DIFFUSION-BASED METHODS FOR ESTIMATING CURVATURE IN DATA

Dhananjay Bhaskar[†], **Kincaid MacDonald**[†], **Dawson Thomas, Sarah Zhao, Kisung You, Jennifer Paige, Yariv Aizenbud, Ian Adelstein**[*] **& Smita Krishnaswamy**[*]
Yale University
New Haven, CT 06510, USA
{ian.adelstein,smita.krishnaswamy}@yale.edu

Bastian Rieck
Institute of AI for Health
Helmholtz Pioneer Campus
Munich, Germany

ABSTRACT

High-throughput high-dimensional data is now being generated in massive quantities in many fields including biology, medicine, chemistry, finance and physics. Researchers have successfully used manifold learning in order to gain insight from such data, particularly in biomedical and single cell data. One such technique, data diffusion geometry, has been useful in understanding manifold intrinsic distances, density, and major non-linear axes or paths through the data. However, a relatively unstudied feature of high-dimensional data is curvature. While curvature is well-defined and easy to compute in low dimensions, it poses computational and conceptual difficulties in high dimensions. Here, we present two techniques to estimate curvature from high-dimensional data starting from data diffusion probabilities. The first technique, *diffusion curvature*, uses the spread or conversely laziness of a random walk to estimate curvature pointwise in data. The second technique, *deep diffusion curvature*, trains a neural network to estimate pointwise curvature. Since these techniques are scalable, we anticipate that they can be used to describe and compare datasets as well as find points in data that represent transitional entities.

1 INTRODUCTION

With the advent of high-throughput high-dimensional data in many fields including biomedicine, social science, physics, and finance, there is an increasing need to understand the structure of data. These investigations usually start with the manifold hypothesis: that high-dimensional data lie on or near a much lower dimensional embedded manifold.

The diffusion geometry and diffusion map framework first introduced by Coifman et al. Coifman & Maggioni (2006) has proven to be a useful framework for understanding geometric features of the data such as manifold-intrinsic data geometry, geodesics, and intrinsic manifold dimensions. However, data curvature, a salient feature of Riemannian geometry, has received less attention. Being an intrinsically smooth quantity, its extension to discrete case is not straightfoward.

In this paper, we use the diffusion operator to build a bridge from the discrete data setting to the smooth notion of curvature. We propose two scalable methods for estimating curvature from high dimensional point-cloud data. The first method, which we call *diffusion curvature* estimates curvature by the tendency of a random walk to stay near its point of origin in positive curvature and diffuse away in negative curvature. Random walk probabilities are computed via a data diffusion operator. The second method uses data diffusion probabilities as input to a neural network that estimates curvature via training on idealized surfaces. We hypothesize that these methods will be useful

[*] co-senior authors; †co-first authors

in introducing curvature as a computable and useful feature for characterizing data from a variety of domains, particularly biomedical domains.

2 Background and Preliminaries

Manifold Assumption and Data Diffusion. A useful assumption in representation learning is that high-dimensional data originates from an intrinsic low-dimensional manifold that is mapped via nonlinear functions to observable high-dimensional measurements; this is commonly referred to as the manifold assumption. Formally, let $\mathcal{M}^d$ be a hidden d dimensional manifold that is only observable via a collection of $n \gg d$ nonlinear functions $f_1, \ldots, f_n : \mathcal{M}^d \to \mathbb{R}$ that enable its immersion in a high-dimensional ambient space as $F(\mathcal{M}^d) = \{\mathbf{f}(z) = (f_1(z), \ldots, f_n(z))^T : z \in \mathcal{M}^d\} \subseteq \mathbb{R}^n$ from which data is collected. Conversely, given a dataset $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ of high-dimensional observations, manifold learning methods assume data points originate from a sampling $Z = \{z_i\}_{i=1}^N \in \mathcal{M}^d$ of the underlying manifold via $x_i = \mathbf{f}(z_i)$, $i = 1, \ldots, N$, and aim to learn a low dimensional intrinsic representation that approximates the manifold geometry of $\mathcal{M}^d$.

To learn a manifold geometry from collected data, we use the popular diffusion maps construction of Coifman & Maggioni (2006). This construction starts by considering local similarities defined via a kernel $\mathcal{K}(x, y)$, $x, y \in F(\mathcal{M}^d)$ that captures local neighborhoods in the data. While the Gaussian kernel is a popular choice, it encodes sampling density information in its computation.

To construct a diffusion geometry that is robust to sampling density variations, we use an anisotropic kernel

$\mathcal{K}(x, y) = {\mathcal{G}(x,y)}/{\|\mathcal{G}(x,\cdot)\|_1^\alpha \|\mathcal{G}(y,\cdot)\|_1^\alpha}$, with $\mathcal{G}(x, y) = e^{-\|x-y\|^2/\sigma}$

as proposed in Coifman & Maggioni (2006), where $0 \leq \alpha \leq 1$ controls the separation of geometry from density, with $\alpha = 0$ yielding the classic Gaussian kernel, and $\alpha = 1$ completely removing density and providing a geometric equivalent to uniform sampling of the underlying manifold.

Next, the similarities encoded by $\mathcal{K}$ are normalized to define transition probabilities $p(x, y) = {\mathcal{K}(x,y)}/{\|\mathcal{K}(x,\cdot)\|_1}$ that are organized in an $N \times N$ row stochastic matrix $\mathbf{P}_{ij} = p(x_i, x_j)$ that describes a Markovian diffusion process over the intrinsic geometry of the data.

Finally, a diffusion map is defined by taking the eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ and (corresponding) eigenvectors $\{\phi_j\}_{j=1}^N$ of $\mathbf{P}$, and mapping each data point $x_i \in X$ to an N dimensional vector $\Phi_t(x_i) = [\lambda_1^t \phi_1(x_i), \ldots, \lambda_N^t \phi_N(x_i)]^T$, where t represents a diffusion-time. In general, as t increases, most of the eigenvalues become negligible; truncated diffusion map coordinates can thus be used for dimensionality reduction.

Graph Laplacian. The diffusion coordinates are closely related to the eigenfunctions of Laplace operators on manifolds, as well as their discretizations as eigenvectors of graph Laplacians.

There are multiple variants of graph Laplacians. The standard graph Laplacian is defined as $\mathbf{L} := \mathbf{D} - \mathbf{W}$, with $\mathbf{D}$ being the diagonal degree matrix of a graph, where $\mathbf{D}_{ii} := \sum_j \mathbf{W}_{ij}$ and $\mathbf{W}_{ij} = \mathcal{K}(x_i, x_j)$. Being a symmetric matrix with positive values, it is diagonalizable. Considering the Laplacian as an operator, it can represent a scalar-valued "signal" $f \colon V \to \mathbb{R}$ on the vertices V of the graph as $\mathbf{L}f = \sum_{v_i \sim v_j} (f(v_i) - f(v_j))$.

Viewed as an operator, this Laplacian is cumbersome, though, because its eigenvalues are unbounded. Thus, one typically uses the *normalized graph Laplacian* instead, which is defined as $\boldsymbol{\mathcal{L}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$.

Note that $\boldsymbol{\mathcal{L}}$ maintains the representation properties of $\mathbf{L}$, and for a real-valued signal g, we have $\langle g, \boldsymbol{\mathcal{L}} g\rangle = \langle f, \mathbf{L}f\rangle = \sum_{v_i \sim v_j} (f(v_i) - f(v_j))^2$, with $f = \mathbf{D}^{-1/2} g$.

Finally, one can see that $\boldsymbol{\mathcal{L}} = \mathbf{I} - \mathbf{D}^{1/2}\mathbf{P}\mathbf{D}^{-1/2}$, and thus it can be verified that the eigenvectors of $\boldsymbol{\mathcal{L}}$ can be written as $\psi_j = \mathbf{D}^{1/2}\phi_j$, with corresponding eigenvalues $\omega_j = 1 - \lambda_j$.

It should be noted that when using the anisotropic kernel with $\alpha = 1$, the data can be treated as being uniformly sampled from the data manifold (Belkin & Niyogi, 2008), in which case these two sets of eigenvectors coincide and the diffusion coordinates can be considered Laplacian eigenvectors.

Asymptotics of the Heat Trace. The eigenvalues of the Laplace-Beltrami operator encode geometry information about a compact Riemannian manifold, as can be seen via the asymptotic expansion of the trace of the heat kernel (Gordon, 2000). Dimension, volume, and total scalar curvature are spectrally determined properties of the manifold. The Bochner formula similarly relates the Laplace-Beltrami operator to the Ricci curvature. Given the relationship between diffusion coordinates and the eigenvectors of the Laplace-Beltrami operator, one can surmise that the diffusion operator also encodes geometric information. These relations motivate our use of the diffusion operator to measure curvature on a data manifold. Appendix A contains additional details in the Riemannian setting.

3 METHODS

3.1 DIFFUSION CURVATURE

To build intuition, we first discuss the case of surfaces (dimension $n = 2$). Three canonical surfaces are the sphere (or surface of a ball), the cylinder, and the saddle. The Gaussian curvature of these surfaces are positive, zero, and negative, respectively. Imagine taking a sticker and trying to adhere it to one of these surfaces. The sticker, having been printed on a flat piece of paper, has zero Gaussian curvature. It will adhere perfectly to the cylinder, will bunch up (there will be too much sticker material) when trying to adhere it to the sphere, and will rip (there will be too little sticker material) when trying to adhere it to the saddle. This example builds intuition for the area comparison definition of Gaussian curvature, where one computes the limiting difference between the area $A(r)$ of a geodesic disk on the manifold and a standard Euclidean disk, i.e., $K = \lim_{r \to 0^+} 12^{\pi r^2 - A(r)}/_{\pi r^4}$. For example, Gaussian curvature will be positive when $\pi r^2 > A(r)$, i.e. when the area of the sticker exceeds the area of the corresponding geodesic disk on the sphere. The Bishop-Gromov volume comparison theorem formalizes how to extend this into dimensions $n > 2$.

Theorem 1 (Bishop-Gromov) *Let (M^n, g) be a complete Riemannian manifold with Ricci curvature bounded below by $(n-1)K$. Let $B(p, r)$ denote the ball of radius r about a point $p \in M$ and $B(p_K, r)$ denote a ball of radius r about a point p_K in the complete n-dimensional simply connected space of constant sectional curvature K. Then $\phi(r) = \frac{Vol\ B(p,r)}{Vol\ B(p_K,r)}$ is a non-increasing function on $(0, \infty)$ which tends to 1 as $r \to 0$. In particular, $Vol\ B(p, r) \leq Vol\ B(p_K, r)$.*

This theorem captures the sticker phenomenon: as curvature increases, the volume of comparable geodesic balls decreases. Positive curvature corresponds to geodesic convergence and smaller volumes, whereas negative curvature corresponds to geodesic divergence (spread) and larger volumes. The discrete nature of the data manifold makes it impossible to compare volumes as the distance scale (radius) goes to zero. However, by comparing volume (or mass) at various diffusion time scales, we are able to capture this relative spreading of geodesics.

Definition 1 *The pointwise* Diffusion Curvature $C(i)$ *is the probability that a random walk starting from a point i ends at one of the k-nearest neighbors of i after t steps of data diffusion, i.e., $C(i) := v_i P^t n_i$, where v_i is the i^{th} row of the identity matrix and n_i is the i^{th} column of the k-nearest neighbor indicator matrix, with entries of 1 for each k^{th} nearest neighbor and zero otherwise.*

We have $C(i) \in [0, 1]$, with larger values indicating higher curvature relative to lower values. It estimates the curvature of a point on a graph by measuring the "laziness" of diffusions emanating from that point. Intuitively, a random walker is more likely to return to their starting point in a region of positive curvature (where paths converge) than negative curvature (where paths diverge). In negatively curved regions the random walker can get lost in the various divergent (disconnected) branches, and has a lower chance of returning to its origin. In positively curved regions, the paths of the random walker exhibit more inter-connectivity, so that the return probability of a walk is higher. As the diffusion operator can be viewed as a collection of transition probabilities of random walks, the "laziness" of powered diffusions serves as an indicator of pointwise curvature.

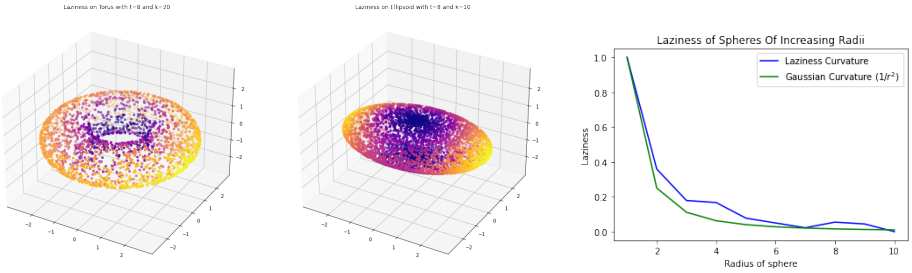

Figure 1: Diffusion curvature on the torus and ellipsoid (lighter points have higher values), and spheres of increasing radii.

3.2 Deep Diffusion Curvature

Having established a theoretical basis for the presence of curvature information in diffusion geometry, we next experimented with a feed-forward neural network to predict Gaussian curvature from diffusion probabilities.

To train the network, we sampled points uniformly (w.r.t area) on unit disks of constant curvature, inspired by the approach of Bubenik et al. (2020). In summary, the probability of a point lying within a disk of curvature r, at uniform density, is equal to the ratio of the area of the disk and the area of the unit disk.

To sample points (in polar coordinates) on the disk, we sample angles θ uniformly from $[0, 2\pi]$ and compute r using the inverse CDF (see Appendix A.3 for equations).

We generated samples for 81 linearly-spaced curvature values in the interval $[-2, 2]$. For each curvature value, we sampled 1000, 2000, and 5000 points on 100 disks. Next, we computed diffusion probabilities of a Dirac centered at the origin, which we used to train a feed-forward neural network with 4 hidden layers to predict curvature.

4 Results and Conclusion

To illustrate diffusion curvature, we tested this method on several toy surfaces including a torus, a hyperboloid, and an ellipsoid. Each dataset was constructed using rejection sampling to mitigate interference from density differences. We performed a single pass of smoothing on the laziness values to correct for spurious local variations arising from dense/sparse regions. As seen in Figure 1, the laziness value correlates well with the expected curvature of each region: the positively-curved outer ring of the torus has a higher laziness value than the saddle-like hole (likewise for the elllipse). We conclude that $C(i)$ visibly reconstructs the curvature of the underlying manifold.

The torus and ellipsoid examples in Figure 1 show that diffusion curvature can measure the relative curvatures of points on a manifold. To test whether this measure can be used to estimate the absolute curvature *across* datasets, we calculated the average diffusion curvature of spheres of increasing radii, by sampling a uniform number of points from a polar cap of fixed height drawn from each sphere. In Figure 1 (right), we plot the ground-truth Gaussian curvature of the spheres alongside the average diffusion curvature (normalized between 0 and 1). The close correspondence of the two measures suggests that, given relatively uniform samplings, the diffusion curvature encodes some degree of absolute curvature, and could be used to draw comparisons across datasets.

Moreover, we tested the deep diffusion curvature network on a torus ($r = 1, R = 3$) and an ellipsoid ($a = 2, b = 2.5, c = 3$) by predicting Gaussian curvature at sampled points (Table 1 and Figure 2 in the Appendix). We fixed the power of the diffusion operator ($t = 5$) and varied the number of point samples in the training and test data to evaluate the model performance at varying sampling densities. For a fixed-step random walk, we expect that the Dirac signal will diffuse more under dense sampling, hence increasing the ability of the model to estimate curvature based on diffusion probabilities. Our experimental findings (Table 1) broadly agree with this hypothesis, with RMSE decreasing with increase in sampling density for both the torus and the ellipsoid. However, despite

# samples on training manifold	# samples on test manifold (samples per unit surface area)					
	Torus ($r = 1, R = 3$)			**Ellipsoid** ($a = 2, b = 2.5, c = 3$)		
	1000 (8.444)	2000 (16.887)	5000 (42.217)	1000 (12.797)	2000 (25.594)	5000 (63.986)
1000	0.540	0.548	0.569	0.815	0.711	0.636
2000	0.542	0.541	0.554	0.794	0.703	0.637
5000	0.537	0.515	0.497	0.776	0.702	0.612

Table 1: RMSE of curvature estimation at varying sampling density of training and test manifolds.

the higher sampling density on the ellipse test cases, the model performs better on the torus. This is likely due to the range of curvature values on which the model is trained and tested. The ellipse curvature varies between 0.06 and 0.36 whereas the curvature of the torus has a broader range, from -0.45 to 0.25. Recall that the model is trained using 81 linearly spaced curvature values between -2 and 2. Therefore, the training data contains 600 samples with curvature within the range of the ellipse and 1400 samples with curvature within the range of the torus.

Finally, we estimated the curvature of a publicly available mass cytometry dataset of induced pluripotent stem cell (iPSC) reprogramming of 2005 mouse embryonic fibroblasts (Zunder et al., 2015). This type of data has a branched hyperbolic structure (Moon et al., 2019), which our model correctly associates with negative curvature (see Figure 3 in the Appendix).

Conclusion. Access to curvature information via diffusion presents an exciting opportunity for understanding the geometry of high-throughput, high-dimensional datasets. Here, we report preliminary results on estimating Gaussian curvature of 'toy' datasets using diffusion curvature and deep diffusion curvature. In future work, we will test this approach on larger benchmark datasets, and investigate the possibility of recovering other notions of curvature, including scalar curvature, sectional curvature, and the Riemann curvature tensor.

REFERENCES

Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, December 2008. ISSN 0022-0000. doi: 10.1016/j.jcss.2007.08.006.

Peter Bubenik, Michael Hull, Dhruv Patel, and Benjamin Whittle. Persistent homology detects curvature. *Inverse Problems*, 36(2):025008, February 2020. ISSN 0266-5611, 1361-6420. doi: 10.1088/1361-6420/ab4ac0. URL `http://arxiv.org/abs/1905.13196`. arXiv: 1905.13196.

Ronald R. Coifman and Mauro Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21(1):53–94, July 2006. ISSN 1063-5203. doi: 10.1016/j.acha.2006.04.004.

Carolyn S Gordon. Chapter 6 - survey of isospectral manifolds. *Handbook of Differential Geometry, Editor(s): Franki J.E. Dillen, Leopold C.A. Verstraelen*, 1:747–778, 2000. ISSN 1874-5741. doi: 10.1016/S1874-5741(00)80009-6.

Shane Lubold, Arun G. Chandrasekhar, and Tyler H. McCormick. Identifying the Latent Space Geometry of Network Models through Analysis of Curvature. Technical Report 28273, National Bureau of Economic Research, Inc, December 2020. URL `https://ideas.repec.org/p/nbr/nberwo/28273.html`. Publication Title: NBER Working Papers.

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing Structure and Transitions in High-Dimensional Biological Data. *Nature biotechnology*, 37(12):1482–1492, December 2019. ISSN 1087-0156. doi: 10.1038/s41587-019-0336-3.

Areejit Samal, R. P. Sreejith, Jiao Gu, Shiping Liu, Emil Saucan, and Jürgen Jost. Comparative analysis of two discretizations of Ricci curvature for complex networks. *Scientific Reports*, 8(1):

8650, June 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-27001-3. Number: 1 Publisher: Nature Publishing Group.

Jayson Sia, Edmond Jonckheere, and Paul Bogdan. Ollivier-Ricci Curvature-Based Method to Community Detection in Complex Networks. *Scientific Reports*, 9(1):9800, July 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-46079-x. Number: 1 Publisher: Nature Publishing Group.

Duluxan Sritharan, Shu Wang, and Sahand Hormoz. Computing the Riemannian curvature of image patch and single-cell RNA sequencing data manifolds using extrinsic differential geometry. *Proceedings of the National Academy of Sciences*, 118(29), July 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2100473118. Publisher: National Academy of Sciences Section: Biological Sciences.

Eli R. Zunder, Ernesto Lujan, Yury Goltsev, Marius Wernig, and Garry P. Nolan. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell*, 16(3):323–337, March 2015. ISSN 1875-9777. doi: 10.1016/j.stem.2015.01.015.

A Appendix

A.1 Asymptotics of the heat trace

Given the strong relationship between diffusion coordinates and the eigenvectors of the Laplace-Beltrami operator, one can believe that these two objects capture similar properties of the manifold. As explored below, the eigenvalues of the Laplace-Beltrami operator relate to the curvature of a compact Riemannian manifold via the asymptotics of the trace of the heat kernel, and this lends credence to use of the diffusion probabilities to measure curvature on a data manifold. This section follows closely the exposition in Gordon (2000).

On a compact Riemannian manifold M we define the Laplace-Beltrami operator Δ to be $\Delta(f) = -div(grad(f))$. We consider the eigenvalues $\Delta(f) = \lambda f$ and note that the spectrum of Δ acting on $L^2(M)$ is a discrete set of non-negative numbers $\{0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots\}$ where each eigenvalue is written as many times as its multiplicity.

For compact Riemannian manifolds the heat kernel (or fundamental solution of the heat equation $u_t + \Delta(u) = 0$) exists uniquely and is given by

$$K(t, x, y) = \sum_j e^{-\lambda_j t} \phi_j(x) \phi_j(y)$$

where the λ_j and ϕ_j are the eigenvalues and associated eigenfunctions of the Laplace-Beltrami operator, with the eigenfunctions normalized to form an orthonormal basis of $L^2(M)$.

We take the trace of the heat kernel and note that

$$Z(t) \equiv \int_M K(t, x, x)\, dx = \sum_{j=0}^{\infty} e^{-\lambda_j t}$$

so that the spectrum $\{\lambda_j\}$ determines the heat trace $Z(t)$.

Minakshisundaram's formula for the asymptotic expansion of $Z(t)$ as $t \to 0^+$ is given by

$$Z(t) = (4\pi t)^{-n/2} \sum_{k=1}^{\infty} a_k t^k$$

where the coefficients a_k are expressed via the metric and its derivatives. The first handful of coefficients have been calculated explicitly:

$$a_0 = vol(M), \qquad a_1 = \frac{1}{6} \int_M \tau, \qquad a_2 = \frac{1}{360} \int_M (5\tau^2 - 2|Ric|^2 - 10|R|^2)$$

where n is the dimension, τ is the scalar curvature, Ric is the Ricci tensor, and R is the curvature tensor. In this way we see that the Laplace-Beltrami operator, via its eigenvalues, determines geometric information about the manifold, including the dimension, volume, and total scalar curvature. In the closed surface setting ($n = 2$) one can use Gauss-Bonnet to additionally recover the genus.

While the spectrum of the Laplace-Beltrami operator certainly contains interesting geometric information, it does not determine the isometry class of the manifold. Indeed, Milnor in 1964 produced the first examples of isospectral, non-isometric manifolds. Many additional examples have been produced, and with each pair one can reveal geometric information that is not encoded in the spectrum. These inaudible properties include the fundamental group, diameter, maximum scalar curvature, and even the local geometry of the manifold. One must therefore be prudent when using diffusion (and its relation to the Laplace-Beltrami operator) to prescribe curvature on a data manifold.

A.2 Bochner formula

The Bochner formula similarly relates the Laplace-Beltrami operator on a Riemannian manifold to its Ricci tensor. It states

$$\frac{1}{2}\Delta|\nabla f|^2 = g(\nabla\Delta f, \nabla f) + |\nabla^2 f|^2 + Ric(\nabla f, \nabla f)$$

where Δ is the Laplace-Beltrami operator, $g(\cdot,\cdot)$ is the metric, ∇ is the gradient, ∇^2 is the Hessian, and Ric is the Ricci tensor. When $f \in C^\infty(M)$ is harmonic, i.e. $\Delta(f) = 0$, the metric term disappears and we have

$$\frac{1}{2}\Delta|\nabla f|^2 = |\nabla^2 f|^2 + Ric(\nabla f, \nabla f).$$

We take these formulas as further evidence that the diffusion operator should contain information about the curvature of its associated data manifold.

A.3 Cumulative Density Functions for Disks of Unit Curvature

Following Bubenik et al. (2020), we sample points uniformly from unit disks of constant curvature. The cumulative probability density is given by:

$$F(r) = \begin{cases} \frac{\pi r^2}{\pi 1^2} = r^2 & \text{if } K = 0 \text{ (Euclidean)}, \\ \frac{\frac{4\pi}{K}\sin^2(\frac{r\sqrt{K}}{2})}{\frac{4\pi}{K}\sin^2(\frac{1\sqrt{K}}{2})} & \text{if } K > 0 \text{ (Spherical)}, \\ \frac{\frac{4\pi}{-K}\sinh^2(\frac{r\sqrt{-K}}{2})}{\frac{4\pi}{-K}\sinh^2(\frac{1\sqrt{-K}}{2})} & \text{if } K < 0 \text{ (Hyperbolic)} \end{cases} \tag{1}$$

To sample points (in polar coordinates) on the disk, we sample angles θ uniformly from $[0, 2\pi]$ and compute r using the respective inverse CDF:

$$r = F^{-1}(u) = \begin{cases} \sqrt{u} & \text{if } K = 0, \\ \frac{2}{\sqrt{K}}\sin^{-1}\left(\sqrt{u}\sin(\frac{\sqrt{K}}{2})\right) & \text{if } K > 0, \\ \frac{2}{\sqrt{-K}}\sinh^{-1}\left(\sqrt{u}\sinh(\frac{\sqrt{-K}}{2})\right) & \text{if } K < 0 \end{cases} \tag{2}$$

B Related Works

We motivated curvature as a way to obtain a more local understanding of the data manifold than that given by topological methods alone. However, Bubenik et al. (2020) show that information typically discarded as noise in persistent homology can actually recover some geometric information. Namely, they show that small scale features in a point cloud can detect the constant curvature of a disk from which they are sampled.

Curvature has been relatively widely studied in the specific discrete setting of graphs. Diffusion methods almost always involve building an initial graph on the data, making this setting particularly

relevant to our work in this paper. Discretizations of curvature, including Ollivier-Ricci and Forman-Ricci curvature (Samal et al., 2018) are constructed by attempting to preserve particular properties from the smooth setting. For example, the spread of geodesics determines the cost of transporting fixed-radius balls across a smooth manifold, so Ollivier-Ricci curvature defines an edge-wise measure of curvature based on optimal transport of probability distributions between nodes. These measures have seen application in detecting hierarchical structure and communities within networks (Sia et al., 2019). On the other hand, Lubold et al. (2020) start with data that is naturally represented as a network and explore how the (constant) curvature of the space affects the ability to appropriately embed the data as a graph.

Sritharan et al. (2021) take an extrinsic approach to determine curvature on a data manifold at each point. Their method involves a local tangent space approximation. After performing a quadratic polynomial fit on the normal coordinates, the elements of the Hessian matrix h^ν for each normal coordinate ν (which coincides with the second fundamental form when the plane is tangent) are easily read off. These determine the Riemann curvature tensor via the Gauss-Codazzi equations

$$R_{ijkl} = (h^\alpha_{jk} h^\beta_{il} - h^\beta_{ji} h^\alpha_{kl}) g_{\alpha\beta}$$

Interestingly, the authors claim that intrinsic approaches based on approximations to the Laplace-Beltrami operator (i.e. the diffusion operator) yield inaccurate results. This is based on the poor convergence of an experiment they ran to determine the average scalar curvature on a sphere by way of the heat trace expansion. However, the success of the approach presented here, which relies on more than just spectral data, indicates that this operator does indeed capture valuable curvature information.

C Additional Figures

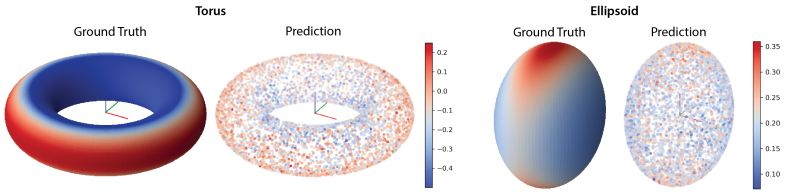

Figure 2: Curvature predicted at 5000 points sampled on a torus ($r = 1.0, R = 3.0$) and ellipsoid ($a = 2.0, b = 2.5, c = 3.0$) using a feed-forward network trained on manifolds of constant curvature.

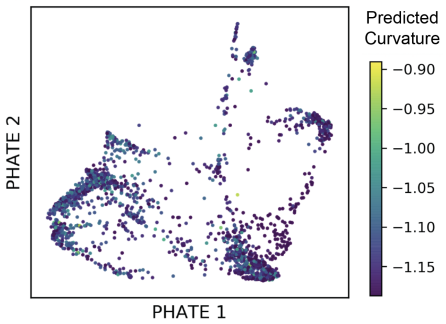

Figure 3: PHATE plot showing curvature estimation of iPSCs using deep diffusion curvature.