W_1(m_x^G, m_x^{\mathcal{M}}) \,|\, \Omega_n\big] \leq (1 + o(1))\,\mathbb{E}\big[W_1(m_x^G, m_x^{\mathcal{M}})\mathbb{1}_{\{D_n\}}\big],$$

and so it suffices to look at $\mathbb{E}\big[W_1(m_x^G, m_x^{\mathcal{M}})\mathbb{1}_{\{D_n\}}\big]$.

Note that on the event D_n,

$$\mathcal{B}_{\mathcal{M}}(x; \delta_n^-) \cap G_n \subseteq \mathcal{B}_G(x; \delta_n), \quad \mathcal{B}_{\mathcal{M}}(x; \delta_n) \cap G_n \subseteq \mathcal{B}_{\mathcal{M}}(x; \delta_n^+) \cap G_n.$$

Let $V_n \subseteq \mathcal{M}$ be any neighborhood of x such that $\mathrm{vol}_{\mathcal{M}}(\mathcal{B}_n) = \Theta(\delta_n^N)$ and

$$\mathcal{B}_{\mathcal{M}}(x; \delta_n^-) \subseteq \mathcal{B}_n \subseteq \mathcal{B}_{\mathcal{M}}(x; \delta_n^+) \cap G_n,$$

where $\mathcal{B}_n = V_n \cap G_n$. Denote by m_n the uniform probability measure on $\mathcal{B}_n$. We will prove that

$$\mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \big] = o(\delta_n^3). \tag{17}$$

Since

$$\mathbb{E}\big[W_1(m_x^G, m_x^{\mathcal{M}}) \mathbb{1}_{\{D_n\}} \big] \leq \mathbb{E}\big[W_1(m_x^G, m_x^+) \mathbb{1}_{\{D_n\}} \big] + \mathbb{E}\big[W_1(m_x^{\mathcal{M}}, m_x^+) \mathbb{1}_{\{D_n\}} \big],$$

applying (17) twice, once with $\mathcal{B}_n = \mathcal{B}_G(x; \delta_n)$ and once with $\mathcal{B}_n = \mathcal{B}_{\mathcal{M}}(x; \delta_n) \cap G_n$, will yield the required result.

Let us write $\mathcal{B}_n^{\pm} := \mathcal{B}_{\mathcal{M}}(x; \delta_n^{\pm}) \cap G_n$ and denote by $m_x^{\pm}$ the uniform probability measure on $\mathcal{B}_n^{\pm}$. To establish (17) we will show that

$$\mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \big] = O\left(\frac{(\delta_n^+)^N - (\delta_n^-)^N}{(\delta_n^+)^{N-1}} \right). \tag{18}$$

Note that by definition of $\delta_n^{\pm}$ we have $(\delta_n^+)^N - (\delta_n^-)^N = O(\xi_n^2 \delta_n^N)$. Therefore, if (18) holds,

$$\mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \big] \leq O\left(\frac{(\delta_n^+)^N - (\delta_n^-)^N}{\delta_n^{N-1}} \right) = O(\delta_n \xi_n^2) = o(\delta_n^3),$$

since $\xi_n = o(\delta_n)$. To establish (18) we condition on $|\mathcal{B}_n^-|$:

$$\mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \big] = \mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \,|\, |\mathcal{B}_n^-| = 0 \big] \cdot \mathbb{P}(|\mathcal{B}_n^-| = 0)$$
$$+ \mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \,|\, |\mathcal{B}_n^-| \geq 1 \big] \cdot \mathbb{P}(|\mathcal{B}_n^-| \geq 1).$$

For the first term we have

$$\mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \,|\, |\mathcal{B}_n^-| = 0 \big] \cdot \mathbb{P}(|\mathcal{B}_n^-| = 0) \leq 2\delta_n^+ \mathbb{P}(|\mathcal{B}_n^-| = 0)$$
$$= O(\delta_n^-) e^{-n\Theta((\delta_n^-)^N)} = O\left(\frac{(\delta_n^+)^N - (\delta_n^-)^N}{(\delta_n^+)^{N-1}} \right),$$

where we used that $\mathbb{E}[|\mathcal{B}_n^-|] = n\,\mathrm{vol}_{\mathcal{M}}(\mathcal{B}_n^-) = n\Theta((\delta_n^-)^N)$. It now suffices to show that

$$\mathbb{E}\big[W_1(m_n, m_x^+) \mathbb{1}_{\{D_n\}} \,|\, |\mathcal{B}_n^-| \geq 1 \big] = O\left(\frac{(\delta_n^+)^N - (\delta_n^-)^N}{(\delta_n^+)^{N-1}} \right). \tag{19}$$

We will do this by constructing a specific transport plan (coupling) between the measures m_n and m_x^+. Define the joint probability mass function on $\mathcal{B}_n \times \mathcal{B}_n^+$:

$$
m(u, v) = \begin{cases} \dfrac{1}{|\mathcal{B}_n^+|} & \text{if } u = v, \\ \dfrac{1}{|\mathcal{B}_n| \cdot |\mathcal{B}_n^+|} & \text{if } v \in \mathcal{B}_n^+ \setminus \mathcal{B}_n, \end{cases}
$$

and observe that $m(u, v)$ is a coupling between m_x^G and m_x^+. Therefore

$$
W_1(m_x^G, m_x^+) \leq \sum_{u \in \mathcal{B}_n} \sum_{v \in \mathcal{B}_n^+} d_{\mathcal{M}}(u, v) m(u, v) = \sum_{u \in \mathcal{B}_n} \sum_{v \in \mathcal{B}_n^+ \setminus \mathcal{B}_n} \frac{d_{\mathcal{M}}(u, v)}{|\mathcal{B}_n| \cdot |\mathcal{B}_n^+|}
$$

$$
\leq 2\delta_n^+ \frac{|\mathcal{B}_n^+| - |\mathcal{B}_n|}{|\mathcal{B}_n^+|} \leq 2\delta_n^+ \frac{|\mathcal{B}_n^+| - |\mathcal{B}_n^-|}{|\mathcal{B}_n^+|} = 2\delta_n^+ \frac{|\mathcal{B}_n^+ \setminus \mathcal{B}_n^-|}{|\mathcal{B}_n^+|}.
$$

Now define $X_n = |\mathcal{B}_n^+ \setminus \mathcal{B}_n^-|$ and $Y_n = |\mathcal{B}_n^-|$. Then X_n and Y_n are independent Poisson random variables satisfying

$$
\frac{|\mathcal{B}_n^+ \setminus \mathcal{B}_n^-|}{|\mathcal{B}_n^+|} = \frac{X_n}{X_n + Y_n}.
$$

It then follows from Lemma 5 that

$$
\mathbb{E}\big[W_1(m_x, m_x^+)\,|\,|\mathcal{B}_n^-| \geq 1\big] \leq O\left(\frac{\delta_n^+ \mathbb{E}[X_n]}{\mathbb{E}[X_n] + \mathbb{E}[Y_n]}\right) = O\left(\frac{\delta_n^+ \operatorname{vol}_{\mathcal{M}}(\mathcal{B}_n^+ \setminus \mathcal{B}_n)}{\operatorname{vol}_{\mathcal{M}}(\mathcal{B}_n^+)}\right).
$$

Equation (19) then follows by noting that $\operatorname{vol}_{\mathcal{M}}(\mathcal{B}_n^+ \setminus \mathcal{B}_n) = \Theta((\delta_n^+)^N - (\delta_n^-)^N)$. □

5.3 Continuous and Discrete Measures on $\mathcal{M}$

5.3.1 Collecting Relevant Known Results

The following is a summary of results on the Wasserstein metric between empirical and uniform measures on the N-dimensional cube. The case $N = 2$ was explicitly stated in [39]. Although the results for $N \geq 3$ are known, they are not stated in the explicit form we need. For completeness we thus include a proof here.

Proposition 6 *Let $X_1, X_2, \ldots$ be independent uniformly distributed random variables on $[0, 1]^N$, let m_n denote the empirical measure*

$$
m_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i = y\}},
$$

and μ the uniform measure on $[0, 1]^N$. Then

$$\mathbb{E}\big[W_1^N(m_n, \mu)\big] = \begin{cases} O\left(\sqrt{\dfrac{\log n}{n}}\right) & \text{if } N = 2, \\ O(n^{-1/N}) & \text{if } N \geq 3. \end{cases}$$

Proof The result for $N = 2$ follows from [39, (1.1)], see also the results in [22, 35]. For $N \geq 3$ we let $Y_1, Y_2, \ldots$ be independent uniformly distributed random variables on $[0, 1]^N$ and define

$$M_n := \inf_\sigma \sum_{i=1}^n \|X_i - Y_{\sigma(i)}\|,$$

where the infimum is taken over all permutations σ of $\{1, 2, \ldots, n\}$. Then, it follows from [38, Lemma 1] that

$$M_n = \sup_{f \in \mathrm{Lip}_1} \left| \sum_{i=1}^n (f(X_i) - f(Y_i)) \right|,$$

where Lip_1 now denotes the set of Lipschitz continuous functions with constant 1, with respect to the Euclidean distance d_N.

Next, we recall the duality formula for the Wasserstein metric on the space $\mathcal{X}$,

$$W_1(\mu_1, \mu_2) = \sup_{f \in \mathrm{Lip}_1} \left\{ \int_{\mathcal{X}} f(x) \, d\mu_1(x) - \int_{\mathcal{X}} f(y) \, d\mu_2(y) \right\}.$$

Since

$$\int_{[0,1]^N} f(z) \, d\mu(z) = \mathbb{E}[f(Y_i)],$$

we have

$$\begin{aligned}
W_1^N(m_n, \mu) &= \sup_{f \in \mathrm{Lip}_1} \left| \frac{1}{n} \sum_{i=1}^n \left(f(X_i) - \int_{[0,1]^N} f(z) \, d\mu(z) \right) \right| \\
&= \frac{1}{n} \sup_{f \in \mathrm{Lip}_1} \left| \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(Y_i)]) \right| \\
&\leq \frac{1}{n} \mathbb{E}\left[\sup_{f \in \mathrm{Lip}_1} \left| \sum_{i=1}^n (f(X_i) - f(Y_i)) \right| \,\Big|\, X_1, \ldots, X_n \right] \\
&= \frac{\mathbb{E}[M_n \mid X_1, \ldots, X_n]}{n},
\end{aligned}$$

and hence

$$\mathbb{E}\big[W_1^N(m_X, \mu)\big] \leq \frac{\mathbb{E}[M_n]}{n}.$$

Finally [38, Thm. 1] implies for $N \geq 3$,

$$\mathbb{E}[M_n] = O(n^{1-1/N}),$$

which then yields

$$\mathbb{E}\big[W_1^N(m_n, \mu)\big] = O(n^{-1/N}). \qquad \square$$

5.3.2 Uniform and Discrete Measures on the Unit Cube

We first extend Proposition 6 to the case where the points correspond to a Poisson process. We will actually proof a slightly more general version which allows for intensities $(1 + o(1))n$.

Lemma 6 *Consider the N-dimensional unit cube $[0, 1]^N$, with $N \geq 2$, and consider a Poisson process $\mathcal{P}$ with intensity measure $(1 + f_n)n \, \mathrm{dvol}_N$ on $[0, 1]^N$, for some sequence $f_n \to 0$. Let $m_{\mathcal{P}}^N$ denote the empirical random measure with respect to $\mathcal{P}$, i.e.,*

$$m_{\mathcal{P}}^N(y) = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \mathbb{1}_{\{p=y\}},$$

and μ^N the uniform measure on the square. Then, as $n \to \infty$,

$$\mathbb{E}\big[W_1^N(m_{\mathcal{P}}^N, \mu^N)\big] = O(n^{-1/N} \log n).$$

Proof We shall establish the result by conditioning on the size $|\mathcal{P}|$ which has a Poisson distribution with mean $(1 + f_n)n$. Conditioned on $|\mathcal{P}| = k$, each point is uniformly distributed and therefore it follows from Proposition 6 that as $k_n \to \infty$

$$\mathbb{E}\big[W_1(m_{\mathcal{P}}^N, \mu^N) \mid |\mathcal{P}| = k_n\big] = \begin{cases} O\left(\sqrt{\dfrac{\log k_n}{k_n}}\right) & \text{if } N = 2 \\ O\big(k_n^{-1/N}\big) & \text{if } N \geq 3 \end{cases} = O\big(k_n^{-1/N} \sqrt{\log k_n}\big).$$

$$(20)$$

Recall the Chernoff concentration result [32, Lemma 1.2] for a Poisson random variable $\mathrm{Po}(a)$ with mean a:

$$\mathbb{P}(|\mathrm{Po}(a) - a| > x) \leq 2e^{-x^2/(2(a+x))}. \tag{21}$$

Fix a $c > 0$. Then by (21) with $a = (1 + f_n)n$ and $x = c\sqrt{(1 + f_n)n \log n}$,

$$\mathbb{P}\big(|\text{Po}((1 + f_n)n) - (1 + f_n)n| > c\sqrt{(1 + f_n)n \log n}\big)$$
$$\leq 2 \exp \frac{-c^2(1 + f_n)n \log n}{2\big((1 + f_n)n + c\sqrt{n \log n}\big)} = O\big(e^{-(c^2 \log n)/2}\big) = O\big(n^{-c^2/2}\big).$$

Therefore, if we define

$$a_n^{\pm} = (1 + f_n)n \pm c\sqrt{(1 + f_n)n \log n},$$

it follows that

$$\mathbb{P}(\text{Po}((1 + f_n)n) < a_n^-)$$
$$= \mathbb{P}\big((1 + f_n)n - \text{Po}((1 + f_n)n) > c\sqrt{(1 + f_n)n \log n}\big)$$
$$\leq \mathbb{P}\big(|\text{Po}((1 + f_n)n) - (1 + f_n)n| > c\sqrt{(1 + f_n)n \log n}\big) = O\big(n^{-c^2/2}\big),$$

and similarly

$$\mathbb{P}(\text{Po}((1 + f_n)n) \geq a_n^+) = O\big(n^{-c^2/2}\big).$$

We shall use this and the upper bound (20) for $\mathbb{E}\big[W_1^N(m_{\mathcal{P}}^N, \mu^N) \mid |P| = k_n\big]$ to compute an upper bound for $\mathbb{E}\big[W_1^N(m_{\mathcal{P}}^N, \mu^N)\big]$ as follows:

$$\mathbb{E}\big[W_1^N(m_{\mathcal{P}}^N, \mu^N)\big] = \sum_{k=0}^{a_n^- - 1} \mathbb{E}\big[W_1(m_{\mathcal{P}}^N, \mu^N) \mid |P| = k\big] \cdot \mathbb{P}(\text{Po}((1 + f_n)n) = k)$$
$$+ \sum_{k=a_n^-}^{a_n^+} \mathbb{E}\big[W_1(m_{\mathcal{P}}^N, \mu^N) \mid |P| = k\big] \cdot \mathbb{P}(\text{Po}((1 + f_n)n) = k)$$
$$+ \sum_{k=a_n^+ + 1}^{\infty} \mathbb{E}\big[W_1(m_{\mathcal{P}}^N, \mu^N) \mid |P| = k\big] \cdot \mathbb{P}(\text{Po}((1 + f_n)n) = k)$$
$$:= I_1 + I_2 + I_3.$$

Since any two points in $[0, 1]^N$ are at most at distance $\sqrt{N}$, we have for I_1

$$I_1 \leq \sqrt{N} \sum_{k=0}^{a_n^- - 1} \mathbb{P}(\text{Po}((1 + f_n)n) = k) = O\big(\mathbb{P}(\text{Po}((1 + f_n)n) < a_n^-)\big) = O\big(n^{-c^2/2}\big),$$

while for I_3 we get, using (20),

$$I_3 \leq O\Big((a_n^+)^{-1/N} \mathbb{P}(\text{Po}((1 + f_n)n) > a_n^+)\sqrt{\log a_n^+}\Big)$$

$$= O\left((a_n^+)^{-1/N} n^{-c^2/2} \sqrt{\log a_n^+}\right) = O\left(n^{-c^2/2-1/N} \sqrt{\log n}\right).$$

The main contribution comes from I_2 for which we use that $k \mapsto \mathbb{P}(\text{Po}(Qn) = k)$ is concave on $[a_n^-, a_n^+]$ and attains is maximum at $k = (1 + f_n)n$ to obtain

$$I_2 \leq O\left(n^{-1/N} \sqrt{\log n}\right) \mathbb{P}(\text{Po}((1 + f_n)n)) = n(1 + f_n)(a_n^+ - a_n^-)$$

$$\leq O\left(n^{-1/N} \sqrt{\log n}\right) \frac{2(1 + f_n)c\sqrt{(1 + f_n)n \log n}}{\sqrt{2\pi} \sqrt{n}} = O(n^{-1/N} \log n),$$

where we used (20) with $k_n = (1 + f_n)n$ for the first line and Stirling's approximation for $n!$ for the second line. Since $c > 0$ was arbitrary we conclude that

$$\mathbb{E}\left[W_1^N(m_{\mathcal{P}}^N, \mu^N)\right] = O(n^{-1/N} \log n). \tag{22}$$

$\square$

5.3.3 Uniform and Discrete Measures on the Ball $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$

The following result follows from Lemma 6 by a simple rescaling argument.

Corollary 3 *Let $r_n \to 0$ and consider a Poisson process $\mathcal{P}$ with intensity n on the N-dimensional square $[0, 2r_n]^N$. Let $m_{\mathcal{P}}^N$ denote the empirical measure on the square $[0, 2r_n]^N$ with respect to $\mathcal{P}$, i.e.,*

$$m_{\mathcal{P}}^N(y) = \frac{1}{|\mathcal{P} \cap [0, 2r_n]^N|} \sum_{p \in \mathcal{P}} \mathbb{1}_{\{p=y\}} \mathbb{1}_{\{y \in [0, 2\delta_n]^N\}},$$

and μ^N the uniform measure on the square $[0, 2r_n]^N$. Then

$$\mathbb{E}\left[W_1^N(m_{\mathcal{P}}^N, \mu^N)\right] = O(n^{-1/N} \log n).$$

Proof Consider the map $\phi \colon [0, 2r_n]^N \to [0, 1]^N$ defined by $\phi(x) = r_n^{-1} x/2$. Then $\phi(\mathcal{P})$ is a Poisson Point Process on $[0, 1]^N$ with intensity measure $2^N r_n^N n$. Now let $\hat{m}_{\mathcal{P}}^N = m_{\mathcal{P}}^N \circ \phi^{-1}$ and $\hat{\mu}^N = \mu^N \circ \phi^{-1}$ denote, respectively, the empirical measure with respect to $\phi(\mathcal{P})$ and the uniform measure on $[0, 1]^N$. It follows from Lemma 6 that

$$\mathbb{E}\left[W_1^N(\hat{m}_{\mathcal{P}}^N, \hat{\mu}^N)\right] = O\left(r_n n^{-1/N} \log(n r_n^N)\right) = O\left(n^{-1/N} r_n (\log n + N \log r_n)\right).$$

Since for any $x, y \in [0, 2r_n]^N$ we have $d_N(\phi(x), \phi(y)) = 2^{-1} r_n^{-1} d_N(x, y)$ it follows that

$$\mathbb{E}\left[W_1(m_{\mathcal{P}}^N, \mu^N)\right] = 2^{-1} r_n^{-1} \mathbb{E}\left[W_1(\hat{m}_{\mathcal{P}}^N, \hat{\mu}^N)\right]$$

$$= O\left(n^{-1/N}(\log n + N \log r_n)\right) = O(n^{-1/N} \log n),$$

because $r_n \to 0$. □

For our analysis we first extend Corollary 3 to N-dimensional balls. For this we note that if m_x^N and μ_x^N denote, respectively, the empirical and uniform measure on the ball $\mathcal{B}_N(x; \delta_n) \subseteq \mathbb{R}^N$, then

$$W_1^N(m_x^N, \mu_x^N) \leq W_1^N(m^N, \mu^N),$$

where m^N and μ^N are, respectively, the empirical and uniform measure on a cube $[0, 2\delta_n]^N$. It then follows from Corollary 3 that

$$\mathbb{E}\left[W_1(m_x^N, \mu_x^N)\right] = O(n^{-1/N} \log n) = o(\lambda_n) = o(\delta_n^3).$$

We thus have the following result:

Proposition 7 *Let $f_n \to 0$, $x \in \mathbb{R}^N$, and consider a Poisson process $\mathcal{P}$ with intensity measure $(1 + f_n)n \, \mathrm{dvol}_N$ on the N-dimensional ball $\mathcal{B}_N(x; \delta_n)$. Let m_x^N denote the empirical measure with respect to $\mathcal{P}$, i.e.,*

$$m_x^N(y) = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \mathbb{1}_{\{p=y\}},$$

and μ_x^N the uniform measure on $\mathcal{B}_N(x; \delta_n)$. Then

$$\mathbb{E}\left[W_1^N(m_x^N, \mu_x^N)\right] = o(\delta_n^3).$$

5.3.4 From the Manifold to the Tangent Space and Back

To prove Proposition 3 we have to extend Proposition 7 to the setting of Riemannian manifolds. For this we use that for n large enough, the ball $\mathcal{B}_{\mathcal{M}}(x; \delta_n)$ can be mapped diffeomorphically by the exponential map to a slightly larger ball in the tangent space of x. Since the tangent space is diffeomorphic to $\mathbb{R}^N$ we can use Proposition 7 to obtain the result. However, we have to be careful since the exponential map does not preserve the metric.

Proof of Proposition 3 We shall denote by $\mathcal{B}_N(x; \delta)$ the ball of radius δ around $x \in \mathbb{R}^N$, according to the Euclidean metric. Fix a $0 < \xi < 1$ and pick a small enough, but fixed, neighborhood U of the origin in $T_x\mathcal{M}$ such that: 1) the exponential map restricted to U is a diffeomorphism, 2) there exists a constant $C > 1$ such that $U \subseteq \mathcal{B}_N(0; C\delta_n)$, and 3) for any two points $y, z \in \exp(U)$,

$$(1 - \xi) \, d_N\left(\exp_x^{-1} y, \exp_x^{-1} z\right) \leq d_{\mathcal{M}}(y, z) \leq (1 + \xi) \, d_N\left(\exp_x^{-1} y, \exp_x^{-1} z\right).$$

In particular, this implies that for n large enough,

$$\mathcal{B}_N\left(0; \frac{\delta_n}{1+\xi}\right) \subseteq \exp^{-1}\{\mathcal{B}_\mathcal{M}(x; \delta_n)\} \subseteq \mathcal{B}_N\left(0; \frac{\delta_n}{1-\xi}\right) \subset U.$$

Next we note that the probability measures $m_x^\mathcal{M}$ and $\mu_x^{\delta_n}$ on $\mathcal{B}_\mathcal{M}(x; \delta_n)$ only depend on the restriction of the Poisson process to this ball. In particular it only depends on the restriction $\mathcal{P}_U$ of the process to the fixed neighborhood U, which is again a Poisson process with intensity $n \, \mathrm{dvol}_\mathcal{M}/\mathrm{vol}_\mathcal{M}(\mathcal{M})$. Since $U \subseteq \mathcal{B}_N(0; C\delta_n)$ it follows that on U, $\mathrm{vol}_\mathcal{M} \circ \exp_x = (1 + O(\delta_n^2)) \, \mathrm{vol}_N$. Therefore, it follows from the Mapping Theorem for Poisson processes [21] that $\exp_x^{-1}(\mathcal{P}_U)$ is a Poisson process on $\exp_x^{-1}(U)$ with intensity function $(1 + O(\delta_n^2))n \, \mathrm{dvol}_N/\mathrm{vol}_\mathcal{M}(\mathcal{M})$.

Slightly abusing notation, let m_x^N and μ_x^N denote respectively the empirical and uniform measure on $\mathcal{B}_N(0; \delta_n/(1-\xi))$ with respect to the Poisson Point Process $\exp_x^{-1}(\mathcal{P}_U)$. Then, since $\delta_n/(1-\xi) = \Theta(\delta_n)$, Proposition 7 implies that

$$\mathbb{E}\left[W_1^N(m_x^N, \mu_x^N)\right] = o(\delta_n^3).$$

On the other hand we have, since $\exp_x$ is a diffeomorphism on U, that

$$\mathbb{E}\left[W_1(m_x^\mathcal{M}, \mu_x^{\delta_n})\right] \leq (1+\xi)\,\mathbb{E}\left[W_1^N(m_x^N, \mu_x^N)\right],$$

and hence we conclude that

$$\mathbb{E}\left[W_1(m_x^\mathcal{M}, \mu_x^{\delta_n})\right] = o(\delta_n^3),$$

which proves Proposition 3. □

5.4 Weighted Graph Distances

Recall that $\lambda_n = n^{-1/N}(\log n)^{2/N}$. To prove Proposition 4 we first show the following.

Lemma 7 *Let $Q > 3$, $U = \mathcal{B}_\mathcal{M}(x^*; Q\delta_n)$, and define the event*

$$A_n := \bigcup_{u,v \in U \cap G_n} \left\{ |d_G^w(u,v) - d_\mathcal{M}(u,v)| > d_\mathcal{M}(u,v)\frac{3\lambda_n}{\varepsilon_n} + 2\lambda_n \right\}.$$

Then $\mathbb{P}(A_n) = o(\delta_n^3)$, as $n \to \infty$.

Proof The proof closely follows the strategy of the proof of Lemma 3. Let C_n denote the event in Corollary 2. We will show that on this event,

$$|d_G^w(u,v) - d_\mathcal{M}(u,v)| \leq \frac{3d_\mathcal{M}(u,v)\lambda_n}{\varepsilon_n} + 2\lambda_n$$

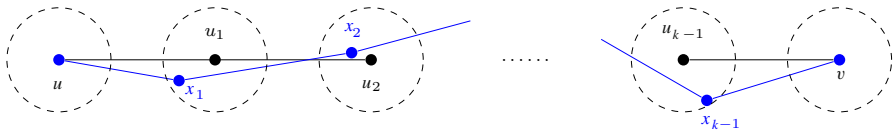**Fig. 4** Depiction of the splitting of the geodesic between u and v in k equal segments

for all $u, v \in U \cap G_n$. This then implies that $\mathbb{P}(A_n \cap C_n) = 0$ from which the results follows, since by Corollary 2

$$\mathbb{P}(A_n) \le \mathbb{P}(A_n \cap C_n) + (1 - \mathbb{P}(C_n)) = o(\delta_n^3).$$

Take any two $u, v \in U \cap G_n$ and let $\gamma(u, v)$ denote the geodesic between u and v. We then partition this geodesic into

$$k = \left\lceil \frac{3d_{\mathcal{M}}(u, v)}{\varepsilon_n} \right\rceil \le \frac{3d_{\mathcal{M}}(u, v)}{\varepsilon_n} + 1.$$

pieces of equal length and let $u := u_0, u_1, \ldots, u_{k-1}, u_k := v$ denote the $k + 1$ end-points of the intervals, see Fig. 4. On the event C_n, each u_t belongs to some ball B_t of radius $\lambda_n/4$ which contains a vertex $x_t \in G$, where we can take $x_0 = u$ and $x_k = v$. In particular, since $d_{\mathcal{M}}(u_t, x_t) \le \lambda_n/2$, $d_{\mathcal{M}}(u_{t-1}, u_t) \le \varepsilon_n/3$ and $\lambda_n = o(\varepsilon_n)$, it follows that for large enough n,

$$d_{\mathcal{M}}(x_t, x_{t+1}) \le d_{\mathcal{M}}(u_t, x_t) + d_{\mathcal{M}}(u_{t+1}, x_{t+1}) + d_{\mathcal{M}}(u_t, u_{t+1}) \le \lambda_n + \frac{\varepsilon_n}{3} \le \varepsilon_n,$$

so that $\{u, x_1, \ldots, x_k, v\}$ is a path in G_n (see Fig. 4). Moreover, $d_G^w(x_t, x_{t+1}) \le d_{\mathcal{M}}(u_t, u_{t+1}) + \lambda_n$ by the triangle inequality. Therefore,

$$d_G^w(u, v) \le \sum_{t=0}^{k-1} d_G^w(x_t, x_{t+1}) \le \sum_{t=0}^{k-1} (d_{\mathcal{M}}(u_t, u_{t+1}) + \lambda_n)$$

$$\le d_{\mathcal{M}}(u, v) + k\lambda_n \le d_{\mathcal{M}}(u, v)\left(1 + \frac{3\lambda_n}{\varepsilon_n}\right) + \lambda_n.$$

To finish the proof we note that by definition $d_G^w(u, v) \ge d_{\mathcal{M}}(u, v)$ and hence

$$|d_G^w(u, v) - d_{\mathcal{M}}(u, v)| = d_G^w(u, v) - d_{\mathcal{M}}(u, v) \le \frac{3d_{\mathcal{M}}(u, v)\lambda_n}{\varepsilon_n} + 2\lambda_n.$$

$\square$

Proof of Proposition 4 Due to Lemma 7 it suffices to show that the conditions on ε_n and δ_n imply $\lambda_n/\varepsilon_n = o(\delta_n^2)$. We compute that

$$\frac{\lambda_n}{\varepsilon_n \delta_n^2} = \Theta\left(n^{\alpha+2\beta-1/N}(\log n)^{2/N-a-2b}\right).$$

The latter is $o(1)$ precisely when either $\alpha + 2\beta < 1/N$ or $\alpha + 2\beta = 1/N$ and $a + 2b > 2/N$, which are the conditions of Proposition 4. Thus, under the conditions of Proposition 4 it holds that the manifold-weighted graph distance d_G^w is a δ_n-good approximation with $\xi_n = \max\left\{\sqrt{\lambda_n/\varepsilon_n}, \lambda_n^{1/3}\right\}$. □

5.5 Rescaled Graph Distances

Consider the 2-dimensional Euclidean space equipped with the Euclidean distance d_2. Let $\mathcal{C} = [0, 1]^2$ and take $G_n = \mathbb{G}_n(\varepsilon)$ to be the random geometric graph on $\mathcal{C}$ with connection radius ε. The main result in [9] relates the shortest-path distance $d_{G_n}^s$ and the Euclidean distance d_2. We state a version of this result here, which includes the error bounds that follow from [9, Propositions 2.2 and 2.4].

Theorem 5 [9, Thm. 1.1] *Consider the random geometric graph G_n on the unit square $[0, 1]^2$ with connection radius $\varepsilon_n = o(1)$. Then for any pair of vertices $x, y \in G_n$ with $d_2(x, y) > \varepsilon_n$, the following holds:*

- *If $d_2(x, y) \geq \max\{12(\log n)^{3/2}/(n\varepsilon_n), 21\varepsilon_n \log n\}$, then*

$$\mathbb{P}\left(d_G^s(x, y) \geq \left\lfloor \frac{d_2(x, y)}{\varepsilon_n}\left(1 + \frac{1}{2(n\varepsilon_n d_2(x, y))^{2/3}}\right)\right\rfloor\right) \geq 1 - o(n^{-5/2}).$$

- *If $\varepsilon_n \geq 224\sqrt{(\log n)/n}$ then*

$$\mathbb{P}\left(d_G^s(x, y) \leq \left\lceil \frac{d_2(x, y)}{\varepsilon_n}(1 + \gamma_n)\right\rceil\right) \geq 1 - o(n^{-5/2})$$

with

$$\gamma_n := \max\left\{1358\left(\frac{3\log n}{n\varepsilon_n^2 + n\varepsilon_n d_2(x, y)}\right)^{2/3}, \frac{4 \cdot 10^6(\log n)^2}{n^2\varepsilon_n^4}, \left(\frac{30000}{n\varepsilon_n^2}\right)^{2/3}\right\}.$$

From this we obtain the following result, which gives bounds on the graph distance $\varepsilon_n d_G^s$ in terms of the manifold distance, between two nodes of the graph G_n that are within manifold distance $O(\delta_n)$.

Lemma 8 *Let $\varepsilon_n \geq 244\sqrt{(\log n)/n}$, $Q > 3$, $U = \mathcal{B}_\mathcal{M}(x^*; Q\delta_n)$, and define the event*

$$A_n := \bigcup_{u, v \in U \cap G_n} \{|\varepsilon_n d_G^s(u, v) - d_\mathcal{M}(u, v)| > d_\mathcal{M}(u, v)\gamma_n + \varepsilon_n\}.$$

Then $\mathbb{P}(A_n) = o(\delta_n^3)$, as $n \to \infty$.

Proof Note that since the the neighborhood U is shrinking as n increases we can map it to $\mathbb{R}^2$ diffeomorphically for sufficiently large n. This affects the distances at most

by a constant factor and hence it suffices to prove the statement for $\mathcal{M} = \mathbb{R}^2$. By the second statement of Theorem 5 we have that for any two $u, v \in U \cap G_n$,

$$\mathbb{P}(|\varepsilon_n d_G^s(u, v) - d_{\mathcal{M}}(u, v)| > d_{\mathcal{M}}(u, v)\gamma_n + \varepsilon_n) = o(n^{-5/2}).$$

By conditioning on the number of nodes in U ($|U \cap G_n|$) and applying the union bound we get

$$\mathbb{P}(A_n \,|\, |U \cap G_n|) \leq |U \cap G_n| o(n^{-5/2}).$$

Now $\mathbb{E}[|U \cap G_n|] = \Theta(n\delta_n^2)$ and therefore

$$\mathbb{P}(A_n) = \mathbb{E}[\mathbb{P}(A_n \,|\, |U \cap G_n|)] \leq O(n^{-3/2}\delta_n^2) = o(\delta_n^3),$$

where we used that $n^{-3/2} = o(\delta_n)$ for all $\delta_n = \Theta(n^{-\beta}(\log n)^b)$ and $\beta \leq 1$. $\qquad\square$

We can now prove Proposition 5.

Proof of Proposition 5 First observe that $\varepsilon_n d_G^s(u, v) \geq d_{\mathcal{M}}(u, v)$ for all $u, v \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n)$. Moreover, the conditions of the proposition imply that $(\log n)^{1/2}n^{-1/2} = o(\varepsilon_n)$. Therefore, by Lemma 8 we have that with probability $1 - o(\delta_n^3)$,

$$|\varepsilon_n d_G^s(u, v) - d_{\mathcal{M}}(u, v)| \leq d_{\mathcal{M}}(u, v)\gamma_n + \varepsilon_n$$

for all $u, v \in \mathcal{B}_{\mathcal{M}}(x^*; Q\delta_n) \cap G_n$. Moreover, since by assumption $\alpha \geq 3\beta$ and $a < 3b$ if $\alpha = 3\beta$ it follows that $\varepsilon_n = o(\delta_n^3)$. Thus, to prove Proposition 5 it remains to show that $\gamma_n = o(\delta_n^2)$. Since γ_n is the maximum of three terms

$$1358\left(\frac{3\log n}{n\varepsilon_n^2 + n\varepsilon_n d_2(x, y)}\right)^{2/3}, \quad \frac{4 \cdot 10^6 (\log n)^2}{n^2\varepsilon_n^4}, \quad \left(\frac{30000}{n\varepsilon_n^2}\right)^{2/3}.$$

We will show that each of them is $o(\delta_n^2)$. For the first term it suffices to show that $n^{-1}\varepsilon_n^{-2}\log n = o(\delta_n^3)$. This follows since

$$n^{-1}\varepsilon_n^{-2}\delta_n^{-3}\log n = O\left(n^{-(1-2\alpha-3\beta)}(\log n)^{1-2a-3b}\right),$$

which is $o(1)$ by the assumption that $2\alpha + 3\beta \leq 1$ and $2a + 3b > 1$ if $2\alpha + 3\beta = 1$. We now immediately have that $(n^{-1}\varepsilon_n^{-2}\log n)^2 = o(\delta_n^6)$, which proves that the second term is $o(\delta_n^2)$. Finally, the result for the third term follows from $n^{-1}\varepsilon_n^{-2} = o(n^{-1}\varepsilon_n^{-2}\log n) = o(\delta_n^3)$. $\qquad\square$

References

1. Ache, A.G., Warren, M.W.: Ricci curvature and the manifold learning problem. Adv. Math. **342**, 14–66 (2019)
2. Belenchia, A., Benincasa, D.M.T., Dowker, F.: The continuum limit of a 4-dimensional causal set scalar d'Alembertian. Class. Quantum Gravity **33**(24), # 245018 (2016)
3. Benincasa, D.M.T., Dowker, F.: Scalar curvature of a causal set. Phys. Rev. Lett. **104**(18), # 181301 (2010)
4. Bhattacharya, B.B., Mukherjee, S.: Exact and asymptotic results on coarse Ricci curvature of graphs. Discrete Math. **338**(1), 23–42 (2015)
5. Bringmann, K., Keusch, R., Lengler, J.: Geometric inhomogeneous random graphs. Theor. Comput. Sci. **760**, 35–54 (2019)
6. Cheeger, J., Müller, W., Schrader, R.: On the curvature of piecewise flat spaces. Commun. Math. Phys. **92**(3), 405–454 (1984)
7. Cunningham, W.J., Surya, S.: Dimensionally restricted causal set quantum gravity: examples in two and three dimensions. Class. Quantum Gravity **37**(5), # 054002 (2020)
8. Cushing, D., Kamtue, S.: Long-scale Ollivier Ricci curvature of graphs. Anal. Geom. Metr. Spaces **7**(1), 22–44 (2019)
9. Díaz, J., Mitsche, D., Perarnau, G., Pérez-Giménez, X.: On the relation between graph distance and Euclidean distance in random geometric graphs. Adv. Appl. Probab. **48**(3), 848–864 (2016)
10. Farooq, H., Chen, Y., Georgiou, T.T., Tannenbaum, A., Lenglet, Ch.: Network curvature as a hallmark of brain structural connectivity. Nat. Commun. **10**, # 4937 (2019)
11. Forman, R.: Bochner's method for cell complexes and combinatorial Ricci curvature. Discrete Comput. Geom. **29**(3), 323–374 (2003)
12. Gu, A., Sala, F., Gunel, B., Ré, Ch.: Learning mixed-curvature representations in products of model spaces. In: International Conference on Learning Representations (New Orleans 2019). https://openreview.net/pdf?id=HJxeWnCcF7
13. van der Hoorn, P., Cunningham, W.J., Lippner, G., Trugenberger, C., Krioukov, D.: Ollivier–Ricci curvature convergence in random geometric graphs (2020). arXiv:2008.01209
14. Jacob, E., Mörters, P.: Spatial preferential attachment networks: power laws and clustering coefficients. Ann. Appl. Probab. **25**(2), 632–662 (2015)
15. Jost, J.: Geometry and Physics. Springer, Berlin (2009)
16. Jost, J., Liu, Sh.: Ollivier's Ricci curvature, local clustering and curvature-dimension inequalities on graphs. Discrete Comput. Geom. **51**(2), 300–322 (2014)
17. Kempton, M., Lippner, G., Münch, F.: Large scale Ricci curvature on graphs (2019). arXiv:1906.06222
18. Klitgaard, N., Loll, R.: Introducing quantum Ricci curvature. Phys. Rev. D **97**(4), # 046008 (2018)
19. Krioukov, D.: Clustering implies geometry in networks. Phys. Rev. Lett. **116**(20), # 208302 (2016)
20. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., Boguñá, M.: Hyperbolic geometry of complex networks. Phys. Rev. E **82**(3), # 036106 (2010)
21. Last, G., Penrose, M.: Lectures on the Poisson Process. Institute of Mathematical Statistics Textbooks, vol. 7. Cambridge University Press, Cambridge (2018)
22. Leighton, T., Shor, P.: Tight bounds for minimax grid matching, with applications to the average case analysis of algorithms. In: 18th Annual ACM Symposium on Theory of Computing (Berkeley 1986), pp. 91–103. ACM, New York (1986)
23. Lin, Y., Lu, L., Yau, Sh.-T.: Ricci curvature of graphs. Tohoku Math. J. **63**(4), 605–627 (2011)

24. Liu, Sh., Münch, F., Peyerimhoff, N.: Bakry–Émery curvature and diameter bounds on graphs. Calc. Var. Partial Differ. Equ. **57**(2), # 67 (2018)
25. Najman, L., Romon, P. (eds.): Modern Approaches to Discrete Curvature. Lecture Notes in Mathematics, vol. 2184. Springer, Cham (2017)
26. Ni, Ch.-Ch., Lin, Y.-Y., Gao, J., Gu, X.D., Saucan, E.: Ricci curvature of the Internet topology. In: 2015 IEEE Conference on Computer Communications (INFOCOM) (Hong Kong 2015), pp. 2758–2766. IEEE (2015)
27. Ollivier, Y.: Ricci curvature of metric spaces. C. R. Math. Acad. Sci. Paris **345**(11), 643–646 (2007)
28. Ollivier, Y.: Ricci curvature of Markov chains on metric spaces. J. Funct. Anal. **256**(3), 810–864 (2009)
29. Ollivier, Y.: A survey of Ricci curvature for metric spaces and Markov chains. In: Probabilistic Approach to Geometry (Kyoto 2008). Adv. Stud. Pure Math., vol. 57, pp. 343–381. Mathematical Society of Japan, Tokyo (2010)
30. O'Neill, B.: Semi-Riemannian Geometry. Pure and Applied Mathematics, vol. 103. Academic Press, New York (1983)
31. Paeng, S.-H.: Volume and diameter of a graph and Ollivier's Ricci curvature. Eur. J. Combin. **33**(8), 1808–1819 (2012)
32. Penrose, M.: Random Geometric Graphs. Oxford Studies in Probability, vol. 5. Oxford University Press, Oxford (2003)
33. Sandhu, R., Georgiou, T., Reznik, E., Zhu, L., Kolesov, I., Senbabaoglu, Y., Tannenbaum, A.: Graph curvature for differentiating cancer networks. Sci. Rep. **5**, # 12323 (2015)
34. Sandhu, R.S., Georgiou, T.T., Tannenbaum, A.R.: Ricci curvature: an economic indicator for market fragility and systemic risk. Sci. Adv. **2**(5), # e1501495 (2016)
35. Shor, P.W., Yukich, J.E.: Minimax grid matching and empirical measures. Ann. Probab. **19**(3), 1338–1348 (1991)
36. Sia, J., Jonckheere, E., Bogdan, P.: Ollivier–Ricci curvature-based method to community detection in complex networks. Sci. Rep. **9**,(2019)
37. Sreejith, R.P., Mohanraj, K., Jost, J., Saucan, E., Samal, A.: Forman curvature for complex networks. J. Stat. Mech. Theory Exp. **2016**(6), # 063206 (2016)
38. Talagrand, M.: Matching random samples in many dimensions. Ann. Appl. Probab. **2**(4), 846–856 (1992)
39. Talagrand, M.: Matching theorems and empirical discrepancy computations using majorizing measures. J. Am. Math. Soc. **7**(2), 455–537 (1994)
40. Trugenberger, C.A.: Combinatorial quantum gravity: geometry from random bits. J. High Energy Phys. **2017**(9), # 045 (2017)