# Diffusion Curvature for Fast, Point-wise, Noise-Resistant Geometric Featurization of Graphs and Pointclouds

Kincaid MacDonald Dhananjay Bhaskar Kaly Zhang
Ian Adelstein Smita Krishnaswamy

2024-05-10

For a number of years now work has been proceeding in order to bring to perfection the crudely conceived idea of a machine that would not only supply inverse reactive current for use in unilateral phase detractors, but would also be capable of automatically synchronizing cardinal grammeters. Such a machine is the "Turbo-Encabulator."

## 0.1 Introduction

One of the most ubiquitous subjects of analysis in data science is the humble point cloud. The points, by themselves, are high dimensional and noisy; it is up to the data scientist to wring sense out of them. Per the Manifold Hypothesis, we assume the points were sampled on or near the surface of a low-dimensional manifold embedded in high-dimensional Euclidean space. Manifold learning methods, like t-SNE, PHATE, and Diffusion Maps [(**maaten2008VisualizingDataUsing?**)](**moon2019VisualizingStructureTransitions?**)(**coifman2006D** endeavor to recover salient features of the underlying manifold, like geodesic distances, population clusterings, and dimension, from its high-dimensional noisy sampling.

Curvature is a particularly troublesome geometric property to translate into the discrete, sampled realm. In smooth Riemannian manifolds, curvature is a *local* phenomenon. It can be obtained by fitting osculating circles of radius limiting to zero, or computed from the manifold's Hessian, using the Second Fundamental Form. None of these translate into the discrete realm. In a sampled manifold, taking a local limit is impossible – one can't "zoom in" past the sampling of points – and we don't have access to the parametrization of the manifold or its tangent bundle, without computationally costly and potentially error-prone estimation. Moreover, as our sampling is likely noisy, the curvature can only be recovered over a sufficiently large neighborhood of points to counter the spurious geometric artifacts created noisy sampling. Thus, in the discrete realm, curvature becomes a “semi-local” phenomenon, in which neither the smallest nor larger scales can be trusted.

There are elegant generalizations of classical curvature to discrete spaces that overcome many of these roadblocks. Ollivier’s *Coarse Ricci Curvature* (CRC) employs optimal transport theory to relate the behavior of a discrete neighborhood to its smooth counterpart (**ollivier2009RicciCurvatureMarkov?**). Sturm’s *displacement convexity of entropy* (DCE) measures the proliferation of midpoints in positive curvature (**sturm2006GeometryMetricMeasure?**). Both methods use optimal transport as the basis of their “semi-local” measurement. Rather than trying to zoom in on a point, they define curvature between pairs of points, approximating, at a coarse scale, a Ricci tangent vector.

Although these techniques are theoretically elegant, general, and applicable to any metric measure space, the setting of noisily sampled point clouds is practically challenging for CRC and DCE. Both methods rely on the graph’s shortest-path lengths as an approximation of the manifold’s ground distance - a perilous assumption when dealing with noisy data. And for large datasets, optimal transport calculations can be computationally prohibitive.

In this paper, we develop Diffusion Curvature, a fast curvature estimate derived solely from the graph diffusion matrix. We first introduced the ideas behind Diffusion Curvature in (Bhaskar et al. 2022), in which we demonstrated its ability to produce an unsigned magnitude of curvature estimation for toy datasets and single-cell data, and proved a correspondence between the ratios of scalar curvature and diffusion curvature. We now present a refined definition which produces *signed* curvature values and prove bounds relating Diffusion Curvature to Ollivier’s coarse Ricci Curvature. We demonstrate Diffusion Curvature’s robustness to noise and sampling artifacts, and position our technique as an adaptation of coarse Ricci curvature particularly suitable for point cloud data.

## 0.2 Background

### 0.2.1 Curvature in the Continuous Setting

### 0.2.2 The Discrete Setting

Within the ambient setting of points $x_i \in \mathbb{R}^D$, the Euclidean distances between the points in our point cloud are not very useful. To perform geometric analysis, we want the manifold’s *geodesic* distances between $x_i, x_{j \in \mathcal{M}}$,. However, manifolds are locally euclidean, so within a sufficiently small neighborhood of $x_i \in \mathcal{M}$ , the euclidean distances are accurate. This is the basis of graph construction: retain only the trustworthy local distances, discard the rest, and then “integrate” over the local neighborhoods to recover features of the global geometry.

A graph $G = (V, E)$ is a collection of $n$ vertices $v_i \in V$ connected by (possibly weighted) edges $e_{ij} \in E$ . It is efficiently represented by a single *adjacency* (or *affinity*) matrix $A \in \mathbb{R}^{n \times n}$, where $A_{ij}$ expresses the degree of connection between the vertices $v_i$ and $v_j$. In a binary adjacency matrix, $A_{ij} = 1$ iff there is an edge between $v_i$ and $v_j$. In a weighted affinity matrix, $0 < A_{ij} < 1$ with a higher affinity indicating a closer connection between the nodes.

One can construct an affinity matrix from a point cloud with the following algorithm: 1. Compute the matrix $D$ of pairwise euclidean distances between points, so that $D_{ij} = \|x_i - x_j\|_2$. 2. Apply a kernel $\kappa$ to the distances to construct the affinity matrix, where $A_{ij} = \kappa(D_{ij})$. This is typically the gaussian kernel:

$$k(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y}{\sigma^2}\right)$$

There are a variety of heuristics for selecting an appropriate kernel bandwidth $\sigma$. In this paper, we use an adaptive kernel bandwidth, in which, when computing $k(D_{ij})$, $\sigma$ is set to the mean distance from the points $x_i$ and $x_j$ to their $k$-th nearest neighbor.

After building our graph affinity matrix $A$, we created a new representation of the point cloud $X$ – turning it from an $n \times D$ matrix of unwieldy ambient coordinates into an $n \times n$ matrix of pairwise connections between points. The challenge is now to reassemble this information of local connectivity to recover the features of $\mathcal{M}$. Graph diffusion does precisely this.

### 0.2.3 Graph Diffusion

The graph diffusion matrix $P$ is a commonly-used method of "integrating" the local connectivity of the graph $A$ into global geometric descriptors of $\mathcal{M}$. Coifman and Lafon (WAWA YEAR) proved a correspondence between iterated graph diffusion $P^t$ and the Neumann heat kernel on $\mathcal{M}$. Their technique, *Diffusion Maps*, uses the euclidean distances between eigencoordinates of $P$ to approximate the geodesic distances on $\mathcal{M}$. The visualization technique *PHATE* (**moon2019VisualizingStructureTransitions?**) constructs a low-dimensional embedding of a point cloud $X$ such that a distance between the transition probabilities $P$ of $X$ is preserved in the embedding. (More on properties of phate, trajectory preservation.) *Diffusion Earth Mover's Distance* (**tongDiffusionEarthMover2021?**) efficiently approximates the transportation distance between distributions on a graph using multi-scale wavelet transform obtained by applying different scales of diffusion. *LEGSNet*'s "learnable geometric scattering" computes tunable scales of diffusion with a graph neural network and achieves state of the art performance on biochemistry graph classification (**tong2020DataDrivenLearningGeometric?**). These are but a few of the many manifold learning techniques based in diffusion.

Constructing the diffusion matrix from the affinity matrix $A$ is straightforward: you simply row-normalize $A$, with an optional step to normalizing by density.

Here is the algorithm presented in Coifman and Lafon (**coifman2006DiffusionMaps?**):

1. (Optional) Compute an *anisotropic density normalization* on $A$, obtaining the anisotropic adjacency matrix $A_\star$.
2. Construct the degree matrix $D$, whose diagonal entries are the rowsums of $A$, i.e. $D_{ii} = \sum_j A_{ij}.$. The other entries are zeros.
3. Define $P = D^{-1}A$, the graph diffusion matrix.

- □ Clean this up: get anisotropic equation, and clarify the role of the self affinity. When is it removed? When is laziness added?

$P$ has several nice properties. The rows $P[i]$ give the transition probabilities of a single step random walk starting at point $x_i$; each row $P[i]$ can be viewed as a probability distribution centered at $x_i$. This is preserved under powers of the matrix. The rows of $P^t$ still sum to 1, and $P^t[i]$ now gives the probability distribution of a $t$-step random walk starting at $x_i$.

Although $P$ is not symmetric, it is conjugate to a symmetric matrix, via $D^{0.5}PD^{-0.5} = D^{-0.5}AD^{-0.5}$, granting it a full basis of real-valued eigenvectors and eigenvalues. These eigenvectors are shared with the normalized graph Laplacian $L = I - D^{-0.5}AD^{-0.5}$. The eigenvalues of $P$ lie between 0 and 1. Powering the matrix $P^t$ thus corresponds to powering the eigenvalues $\lambda_i^t$ of $P$, via diagonalization

$$P^t = \Psi \Lambda^t \Psi^T$$

This is similar to applying a low-pass filter to the graph. As $t$ increases, the smallest eigenvalues decay fastest under repeated powering, and their corresponding eigenvector vanishes from the eigenbasis – leaving only the largest $\lambda_i$, whose eigenvectors trace global geometric features.

This is a remarkable feature of the diffusion matrix: the ability to "denoise" itself by iterating the random walk over larger time scales. Intuitively, the paths through the data most robustly trafficked by random walkers are those supported by multiple high-probability connections from independent starting points.

## 0.3 Related Work

### 0.3.1 Foreman Ricci Curvature

### 0.3.2 Hickock's Curvature

### 0.3.3 Ollivier-Ricci Curvature

Developed by Yann Ollivier in 2007, *Coarse Ricci Curvature* (or sometimes, "Ollivier Ricci Curvature") is a direct translation of Ricci curvature to discrete metric spaces like graphs (**ollivier2009RicciCurvatureMarkov?**). Several classical properties of Ricci curvature can be extended to the graph setting using Coarse Ricci Curvature. Ollivier has, for instance, proven versions of concentration inequalities, Bonnet Myers (more). Coarse Ricci Curvature has, in this way, become something of a bridge between continuous and coarse geometry. The basis of this bridge is optimal transport, and specifically, the 1-Wasserstein distance.

In the Riemannian setting, Ricci curvature captures the phenomenon that, in positive curvature, "small spheres are closer (in transportation distance) than their centers are" (CITE 43 in ORC Paper). On the sphere, for instance, imagine two circles placed at the north and south poles: every point is closer to the corresponding point on the opposite pole than the centers. In negatively curved spaces, the discrepancy reverses, while in a flat space, the average distance between the points of the circles is the distance between the centers.

Coarse Ricci Curvature captures a similar phenomenon on graphs. Instead of spheres, it uses locally-centered probability distributions defined by random walks. And to measure the distance between these walks, it uses the 1-Wasserstein (or Earth Mover's) distance. We'll briefly define each.

The 1-Wasserstein distance is a measure of the distance between probability distributions. Given distributions $\mu_x$ and $\mu_y$ over some shared space $X$, the Wasserstein distance quantifies the smallest amount of "work" needed to transform one distribution into another, by transporting probability "mass" between pairs of points over the ground metric $d(x, y)$:

> [!Info] Definition The 1-Wasserstein distance between distributions $\mu_x$ and $\mu_y$ is
>
> $$W_1(\mu_x, \mu_y) := \inf_{\xi \in \Pi(\mu_x, \mu_u)} \int \int d(x, y)\, d\xi(x, y)$$
>
> where the "transportation plan" $\xi$ is drawn from the space $\Pi(\mu_x, \mu_y)$ of joint probability distributions over $X \times X$ which project onto $\mu_x$ and $\mu_y$.
>
> In the discrete setting, this translates naturally into an infimum over a summation.

$$W_1(\mu_x, \mu_y) := \inf_{\xi \in \Pi(\mu_x, \mu_y)} \sum_{x \in X} \sum_{y \in X} d(x, y) \xi(x, y)$$

What is the analog on a graph of a "small sphere" around a point? Ollivier replaces spheres with a family of measures $m_x(\cdot)$ defined for each point $x$, where 1. Each $\mu_x(\cdot)$ depends measurably on $x$, i.e. the map $x \to \mu_x$ is measurable. 2. Each $\mu_x(\cdot)$ has finite first moment, or *Jump*, i.e. for some $o \in X$ $\int d(o, y)\mu_x(y)\, dx < \infty$.

The *Jump* $J(\mu_x)$ of a measure, a measure of its concentration around a central point, is a concept to which we'll return.

$$J(\mu_x) = \int_{y \in X} d(x, y) \mu_x(y)\, dx$$

In graphs, Ollivier defines these $\mu_x$ as the probability distributions created by a single-step random walk from the point $x$. With a transition probability $\alpha$, and equal probability of moving to each of $x$'s neighbors on the graph, $\mu_x(x) = (1-\alpha)$ and $m_x(y) = \alpha$ if $y \in N(x)$ or 0 otherwise. This is analogous to defining $m_x = Pe_x$, if $P$ is the diffusion matrix created from a binary adjacency matrix. Note, however, that there is nothing limiting us to binary adjacency matrices, or even single steps of diffusion; the two conditions above are equally satisfied by weighted adjacency matrices and $t$-step diffusions, and in sparse or noisy graphs, this may be desirable.

Definition

The *Coarse Ricci Curvature* between $x$ and $y$ is

$$\kappa(x,y) := 1 - \frac{W_1\left(m_x, m_y\right)}{d(x,y)}$$

There are a number of provisos attached to this definition, which tries to approximate a continuous phenomenon within discrete constraints. These constraints, and the relationship between Ricci and Ollivier's coarse Ricci curvature are illustrated Ollivier's Example 2.6 (**ollivier2009RicciCurvatureMarkov?**):

Let $(X,d)$ be a smooth Riemannian manifold of dimension $d$ and let vol be the Riemannian volume measure. Let $\epsilon > 0$ small enough and consider the ball of radius $\epsilon$ around each point $x$. Let $x, y \in X$ be two sufficiently close points. Let $v$ be the unit tangent vector at $x$ directed towards $y$. The coarse Ricci curvature along $v$ is then

$$\kappa(x,y) = \frac{\epsilon^2 \mathrm{Ric}(v,v)}{2(d+2)} + o(\epsilon^3 + \epsilon^2 d(x,y))$$

Hence the coarse Ricci curvature applied to a manifold recovers the Ricci curvature, up to a scaling factor contingent on dimension, and plus an error term that grows with the radius of ball and distance between points.

Ollivier's choice not to scale $\kappa(x,y)$ by dimension is interesting, and likely motivated by his application of coarse Ricci curvature to graph-like spaces for which dimension isn't clearly defined, like social networks. Within our domain of point-cloud data, incorporating dimension may be desirable; without it, spaces of high dimension can be conflated with spaces of low negative curvature but high dimension.

A result on coarse Ricci curvature which will prove useful concerns the *contraction (or expansion) of measure* that occurs under diffusion in spaces of positive (or negative) curvature.

Proposition 20 (**ollivier2009RicciCurvatureMarkov?**)

Let $(X,d,m)$ be a metric space with a random walk. Let $\kappa \in \mathbb{R}$. Then we have $\kappa(x,y) \geq \kappa$ for all $x, y \in X$ iff for any two probability distributions $\mu, \mu' \in \mathcal{P}(X)$ one has

$$W_1(\mu \star m, \mu' \star m) \leq (1-k) W_1(\mu, \mu')$$

Where

$$\mu \star m := \int_{x \in X} d\mu(x) m_x \, dx$$

## 0.4 Methods

The Coarse Ricci Curvature and Displacement Convexity of Entropy are both based on an invariant property of flat spaces, the deviation from which allows for a measure of curvature even without a clear definition of factors which normally interfere with curvature, like dimension. However, the elegance of these invariance-based definitions comes at a computational cost. Both methods require optimal transport calculations for (a subset) of each pair of edges, a complexity growing with $O(n^2)$. Additionally, the transport distance based on the graph shortest-path distance is highly susceptible to noise.

This motivates our formulation of graph curvature, which uses a *comparison* between spaces in lieu of an invariant. This gives *Diffusion Curvature* a number of nice properties:

1. **Point-wise, instead of edge-wise**. In many applications, knowledge of the Ricci curvatures is unnecessarily detailed. Diffusion curvature answers the question: *is the region around this point positively or negatively curved, and in what magnitude, relative to other points on the manifold?*

2. **Diffusion-based denoising**. When edges may be spurious, and random sampling creates deceptive pockets of density, Diffusion Curvature inherits the denoising obtained by powering the diffusion operator.

3. **Extremely fast**. Designed for point clouds with millions of points, the pointwise Diffusion Curvature of the entire dataset can be calculated with a handful of sparse matrix multiplications. In addition, Diffusion Curvature can be *GPU accelerated.* This combination allows computing in minutes what takes existing methods hours.

4. **Differentiable**. Diffusion Curvature natively supports backpropogation, and can easily be incorporated into a loss function for a neural network, e.g. to regularize a latent space to have low (or high) curvature. We provide a PyTorch implementation of the algorithm for this purpose. [LINK TO GITHUB]

The core intuition of Diffusion Curvature is that the *laziness* of random walks on a graph is a proxy for the curvature of the underlying manifold. Picture a "random walker" drunkenly traversing a sphere. If he manages, over the course of several steps, to wander to the opposite pole, he has many ways of getting back to where he started. By contrast, if he begins on the top of a (negatively curved) saddle and wanders down one side, any path aside from exactly retracing his steps incurs a steep distance penalty. On the sphere, the random walker is more likely to find his way home: his walks are "lazier".

Our previous paper (Bhaskar et al. 2022) measured this laziness directly, as the return probability within a k-neighborhood of the starting point. This required tuning the parameter $k$, and neglected the information provided by the random walk probabilities outside of this k-neighborhood. We now refine this definition in two ways.

First, we refine our previous definition of diffusion laziness using the Wasserstein distance between a dirac $\delta_x$ and its $t$ step diffusion $p_X^t(x)$,

$$-W_1\left(\delta_x, p_X^t(x)\right)$$

{eq-laziness} This provides a more sensitive and parameter-free measure of the 'spread' of diffusion.

Second, by comparing the laziness of diffusion between two manifolds – one of known curvature – we make our method signed.

Given samples $X \subseteq M$ and a flattening map $\Phi : X \to \mathbb{R}^d$, The $t$-step *Diffusion Curvature* of $x$ is

$$k_t(x) = W_1\left(\delta_x, p_{\Phi(x)}^t(x)\right) - W_1\left(\delta_x, p_X^t(x)\right)$$

{eq-definition}

- ☐ Use an arbitrary distance? Or a p-Wasserstein?

Where $p_X^t$ is the t-step random walk over $X$, and $p_{\Phi(X)}^t$ is the same over the flattened points $\Phi(X)$. Depending on the ground distance used for the Wasserstein-1 transport distance, and the definition of flattening map employed, one obtains any of a family of diffusion curvatures.

Our best results come a very simple formulation of both, where the ground distance is chosen such that $W_1$ becomes the Shannon Entropy, and our flattening map simply uniformly samples points from the Euclidean space $\mathbb{R}^d$ of the same dimension as $M$. This produces what we'll call the *Entropic Diffusion Curvature*,

$$\kappa e_t(x_i) = H(P^t[i]) - H(P_E^t[i])$$

{eq-entropic-definition} which evokes the use of entropy to measure heat diffusion.

This definition rests on two claims: that the laziness of diffusion reveals the (unsigned) magnitude of curvature, and that by comparing diffusion across manifolds, we can convert this into a signed curvature. The next two sections will analyze this first claim. We provide a motivating example recovering a Bishop-Gromov type volume comparison from our entropic definition. We then analyze diffusion laziness within Ollivier's framework of metric measure theory, and bound the diffusion laziness from above by the Ollivier-Ricci curvature.

### 0.4.1 A Motivating Example: Entropic Diffusion Curvature Recovers a Bishop-Gromov Volume Comparison

Among the most obvious manifestations of curvature is differences in volume: the higher the curvature, the smaller the volume of geodesic balls of the same radius. This is formalized by Bishop-Gromov's Volume Comparison Theorem.

THEOREM!

One route to discrete curvature is thus estimating volumes [cite Hickok]. However, standard methods of estimating volumes in discrete, sampled spaces like point clouds or graphs rely on more complex geometric quantities, like density estimation or approximated geodesic distances, which are not only susceptible to noise, but often themselves dependent on the curvature![1] One wonders: can we use the tools of diffusion geometry to robustly estimate volume?

Indeed, this is one way to interpret diffusion laziness: as an inverse volume measurement. [Cite Huguet Heat Geo] motivates this with a result from Saloff Costes et al. (**saloff-costeHeatKernelIts?**). For manifold satisfying the Parabolic Harnock Inequality (e.g. all manifolds of positive curvature), heat diffusion $m_x^t$ on the manifold is bounded by

$$\frac{c_1}{V(x,\sqrt{t})}\exp\left(-\frac{d(x,y)^2}{c_2 t}\right) \leq m_x^t(y) \leq \frac{c_3}{V(x,\sqrt{t})}\exp\left(-\frac{d(x,y)^2}{c_4 t}\right)$$

In other words, heat diffusion on non-Euclidean manifolds behaves approximately like Euclidean heat diffusion scaled by the local volume.

Powering the diffusion matrix approximates heat diffusion [cite Coifman]. We can thus relate our measure of the 'spread' or 'laziness' of this diffusion to the manifold's volume.

The Shannon entropy of the diffusion $m_x$ over the bounded neighborhood of $X$ with non-zero diffusion mass, can be written

$$H\left(m_x^t\right) = -\int_{y_\epsilon x} m_x^t(y)\ln\left(m_x^t(y)\right)$$

Here we use Shannon's differential entropy.Though this formulation encounters pathologies on unbounded domains, it approximates the entropy a signal on a sampled region as the number of samples goes to infinity.

We'll simplify the notation by combining the integral and distribution into an expected value over $m_x^t$

$$= -\mathbb{E}_{m_x}\ln\left(m_x(y)\right)dy$$

[1] Consider kernel density estimation, where one places gaussians on each data point and sums up the probability assigned to each area. The higher the curvature, the higher the reported density, since at the end of an ellipsoid the gaussians have greater overlap with each other than in its center – even though the intrinsic density of the surface is uniform.

Recalling the above result, there are constants $c_1 \dots c_4$ such that

$$\frac{c_1}{V(x,\sqrt{t})}\exp\left(-\frac{d(x,y)^2}{c_2 t}\right) \le m_x^t(y) \le \frac{c_3}{V(x,\sqrt{t})}\exp\left(-\frac{d(x,y)^2}{c_4 t}\right)$$

hence we can approximate $H(m_x^t)$ by

$$\begin{aligned}
&= -\mathbb{E}_{m_x^t}\ln\left(\frac{1}{V(y,\sqrt{t})}\exp\left(\frac{-d(x,y)^2}{4t}\right)\right)\\
&= -\mathbb{E}_{m_x^t}\ln\left(\frac{1}{V(y,\sqrt{6})}\right) - \frac{d(x,y)^2}{4t}\\
&= \mathbb{E}_{m_x^t}\ln(V(y,\sqrt{t})) + \mathbb{E}_{m_x^t}\frac{d(x,y)^2}{4t}
\end{aligned}$$

By assumption, the volumes $V(y,\sqrt{t})$ equal some constant $V_m(\sqrt{t})$ over the support of $m_x^t$. Also note that the right hand term is precisely $\frac{1}{4t}W_2(\delta_x, m_x^t)$, so

$$H(m_x^t) \simeq \ln(V_m(\sqrt{t})) + \frac{1}{4t}W_2(d_x, m_x^t)$$

The Diffusion Curvature is then

$$\frac{1}{t}(H(m_E^t) - H(m_x^t)) \simeq \frac{1}{t}\ln(V_E(\sqrt{t})) - \frac{1}{t}\ln(V_m(\sqrt{t})) + \frac{1}{4t^2}(W_2(d_x, m_E^t) - W_2(d_x, m_x^t))$$

This presents us with two comparisons: one between the average volume within $N_t(x)$ both in Euclidean space and on the manifold, and one between the Wasserstein spread of diffusion in Euclidean space and on the manifold. Note that both exhibit the same inverse relationship to curvature. If the curvature of $m$ is greater than in euclidean space (e.g. positive), the volume of a ball is smaller – just as the spread of diffusion is diminished. In negative curvature, we have the opposite.

### 0.4.2 Bounding diffusion laziness by Ollivier-Ricci curvature

Although diffusion curvature was designed for sampled manifolds, it extends naturally to the realm of more general graphs. Indeed, diffusion curvature can be seen as an node-wise adaptation of Ollivier Ricci curvature, the theoretically richest and highest performing edge-wise graph curvature [citations to Bastian's paper on ORC expressivity]. Here we review the definitions of Ollivier-Ricci curvature and present theoretical connections between it and diffusion curvature. These motivate the use of diffusion curvature as substitute for Ollivier Ricci curvature on large, or noisy graphs where the latter method stumbles. [Present some of these in experiments]

Ollivier Ricci curvature is based in optimal transport, formalizing the classical intuition that in spaces of positive curvature, “spheres are closer than their centers” [cite ollivier]. In place of spheres, Ollivier uses diffused diracs; for distance, he uses the Wasserstein-1 distance with the shortest-path metric. Given points $x, y$ in the metric space, with 1-step diffusions $m_x$ and $m_y$, the Ollivier Ricci curvature of the edge between $x$ and $y$ is:

$$\kappa(x,y) := 1 - \frac{W_1\left(m_x, m_y\right)}{d(x,y)}$$

Local curvature always involves some comparison. Here, the comparison is between the manifold’s optimal transport distance (the numerator), and the Euclidean optimal transport distance (which, helpfully, is just the distance in the denominator. You’ll notice the similarity to diffusion curvature, which replaces the $W_1$ distance between two diffused distributions with the $W_1$ distance between a point and its $t$-step diffusion. We do this for both the manifold (in the numerator) and a Euclidean comparison space (in the denominator).

$$k_d(x,y) = 1 - \frac{W_1^M(\delta_x, m_x^t)}{W_1^E(\delta_x, m_x^t)}$$

The basis of diffusion curvature’s comparison is the “spread” of diffusion as measured by $W_1(\delta_x, m_x^t)$. Intuitively, in regions of positive curvature, diffusion spreads less; neighborhoods with higher interconnectivity make random walks lazier. We can formalize this with reference to Ollivier Ricci curvature:

**Proposition 1.** Let $(X, d, m)$ be a metric space equipped with a random walk, with coarse Ricci curvature bounded from below by some $k$ such that $\kappa(x,y) \geq k$ for all $x, y \in X$. The Wasserstein Spread of Diffusion of a $t$ step diffusion in $X$ is bounded above by

$$W_1(\delta_x, m_x^t) \leq W_1\left(\delta_x, m_x\right) \frac{(1-(1-k)^t)}{k}$$

In particular, if $k > 0$ then $W_1(\delta_x, m_x^t) \leq \frac{W_1(\delta_x, m_x)}{k}$, and if $k = 0$, then $W_1(\delta_x, m_x^t) \leq tW_1\left(\delta_x, m_x\right)$.

*Proof.* First, we bound $W_1(m_x^t, m_x^{t+1})$ using Ollivier’s Proposition 20 [citation needed] on concentration of measure. The proposition states that a lower bound on curvature, such as we have, implies that

$$W_1(\mu \star m, \mu' \star m) \leq (1-k) W_1(\mu, \mu')$$

where here $\mu, \mu'$ are two probability distributions and $m$ is a random walk. This provides an easy lemma:

**Lemma 1.** Let $(X, d, m)$ be a metric space with a random walk. Suppose there is some $k \in \mathbb{R}$ such that the coarse Ricci curvature $\kappa(x,y) \geq k$ for all $x, y \in X$. Then:

$$W_1\left(m_x^t, m_x^{t+1}\right) \leq (1-k)^t W_1\left(\delta_x, m_x\right)$$

*Proof.* We proceed by induction. For $t = 0$, the above is true, as $W_1(m_x^0, m_x^1) = W_1(\delta_x, m_x^1) = W_1(\delta_x, m_x)$. Suppose it holds for $t-1$, e.g.

$$W_1\left(m_x^{t-1}, m_x^t\right) \leq (1-k)^{t-1} W_1\left(\delta_x, m_x\right)$$

Consider $W_1\left(m_x^{t-1} \star m, m_x^t \star m\right)$, the application of another step of diffusion. By Ollivier's Proposition 20, this distance is bounded above by

$$W_1\left(\mu_1 \star m, \mu_2 \star m\right) \leq (1-k) W_1\left(\mu_1, \mu_2\right)$$

So

$$W_1\left(m_x^{t-1} \star m, m_x^t \star m\right) \leq (1-k) W_1\left(m_x^{t-1}, m_x^t\right)$$

which, since the statement holds for $t-1$, yields

$$W_1\left(m_x^t, m_x^{t+1}\right) \leq (1-k)^t W_1\left(\delta_x, m_x\right)$$

To use this lemma, we can decompose our $t-$step diffusion into a sum of single step diffusions. By the triangle inequality,

$$W_1(\delta_x, m_x^t) \leq W_1\left(\delta_x, m_x\right) + W_1(m_x, m_x^2) + \cdots + W_1(m_x^{t-1}, m_x^t)$$

By Lemma 1,

$$\leq W_1\left(\delta_x, m_x\right)\left(1 + (1-k) + (1-k)^2 + \ldots + (1-k)^{t-1}\right)$$

This truncated series is equal to $\frac{1-(1-k)^t}{1-(1-k)} = \frac{(1-(1-k)^t)}{k}$. If $k > 0$, then as $t \to \infty$, the infinite sum $\sum_{i=0}^t (1-k)^i$ converges to the geometric series $\frac{1}{1-(1-k)} = \frac{1}{k}$. Because the sum is monotonically increasing with $t$, the partial sum is upper bounded by the infinite sum. It follows that

$$W_1(\delta_x, m_x^t) \leq \frac{W_1\left(\delta_x, m_x\right)}{k}$$

If $k = 0$, then obviously $\sum_i^t (1-k)^i = t$.

The above asserts that the "spread" of diffusion is bounded from above by the Ollivier Ricci curvature of the region. To turn this measurement of spread into a signed curvature, we have only to introduce a comparison: the dimensional constant of the Euclidean spread of diffusion, $W_1^E(\delta_x, m_x^t)$. Using Proposition 1, we can directly bound $k_D$ in terms of the Ollivier Ricci curvature:

$$1 - \frac{W_1^M\left(\delta_x, m_x\right) \frac{(1-(1-k)^t)}{k}}{W_1^E(\delta_x, m_x^t)} \leq k_D$$

A simple approximation of $W_1^E(\delta_x, m_x^t)$ is obtained by scaling the spread of a single-step diffusion by a factor of t, i.e. $tW_1^E(\delta_x, m_x)$. Using this, and noting that $W_1^E(\delta_x, m_x) \approx W_1^M(\delta_x, m_x)$,

$$1 - \frac{(1-(1-k)^t)}{tk} \leq k_D$$

**Note: this analysis sidesteps the reality that tJ is a (possibly egregious) *overestimate* of the euclidean WSD. We can't actually just plug it in without pushing some indeterminate space inbetween the sides of the inequality**

### 0.4.3 Comparing Diffusion Across Manifolds

For a single graph, the laziness of diffusion is linked to graph volume, and theoretically bounded by curvature: within that graph, the higher the curvature, the higher the diffusion laziness. In analogy to Bishop-Gromov, we can convert this unsigned magnitude-of-curvature estimation to a signed curvature with a comparison between spaces. In the continuous realm, this is easy: volume on one surface has the same units as volume on another. But in the point cloud setting, there is no common scale; it is *assumed* that the scale differs between and possibly even *within* a point cloud. Moreover, differences in graph construction can make diffusion spread faster independently from the local geometry – much as changing the 'size' of one's 'measuring stick' could give the appearance of larger volumes. Diabolically, common techniques for constructing graphs from variably-sampled data interfere with the speed *most* in areas with high negative or positive curvature, obscuring the geometric information in the diffusion operator.

Our challenge is to construct graphs to allow comparison of their diffusions.

#### 0.4.3.1 Standard methods of graph construction obscure geometric information

To turn a point cloud into a graph, one places a kernel (e.g. a gaussian) at each point, and assigns edges to surrounding points with weight given by the kernel. The main parameter is the bandwidth of each kernel: make it too high, and everything is connected; too low, and points are isolated. Moreover, the distances between a point and its neighbors may vary. Where the points are sparse, one wants a higher bandwidth. Where dense, a lower bandwidth.

#### 0.4.3.2 The Curvature Agnostic Kernel

## 0.5 Results

Curvature is an easy quantity to find – indeed, at least two of the many recent curvature methods were defined *accidentally* [cite Steinerberger, myself]! The challenge is to make one's definition robust to the bogeyman of real-world data, especially 1) noise and 2) high dimensions. Our benchmarks of existing methods, including Hickok & Blumberg's, Ollivier Ricci Curvature, Adal-PCA [and Sritharan], [citations] reveal that all deteriorate quickly under noise, and most *fail entirely* in high-dimensional data. The dimension-independent, noise-resilient properties of the diffusion operator endow diffusion curvature with good performance on both of these axes.

But first, a sanity check. Does diffusion curvature recover the Gaussian curvature of basic 2-manifolds? Figure WAWA shows the diffusion curvature of a Torus, Saddle, and Ellipsoid, along with a scatter plot correlating diffusion curvature and Gaussian curvature. Figure WAWA2 places this in the context of existing methods.

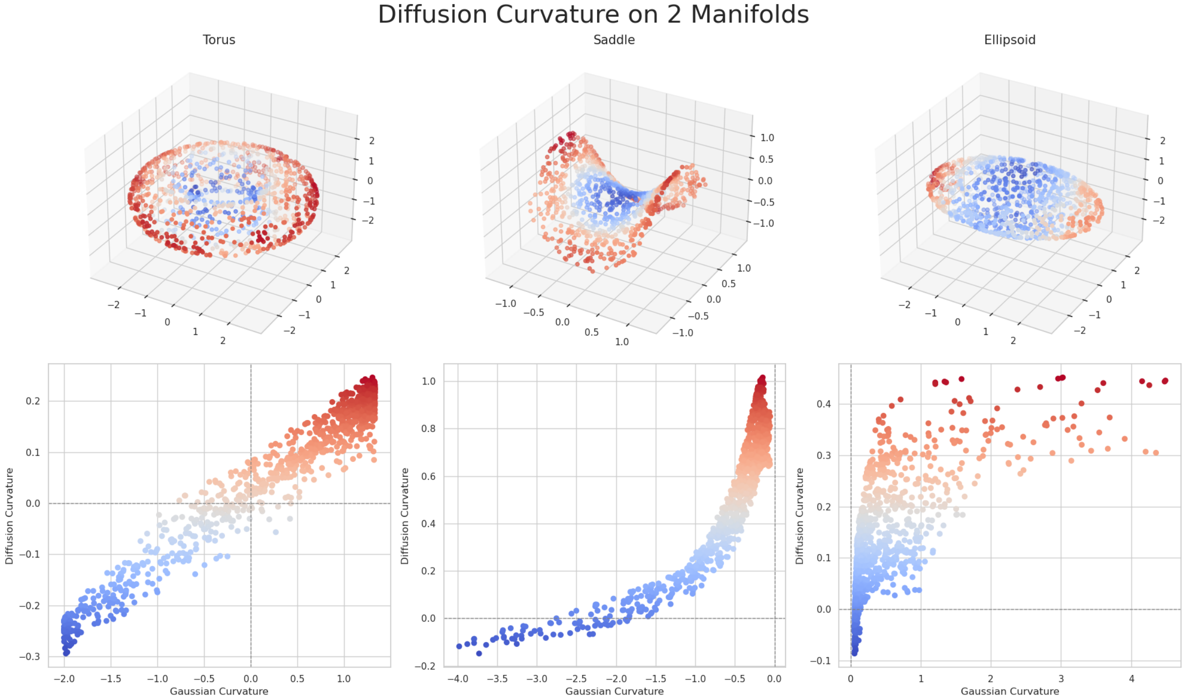

Figure 1: Diffusion Curvature vs Gaussian Curvature on 2-Manifolds.

Source: Standard libraries

Though the correlation is strong – passing the 'sniff test' – this low-dimensional validation highlights two subtleties of our method. First, diffusion curvature, being an intrinsic, graph-based measurement, is susceptible to edge effects to a greater degree than extrinsic methods.

When diffusion hits the edges of the saddle, it rebounds, creating the false appearance of positive curvature.

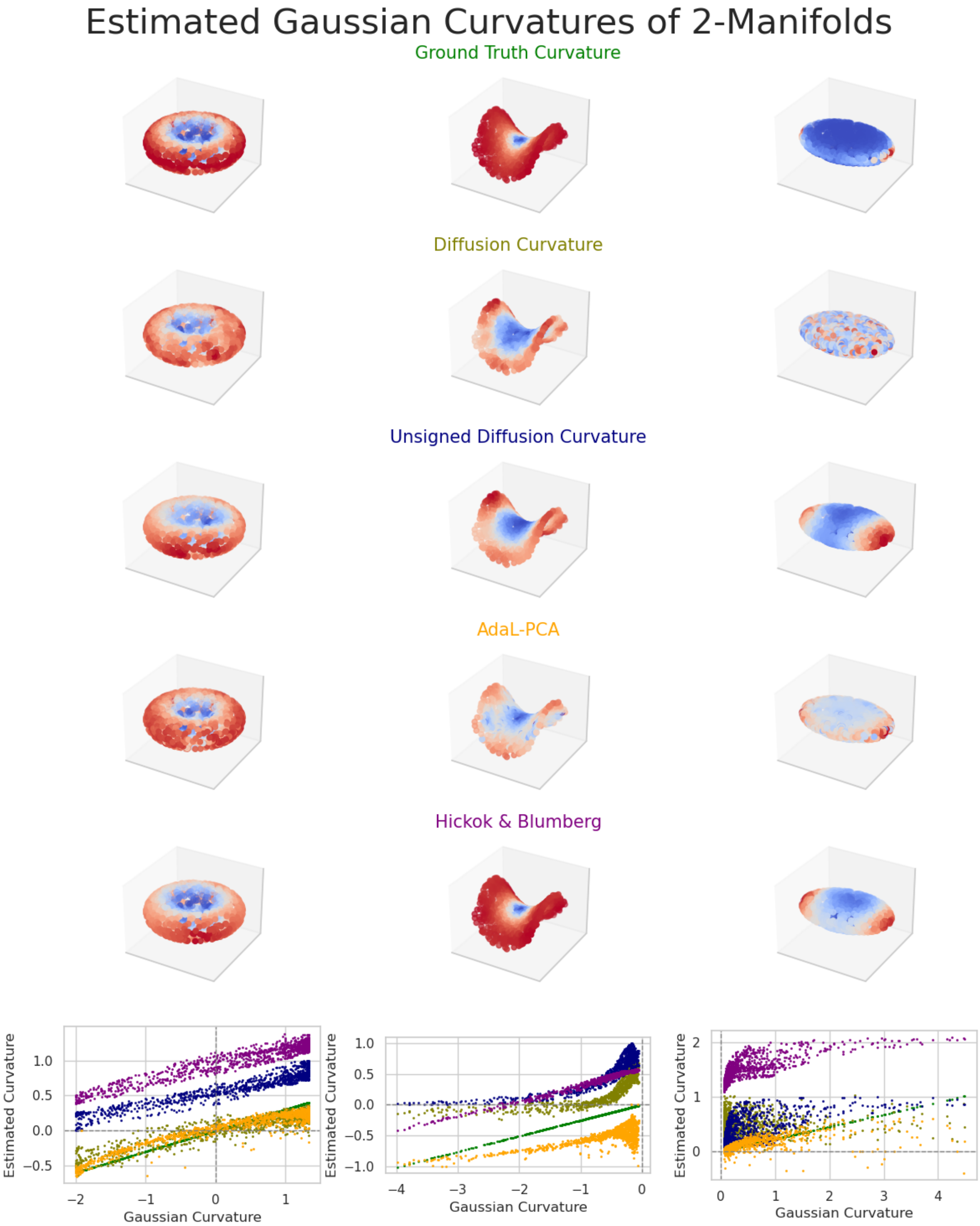

Figure 2: Diffusion Curvature vs Gaussian Curvature on 2-Manifolds.

Source: Standard libraries

The second subtly emerges in the context of related methods. Everything pictured does an excellent job of coloring the toy manifolds. But looking beyond this, *what matters?* Most methods don't care about the precise magnitude – or even exactly matching the correlations, as in the ellipsoid… %% mean vs gaussian curvature, and uniqueness of definitions%%

All of these methods in Fig WAWA2 perform well on well-sampled noiseless toy manifolds. In Table WAWA3, we see the results of adding Gaussian noise to each manifold.

![[CleanShot 2024-04-28 at 23.30.38_147B95A7.png]]

### 0.5.1 Differentiating Sign in High Dimensions

Most existing methods only quantify their performance in high dimensions on one or two test cases.

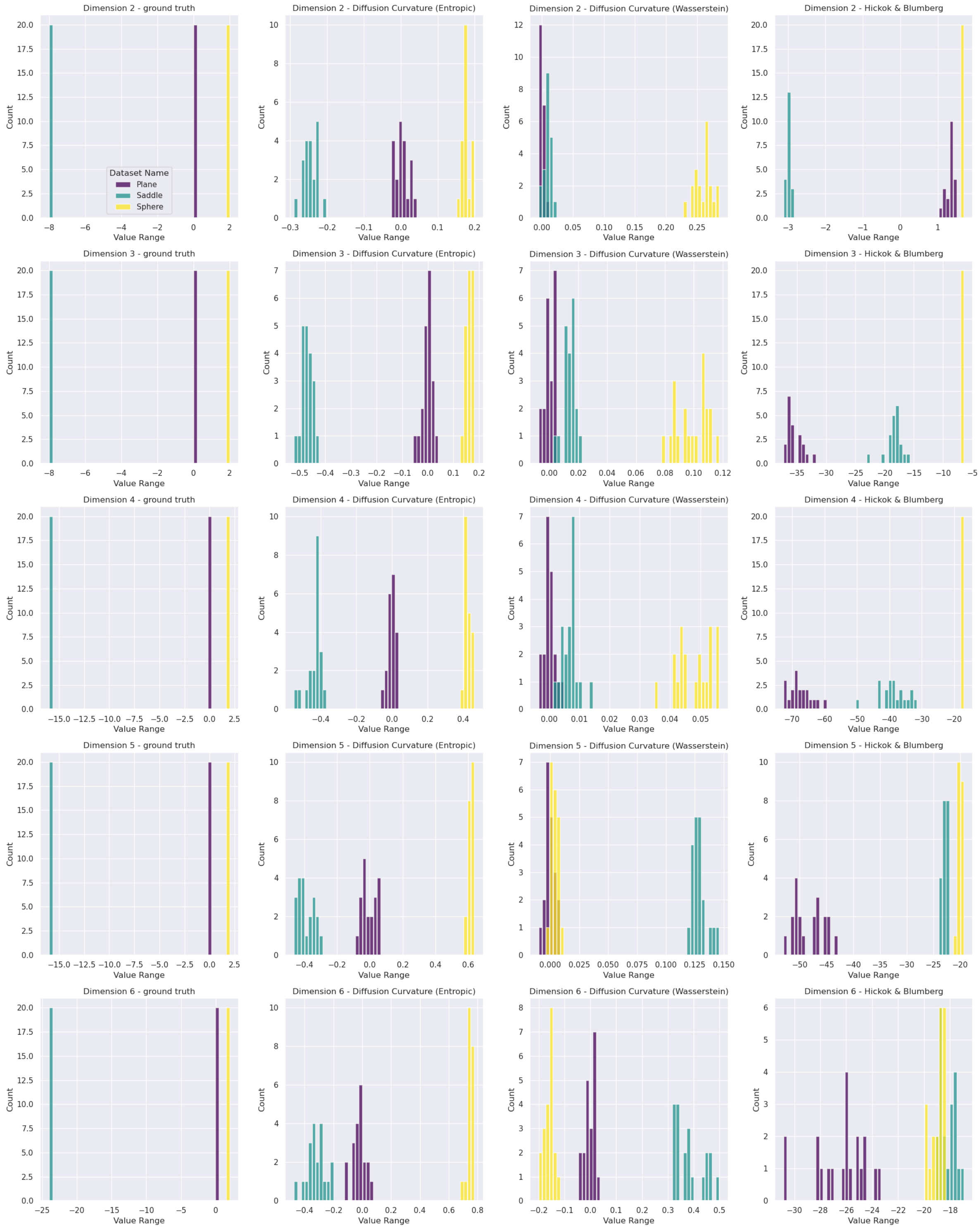

Figure 3: Predicted curvatures of Saddles and Spheres in dimensions 2-6. Diffusion Curvature robustly distinguishes between the signs of the data, even in high dimensions, and with relative sparsity.

Source: Standard libraries

### 0.5.2 Loss Landscapes

### 0.5.3 Curvature as a TDA Filtration

## 0.6 Conclusion

## References

Bhaskar, Dhananjay, Kincaid MacDonald, Oluwadamilola Fasina, Dawson Thomas, Bastian Rieck, Ian Adelstein, and Smita Krishnaswamy. 2022. "Diffusion Curvature for Estimating Local Curvature in High Dimensional Data." *Advances in Neural Information Processing Systems* 35: 21738–49. https://proceedings.neurips.cc/paper_files/paper/2022/hash/88438dc62fc5c8777e2b5f1b4f6d37a2-Abstract-Conference.html.