Received: 25 April 2017 | Accepted: 13 February 2018 | Published on: 14 November 2018

DOI: 10.1002/rsa.20800

RESEARCH ARTICLE

WILEY

Random Čech complexes on Riemannian manifolds

Omer Bobrowski[1] | **Goncalo Oliveira**[2]

[1]Viterbi Faculty of Electrical Engineering, Technion - Israel Institue of Technology Haifa, Israel

[2] Universidade Federal Fluminense, Instituto de Matemática, Niterói, Rio de Janeiro, Duke Univerisity, Durham NC, USA

Correspondence

Omer Bobrowski, Viterbi Faculty of Electrical Engineering, Technion – Israel Institue of Technology, Haifa, Israel.

Email: omer@ee.technion.ac.il

Abstract

In this paper we study the homology of a random Čech complex generated by a homogeneous Poisson process in a compact Riemannian manifold M. In particular, we focus on the phase transition for "homological connectivity" where the homology of the complex becomes isomorphic to that of M. The results presented in this paper are an important generalization of [7], from the flat torus to general compact Riemannian manifolds. In addition to proving the statements related to homological connectivity, the methods we develop in this paper can be used as a framework for translating results for random geometric graphs and complexes from the Euclidean setting into the more general Riemannian one.

KEYWORDS

cech complexes, homological connectivity, random topology, riemannian manifolds

1 | INTRODUCTION

1.1 | Motivation

In this paper we continue the work in [7], and extend the results on the homological connectivity (or the vanishing of homology) in random Čech complexes from the d-dimensional flat torus $\mathbb{T}^d$, to general d-dimensional compact Riemannian manifolds.

The study of random simplicial complexes originated in the seminal result of Erdős and Rényi [14] on the phase transition for connectivity in random graphs $G(n, p)$ (with n vertices, and where edges are included independently and with probability p) . In their paper, Erdős and Rényi studied these graphs in the limit when $n \to \infty$ and $p = p(n) \to 0$, and showed that the phase transition for connectivity occurs around $p = \log n/n$, when the expected degree is approximately $\log n$.[1] Over the past

[1]Note that this is the more familiar formulation of the model, while the one in [14] is slightly different, yet equivalent.

decade, a body of results was established for higher dimensional generalizations of the $G(n, p)$ graph. In these generalizations, graphs are replaced by simplicial complexes, where in addition to vertices and edges we may include triangles, tetrahedra and higher dimensional simplexes. For example, in the *Linial-Meshulam k-complex* [30, 31], one starts with the full $(k-1)$-skeleton on n vertices, and then attaches the k-faces independently and with probability p. In a different model, called the *clique complex* [23, 25], one starts with a random graph $G(n, p)$ and then adds a k-face for every $(k+1)$-clique in the graph. We will refer to these generalizations as random *combinatorial* complexes (see [26] for a survey). It turns out that the Erdős -Rényi threshold for connectivity can be generalized to that of "homological connectivity," where the higher homology groups H_k become trivial.

In parallel to the study of combinatorial complexes, a line of research was established for random *geometric* complexes [3, 5, 24, 42, 43]. This type of complexes generalizes the model of the *random geometric graph* $G(n, r)$ (introduced by Gilbert [18]), where vertices are placed at random in a metric-measure space, and edges are included based on proximity [38]. The main differences between the geometric and the combinatorial models are twofold. First, in geometric complexes edges and faces added are no longer independent. Second, as we shall see in this paper, in geometric complexes the topology of the underlying metric space plays an important role in the behavior of the complex, whereas in the combinatorial models there is no underlying structure. In this paper we focus on the random Čech complex $\mathcal{C}(n, r)$ generated by taking a random point process of size n, and asserting that $k+1$ points span a k-simplex, if the balls of radius r around them have a nonempty intersection.

Similarly to the study of combinatorial models, we examine the behavior of geometric complexes in the limit when $n \to \infty$ and $r = r(n) \to 0$. The work in [7] established the first rigorous statement about the phase transition describing homological connectivity in random Čech Complexes generated over the d-dimensional flat torus $\mathbb{T}^d$. The main goal of this paper is to extend these results from the flat torus to general d-dimensional compact Riemannian manifolds. Note, that as opposed to the combinatorial models, for the Čech complex, we do not expect homology to become trivial in the limit, but rather to become isomorphic to that of the underlying manifold.

In addition to its mathematical value, the study of random geometric complexes is motivated by applications in engineering and statistics. A rising area of research called *Topological Data Analysis* (TDA), or *Applied Topology*, focuses on utilizing topology in data analysis, machine learning, and network modeling (see [9, 41, 44] for an introduction). The main idea is to use topological features (eg, homology, Euler characteristic, persistent homology) as a "signature" for various types of complex high-dimensional data.

Geometric complexes are used often in TDA to translate data points into a combinatorial-topological space, which in turn can be fed into a software algorithm that calculates its relevant topological properties. It is therefore desired to develop a solid statistical theory for geometric complexes (see eg, [2, 6, 10, 35]), and an imperative part of this effort is to develop its probabilistic foundations (see eg, [1, 4, 12, 37]). Most of the results on random geometric complexes and graphs to date have been studied for point processes in a Euclidean space. In applications, however, it is commonly assumed that the data lie on (or near) a manifold. The results and methods in this paper provide an important gateway to analyzing such cases. In particular the threshold for homological connectivity is related to the problem of "topological inference" [35, 36], where the goal is to recover topological properties of a manifold from a finite sample.

1.2 | Main result

The main result of this paper is the generalized version of Theorem 5.4 in [7]. In the following we assume that M is a d-dimensional compact Riemannian manifold, and $\mathcal{P}_n$ is a homogeneous Poisson process on M with intensity n (see definition in Section 2.3). With no loss of generality, and to shorten notation, we will assume that $\mathrm{Vol}(M) = 1$ as this will only affect the results by an overall scaling constant. In this case, we define $\Lambda := n\omega_d r^d$, where ω_d is the volume of a unit ball in the d-dimensional Euclidean space $\mathbb{R}^d$. For small $r > 0$ this quantity approximates the expected number of points inside a ball of radius r, and can be thought of as measure of density for the process (when $\mathrm{Vol}(M) \neq 1$ this should be $n\omega_d r^d / \mathrm{Vol}(M)$). The following result is the main theorem of this paper, which is the Riemannian analog of Theorem 5.4 in [7], describing the phase transition for homological connectivity in terms of Λ.

Theorem 1.1 *Suppose that as $n \to \infty$, $w(n) \to \infty$. Then, for $1 \le k \le d-1$*

$$\lim_{n\to\infty} \mathbb{P}\left(H_k(C(n,r)) \cong H_k(M)\right) = \begin{cases} 1 & \Lambda = \log n + k \log\log n + w(n), \\ 0 & \Lambda = \log n + (k-2)\log\log n - w(n), \end{cases}$$

Notice that in [7] we further required that $w(n) \gg \log\log\log n$. The generalized proof we use in this paper allows us to avoid this condition. While the result in Theorem 1.1 is not tight, it does provide a good estimate for the exact threshold. For instance, we can deduce that the exact threshold for H_k occurs at $\Lambda = \log n + \alpha_k \log\log n$, with $\alpha_k \in [k-2, k]$ (see the discussion in Section 10). Notice that there are two homology groups which are not covered by this theorem—H_0 and H_d. For H_0, which describes the connectivity of the underlying random geometric graph, it is known that the threshold occurs around $\Lambda = 2^{-d} \log n$.[2] As for H_d, using the Nerve Lemma 2.2 and coverage arguments [16], the threshold can be shown to be $\Lambda = \log n + (d-1)\log\log n$.

Comparing to the results on combinatorial complexes in [25, 30], the main difference here is that the phase transition for connectivity cannot be considered as a special case of the higher dimensional homological connectivity, but rather occurs much earlier. Thus, we observe two main stages—the first one is for connectivity (H_0), and the second one for all other homology groups (H_k, $k \ge 1$). Within the second stage we observe that the different homologies vanish in an orderly fashion. These observations are discussed in detail in [7].

1.3 | Outline of the proof

The proof of Theorem 1.1, has a similar outline to the one in [7], but with considerable geometric adjustments required for the Riemannian case. In fact, the approach we use here for addressing the general setting turns out to be powerful by allowing us to (1) weaken some of the conditions required in [7], and (2) prove many other statements for random Čech complexes in the Riemannian setting.

In the remaining of this section we wish to outline the steps required for the proof. The first of these is based on the following Proposition about the expected Betti numbers of $C(n,r)$, denoted by $\beta_k(r)$. This is a Riemannian analogue of Proposition 5.2 in [7].

Proposition 1.2 *Let $\Lambda \to \infty$ and $r \to 0$, in such a way that $\Lambda r \to 0$. Then, for every $1 \le k \le d-1$ there exist constants a_k, b_k (that depend on the metric g as well), such that*

$$a_k n \Lambda^{k-2} e^{-\Lambda} \le \mathbb{E}\left\{\beta_k(r)\right\} \le \beta_k(M) + b_k n \Lambda^k e^{-\Lambda}.$$

[2]This result is proved only for the torus, but similar techniques as we use in this paper could be used to extend it to the general Riemannian case as well.

Notice that the original version in [7], stated for the flat torus, does not require the assumption that $\Lambda r \to 0$. Even though this condition is necessary to extend the statements to the non-flat case, it does not affect the result in any of the radii of interest for us, since for $\Lambda \gg \log n$ the manifold is covered with high probability which implies that $C(n, r) \simeq M$.

The proof of proposition 1.2 will be split between Section 6 for the upper bound, and Section 7 for the lower bound. The proof of the upper bound makes use of a special variant of Morse theory for the distance function, which we will discuss in detail later. The main idea is to use the Morse inequalities to obtain an upper bound on the Betti numbers $\beta_k(r)$ from the number of index k critical points of an appropriate Morse function. To prove the lower bound, we search for special configurations named "Θ-cycles" (introduced in [7]), which are guaranteed to generate nontrivial k-cycles in homology. Then, by counting these we obtain a lower bound on $\beta_k(r)$. The proof of Proposition 1.2 also covers the upper part of the phase transition in Theorem 1.1. To prove the lower part, in addition to the first moment bound provided in proposition 1.2, we need to use a second moment argument, which will be discussed in Section 8. In Section 9 we will put together all the parts of the proof for Theorem 1.1. Before we get to the proofs, we need to provide various statements that will allow us to carry out the calculations in the general Riemannian setting. This will be done in Sections 2–5.

2 | PRELIMINARIES

2.1 | Homology

In this section we introduce the concept of homology in an intuitive rather than a rigorous way. A comprehensive introduction to the topic can be found in [21, 34]. The reader familiar with the fundamentals of algebraic topology is welcome to skip to Section 2.2.

Let X be a topological space, the *homology* of X is a sequence of abelian groups denoted $\{H_i(X)\}_{i=0}^d$. In this paper, we will assume that homology computed with field coefficients, and then the homology groups $H_*(X)$ are in fact vector spaces. This sequence of vector spaces encapsulates topological information about X in the following way. The basis elements of $H_0(X)$ correspond to the connected components of X, and for $k \geq 1$ the basis elements of $H_k(X)$ correspond to (nontrivial) k-dimensional cycles. Loosely speaking, a nontrivial k-dimensional cycle in a manifold M can be thought of as the boundary of a $k+1$-dimensional solid, such that the interior of the solid is not part of M. The Betti numbers are defined as $\beta_k(X) = \dim(H_k(X))$, namely they are the number of linearly independent k-dimensional cycles in X.

For example, for the d-dimensional sphere $\mathbb{S}^d$ we have $\beta_0(\mathbb{S}^d) = \beta_d(\mathbb{S}^d) = 1$, while $\beta_k(\mathbb{S}^d) = 0$ for $k \neq 0, d$. Another example is the 2-dimensional torus $\mathbb{T}^2$ that has $\beta_0(\mathbb{T}^2) = 1$ (a single component), $\beta_1(\mathbb{T}^d) = 2$ (two essential "loops"), and $\beta_2(\mathbb{T}^2) = 1$ (the boundary of the 3D solid) (see Figure 1).

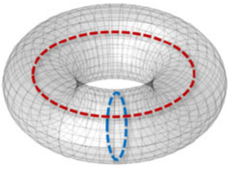

FIGURE 1 The 2D torus $\mathbb{T}^2$. We observe the two loops generating H_1, while the entire 2D surface generates H_2 [Colour figure can be viewed at wileyonlinelibrary.com]

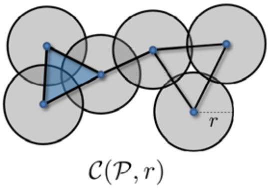

FIGURE 2 A Čech complex generated by a set of points in $\mathbb{R}^2$. The complex has six vertices (0-simplexes), seven edges (1-simplexes) and a single triangle (a 2-simplex) [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | The Čech complex

The Čech complex we study in this paper, is an abstract simplicial complex constructed from a finite set of points in a metric space in the following way.

Definition 2.1. Let $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$ be a collection of points in a metric space, and let $r > 0$ and let $B_r(x)$ be the ball of radius r around x. The Čech complex $C_r(\mathcal{P})$ is constructed as follows:

1. The 0-simplexes (vertices) are the points in $\mathcal{P}$.
2. A k-simplex $[p_{i_0}, \dots, p_{i_k}]$ is in $C_r(\mathcal{P})$ if $\bigcap_{j=0}^{k} B_r(p_{i_j}) \neq \emptyset$.

Associated with the Čech complex $C_r(\mathcal{P})$ is the union of balls used to generate it (in the underlying metric space), which we define as

$$B_r(\mathcal{P}) = \bigcup_{p \in \mathcal{P}} B_r(p). \tag{2.1}$$

The spaces $C_r(\mathcal{P})$ and $B_r(\mathcal{P})$ are of a completely different nature. Nevertheless, the following lemma claims that they are very similar in the topological sense. This lemma is a special case of a more general topological statement originated in [8] and commonly referred to as the "Nerve Lemma."

Lemma 2.2 (The Nerve Lemma) *Let $C_r(\mathcal{P})$ and $B_r(\mathcal{P})$ as defined above. If for every $p_{i_1}, \dots, p_{i_k}$ the intersection $B_r(p_{i_1}) \cap \cdots \cap B_r(p_{i_k})$ is either empty or contractible (homotopy equivalent to a point), then $C_r(\mathcal{P}) \simeq B_r(\mathcal{P})$, and in particular,*

$$H_k(C_r(\mathcal{P})) \cong H_k(B_r(\mathcal{P})), \quad \forall k \geq 0.$$

Consequently, for sufficiently small r, we will sometimes be using $B_r(\mathcal{P})$ to make statements about $C_r(\mathcal{P})$. This will be very useful especially when coverage arguments are available. Note that in Figure 2 indeed both $C_r(\mathcal{P})$ and $B_r(\mathcal{P})$ have a single connected component and a single hole.

2.3 | The Poisson process

In order to make a Čech complex $C_r(\mathcal{P})$ random, we generate the point set $\mathcal{P}$ at random. In this paper we will use the model of a homogeneous Poisson process. Let M be a compact Riemannian manifold, and let $X_1, X_2, \dots,$ be a sequence of i.i.d. (independent and identically distributed) random variables, distributed uniformly on M with respect to the Riemannian volume measure. Let $N \sim \text{Poisson}(n)$ be a Poisson random variable, independent of the X_i-s, and define the spatial Poisson process $\mathcal{P}_n$ as

$$\mathcal{P}_n = \{X_1, X_2, \dots, X_N\}.$$

In other words, we generate a random number N and then pick the first N random points generated on M. Notice that we have $\mathbb{E}\{|\mathcal{P}_n|\} = n$, so while the number of points is random its expected value is n. For every subset $A \subset M$ we define $\mathcal{P}_n(A) = |\mathcal{P}_n \cap A|$, that is, the number of points lying in A. The process $\mathcal{P}_n$ can be equivalently defined as the process that satisfies the following properties:

1. For every subset $A \subset M$ the number of points in A has a Poisson distribution. More specifically: $\mathcal{P}_n(A) \sim \text{Poisson}\,(n\,\text{Vol}(A)/\,\text{Vol}(M))$.

2. If $A, B \subset M$ are two disjoint sets ($A \cap B = \emptyset$) then the variables $\mathcal{P}_n(A), \mathcal{P}_n(B)$ are independent.

The last property of the Poisson process is known as "spatial independence" and it is the main reason why the Poisson process such a convenient model to analyze.

In this paper we study the random Čech complex $C(n, r) := C_r(\mathcal{P}_n)$. To shorten notation, from here on we will use C_r to refer to $C(n, r)$. As mentioned earlier, throughout the paper we will assume that $\text{Vol}(M) = 1$, to simplify the calculations.

3 | RIEMANNIAN GEOMETRY INGREDIENTS

In this section we wish to provide the reader with a brief review of Riemannian geometry, which will be used throughout this paper. There are many good textbooks on the subject, for a comprehensive introduction see for example [11,39], or [28].

3.1 | A quick introduction to Riemannian geometry

A Riemannian manifold is a pair (M, g), where M is a smooth manifold and a smoothly varying metric $g : T_pM \times T_pM \to \mathbb{R}$, where T_pM is the tangent space to M at a point p. The metric yields a smoothly varying inner product in the tangent space T_pM at any point $p \in M$. Hence, it can be used to define the norm of a tangent vector, as well as angles between tangent vectors at the same point. Using the notion of norm $|v| = \sqrt{g_p(v, v)}$ of a vector $v \in T_pM$ one can define the length $\ell(\gamma)$ of a path $\gamma : I \subset \mathbb{R} \to M$ by simply integrating its velocity, that is:

$$\ell(\gamma) = \int_I |\dot{\gamma}(t)|\ dt,$$

where $\dot{\gamma}(t) \in T_{\gamma(t)}M$ denotes the tangent vector to γ at $\gamma(t)$, for $t \in I$. Given two points $p, q \in M$ one can define the distance $\text{dist}(p, q)$ as the infimum of the length over all paths connecting p and q. A Riemannian manifold is called *complete* if for any two points $p, q \in M$ there is a curve connecting them which minimizes the distance, that is, the infimum length is achieved. These minimizing curves are called *geodesics*. Notice that every compact manifold is also complete, and moreover, by the theorem of Hopf and Rinow [22] it is also a complete metric space. Hence, the definitions of the Poisson process and the Čech complex described above are valid.

The local invariants of a Riemannian metric g are encoded in its curvature. We shall now introduce this in a way that will be useful for us later. Given a point $p \in M$ one can consider the geodesics that pass through p at time $t = 0$ with velocity vector as a given vector $v \in T_pM$. The geodesic equations are a system of second order differential equations and a simple application of Picard's existence and uniqueness theorem shows that these geodesic are unique, and we denote them by $\gamma_v(t)$. We can then consider a map $\exp_p : T_pM \to M$, called the exponential map at p, which assigns to each vector $v \in T_pM$ the position at which the unique geodesic starting at p with velocity v is at time 1, that is, $\exp_p(v) = \gamma_v(1)$. The derivative of $\exp_p$ at the origin in T_pM is the identity and so, by the inverse function theorem, $\exp_p$ is a local diffeomorphism of a small ball around $0 \in T_pM$ to a small neighborhood of $p \in M$. As a consequence, one can use the exponential map to define local coordinates around p. For instance, fixing an orthonormal basis for T_pM one obtains coordinates on T_pM, which can be regarded as local coordinates $(x^1, \dots, x^d)$ around p. These are the so called *geodesic normal coordinates* and have the following properties. The point p corresponds in this coordinates to the point $(0, \dots, 0)$. Any geodesic through p corresponds to a straight line passing through the origin $(0, \dots, 0)$ and intersects the spheres $S_r(p) = \{(x^1, \dots, x^d)\ |(x^1)^2 + \cdots + (x^d)^2 = r^2\}$ orthogonally. In these coordinates, the metric can be written as $g = g_{ij}dx^i \otimes dx^j$, with

$$g_{ij} = \delta_{ij} + \frac{1}{3}R_{iklj}x^k x^l + O(|x|^3), \tag{3.1}$$

where δ_{ij} is the Kronecker delta. The second order terms, that is, R_{iklj}, in the Taylor expansion above, form the *Riemann curvature tensor* at p. We note that here, as well as later in the paper, we use Einstein's notation for summation, where by a_ib^i we mean the sum $\sum_i a_ib^i$. We refer the reader to any textbook on Riemannian geometry for all that was described above. In what follows we shall give a self contained exposition of all the geometric input needed. This will be mostly derived in the rest of this section and the appendix.

3.2 | Notation

In this paper we restrict to the case when (M, g) is a compact Riemannian manifold. For convenience, we will sometimes use $\langle \cdot, \cdot \rangle_p$ to denote the inner product of tangent vectors at a point $p \in M$ (instead of g). Next, for $\mathcal{P} \subset M$ and $r > 0$, we define the following:

- $B_r(p)$ = the closed ball of radius r around p.
- $S_r(p)$ = the sphere of radius r around p.
- $B_r(\mathcal{P}) = \bigcup_{p\in\mathcal{P}} B_r(p)$ - the union of balls.
- $B_r^{\cap}(\mathcal{P}) = \bigcap_{p\in\mathcal{P}} B_r(p)$ - the intersection of balls.

3.3 | Approximations of Riemannian volumes

Throughout the proofs in this paper, we will need to approximate volumes for subsets of Riemannian manifolds, and compare them to their Euclidean counterparts. In this section we introduce the Riemannian normal coordinates and use them to provide the main inequalities that we will use for that purpose. The proofs for the statements in this section appear in Appendix A. For an exposition of the background leading to these see for example [28, 39]. A particularly nice exposition of Riemannian normal coordinates can be found in [40].

Let $p \in M$ and $(x^1, \dots, x^d)$ be geodesic normal coordinates centered at p. Then, in the domain of definition of these coordinates the metric can be written as $g = g_{ij}dx^i \otimes dx^j$, with g_{ij} as in (3.1). Then, there is a canonical measure on M induced by the Riemannian density which is given by

$$\left|\mathrm{dvol}_g\right| = \sqrt{\left|\det(g_{ij})\right|}\,\left|\mathrm{dvol}_{g_E}\right|,$$

where $\left|\mathrm{dvol}_{g_E}\right| = \left|dx^1 \wedge \cdots \wedge dx^d\right| = dx_1 \dots dx_n$ is the density associated with the Euclidean metric with $g_E = \delta_{ij}dx^i \otimes dx^j$. For more on densities in Riemannian manifolds see, for example, in [29, pp.

304-306]. In this section we will prove various inequalities that relate geometric quantities associated with g to those associated with g_E. The first ingredient we need is the computation of the Riemannian density $\sqrt{|\det(g_{ij})|}$ using Equation (3.1),

$$\sqrt{|\det(g_{ij})|} = 1 - \frac{Ric_{ij}}{3} x^i x^j + O(|x|^3), \tag{3.2}$$

where $Ric_{ij} = -\sum_k R_{ikkj}$ is known as the *Ricci curvature tensor* at p. Using this expression for the Riemannian density it is easy to show that the volume of a normal ball $B_r(p) = \{(x^1, \dots, x^d) \mid (x^1)^2 + \cdots + (x^d)^2 \leq r^2\}$ can be written as:

$$\mathrm{Vol}(B_r(p)) = \omega_d r^d \left(1 - \frac{s(p)}{6(d+2)} r^2 + O(r^3)\right), \tag{3.3}$$

where ω_d is the volume of an Euclidean d-dimensional ball of radius 1, and $s(p)$ denotes the *scalar curvature* $s(p) = \sum_i Ric_{ii}$ at p. Similarly, one can compute the volume of a normal sphere to be

$$\mathrm{Vol}(S_r(p)) = d\omega_d r^{d-1} \left(1 - \frac{s(p)}{6d} r^2 + O(r^3)\right). \tag{3.4}$$

In the following, we will make use of Equations (3.3) and (3.4) in order to get bounds on these volumes for small values of r. Notice that we can write $|\mathrm{dvol}_g|$ in polar coordinates as $|\mathrm{dvol}_g| = dr\,|\mathrm{dvol}_{S_r(p)}|$, where $\mathrm{dvol}_{S_r(p)}$ denotes the volume form of the induced metric on the $S_r(p)$, and by $|\mathrm{dvol}_{S_r(p)}|$ we simply mean that we regard it as a density. In the Euclidean case, $\mathrm{dvol}_{g_E} = r^{d-1} dr \wedge \mathrm{dvol}_{\mathbb{S}^{d-1}}$, where $\mathbb{S}^{d-1}$ is the unit round sphere. The following result compares $\mathrm{dvol}_{S_r(p)}$ with $r^{d-1}\,\mathrm{dvol}_{\mathbb{S}^{d-1}}$, for a given Riemannian metric g.

Lemma 3.1 *Let (M, g) be a compact Riemannian manifold. Denote by $|Ric_p| = \sup_{v \in T_pM \setminus 0} \frac{|Ric(v,v)|}{|v|^2}$ the norm of the Ricci tensor at $p \in M$. Then, for any $\nu > 0$ there exists $r_\nu > 0$ such that for all $r \leq r_\nu$ and for all $p \in M$, we have*

$$r^{d-1}\left(1 - \frac{|Ric_p| + \nu}{3} \cdot r^2\right) |\mathrm{dvol}_{\mathbb{S}^{d-1}}| \leq |\mathrm{dvol}_{S_r(p)}| \leq r^{d-1}\left(1 + \frac{|Ric_p| + \nu}{3} \cdot r^2\right) |\mathrm{dvol}_{\mathbb{S}^{d-1}}|,$$

on $B_r(p)$. Moreover, r_ν depends continuously on ν.

Corollary 3.2 *For $\nu > 0$, let $s_{min}(\nu) = \inf_{p\in M} \frac{s(c)}{6(d+2)} - \nu$ and $s_{max}(\nu) = \sup_{p \in M} \frac{s(c)}{6(d+2)} + \nu$. Then, for all $\nu > 0$ there exists $r_\nu > 0$, such that for all $r \leq r_\nu$*

$$\omega_d r^d \left(1 - s_{max}(\nu) r^2\right) \leq \mathrm{Vol}(B_r(p)) \leq \omega_d r^d \left(1 - s_{min}(\nu) r^2\right).$$

Lemma 3.3 *Let (M, g) be a compact Riemannian manifold. Then, for all $\nu > 0$ there is $r_\nu > 0$, such that for all $p \in M$ and $r \leq r_\nu$, we have*

$$(1 - \nu r^2)\,|\mathrm{dvol}_{g_E}| \leq |\mathrm{dvol}_g| \leq (1 + \nu r^2)\,|\mathrm{dvol}_{g_E}|,$$

on $B_r(p)$, for all $p \in M$.

Note, that we can take r_ν in both lemmas above to be the same (by taking the minimum of the two). The following lemma is an immediate consequence of the above. The last comparison we will need is between the union of Riemannian balls and Euclidean ones centered at the same points. The statement requires a little bit of notation. For small $r > 0$ and p_1, p_2 at distance at most $2r$ from each other, we let p be the midpoint of the minimizing geodesic connecting p_1 to p_2. Next, fix the Riemannian normal coordinates $(x^1, \ldots, x^d)$ centered at p, which induce a Euclidean metric $g_E = \delta_{ij} dx^i \otimes dx^j$. In the domain of definition of g_E we will use $B_s^E(q)$ to denote the s-ball centered at q, where s is measured using the metric g_E.

Lemma 3.4 *Let (M, g) be a compact Riemannian manifold. Then, there exist $\nu > 0$ and $r_\nu > 0$ such that for every $r < r_\nu$ and any two points p_1, p_2 with* $\mathrm{dist}(p_1, p_2) < 2r$ *we have,*

$$\left(B^E_{(1-\nu r)r}(p_1) \cup B^E_{(1-\nu r)r}(p_2)\right) \subset (B_r(p_1) \cup B_r(p_2)) \subset \left(B^E_{(1+\nu r)r}(p_1) \cup B^E_{C(1+\nu r)r}(p_2)\right).$$

Note that the values of r_ν in the previous two lemmas can be chosen independently of each other.

4 | MORSE THEORY FOR THE DISTANCE FUNCTION

In this section we use the results in [17] to develop the framework which will allow us to justify various Morse theoretic arguments later. For an introduction to Morse theory see [33]. Briefly, developing a Morse theoretic framework will allow us to study the homology of the Čech complex by examining critical points for the corresponding distance function (defined below).

Recall that for $x, y \in M$ we define $\mathrm{dist}(x, y)$ to be the geodesic distance between x and y with respect to the metric g. For $p \in M$ we define $\rho_p : M \to \mathbb{R}_0^+$ to be the function $\rho_p(x) := \mathrm{dist}(p, x)$. If $\mathcal{P} \subset M$ is a finite set, we define the distance function $\rho_{\mathcal{P}} : M \to \mathbb{R}_0^+$ as:

$$\rho_{\mathcal{P}}(x) := \min_{p \in \mathcal{P}} \rho_p(x).$$

We shall now establish Morse theory for the function $\rho_{\mathcal{P}}$, using the framework of Morse theory for min-type functions developed in [17]. We start by showing that for any finite set $\mathcal{P}$ the function $\rho_{\mathcal{P}}^2$ is a *Morse min-type function*, namely that at every point we can write $\rho_{\mathcal{P}}^2$ as a minimum of finitely many smooth Morse functions.

Lemma 4.1 *Given a compact Riemannian manifold (M, g) there exists $r_{\mathrm{mt}} > 0$ such that for every $\mathcal{P} \subset M$ the function $\rho_{\mathcal{P}}^2$ is a Morse min-type function on $B_{r_{\mathrm{mt}}}(\mathcal{P})$.*

Proof For any $p \in M$, there exists $r_p > 0$ such that the function $\rho_p^2(\cdot)$ is smooth, Morse, and strictly convex on $B_{r_p}(p)$. Since the metric g is smooth we can choose r_p continuously in $p \in M$, and since M is compact we can define $r_{\mathrm{mt}} := \min_{p \in M} r_p > 0$. The result follows. ■

Now that we established that $\rho_{\mathcal{P}}^2$ is a Morse min-type function, we want to explore its critical points. Similarly to the Euclidean case (cf. [3]), critical points of index k are generated by subsets $\mathcal{Y} \subset \mathcal{P}$ containing $k + 1$ points, and are located at the “center” of the set. While in the Euclidean case the center of $\mathcal{Y}$ is simply taken to be the center of the unique $(k-1)$-sphere that contains $\mathcal{Y}$, in the general Riemannian case we need to carefully define the notion of a center.

4.1 | The center and radius of a set

Let $\mathcal{Y}$ be a finite subset of M and define:

$$E(\mathcal{Y}) := \{x \in M \mid \rho_{p_1}(x) = \rho_{p_2}(x) = \cdots = \rho_{p_k}(x))\},$$
$$E_r(\mathcal{Y}) := E(\mathcal{Y}) \cap B_r(\mathcal{Y}).$$

In other words, $E(\mathcal{Y})$ is the set of all points that are equidistant from $\mathcal{Y}$, and $E_r(\mathcal{Y})$ is its restriction to a small neighborhood around $\mathcal{Y}$. We say that a subset $\mathcal{Y}$ is *generic* if $E(\mathcal{Y}) \neq \emptyset$, and the zero level sets of the $k-1$ functions $\rho^2_{p_i}(\cdot) - \rho^2_{p_1}(\cdot)$ intersect transversely[3] in $B^{\cap}_{r_{\mathrm{mt}}}(\mathcal{Y})$. This guarantees that for a generic set the intersection $E_{r_{\mathrm{mt}}}(\mathcal{Y})$ is a smooth submanifold of dimension $d-k+1$.

Lemma 4.2 *There exists a positive $r_{\max} < r_{mt}$ such that if $\mathcal{Y} \subset M$ is a finite generic subset with $E_{r_{\max}}(\mathcal{Y}) \neq \emptyset$, then there exists a unique point $c(\mathcal{Y}) \in M$ such that for all $p \in \mathcal{Y}$,*

$$\rho_p(c(\mathcal{Y})) = \inf_{x \in E(\mathcal{Y})} \rho_{\mathcal{Y}}(x). \tag{4.1}$$

In that case, we define $\rho(\mathcal{Y}) := \rho_{\mathcal{Y}}(c(\mathcal{Y}))$, and we will refer to $c(\mathcal{Y})$ and $\rho(\mathcal{Y})$ as the center and radius of the set $\mathcal{Y}$, respectively.

Proof Since $\mathcal{Y} \subset M$ is generic, $E_{r_{\mathrm{mt}}}(\mathcal{Y})$ is a nonempty smooth submanifold of dimension $d-k+1$. Moreover, in $E_{r_{\mathrm{mt}}}(\mathcal{Y})$ we know that $\rho_{\mathcal{Y}} \le r_{\mathrm{mt}}$, with equality attained only on the boundary. Thus, as $E_{r_{mt}}(\mathcal{Y})$ is bounded and $\rho(\mathcal{Y})$ is continuous, there is a minimum point for $\rho(\mathcal{Y})$ in $E_{r_{\mathrm{mt}}}(\mathcal{Y})$, denoted $c(\mathcal{Y})$. Moreover, since in $E(\mathcal{Y}) \backslash E_{r_{\mathrm{mt}}}(\mathcal{Y})$ we have $\rho_{\mathcal{Y}} > r_{\mathrm{mt}}$, then $c(\mathcal{Y})$ is actually a global minimum of $\rho_{\mathcal{Y}}|_{E(\mathcal{Y})}$. In addition, it follows from the definition of $E(\mathcal{Y})$ that for any $p \in \mathcal{Y}$ we have $\rho_{\mathcal{Y}}|_{E(\mathcal{Y})} \equiv \rho_p|_{E(\mathcal{Y})}$, and so Equation (4.1) holds.
To prove that $c(\mathcal{Y})$ is unique, we use Lemma .1 in Appendix C. This lemma shows that for $r' \le r_{\mathrm{mt}}$ sufficiently small $E_{r'}(\mathcal{Y})$ approaches a totally geodesic submanifold, as the Riemannian distance functions approach the Euclidean ones, and therefore the restriction of ρ^2_p to $E_{r'}(\mathcal{Y})$ is strictly convex. As a consequence, and given that the minimums of ρ_p and ρ^2_p are the same, ρ_p has a unique minimum in $E_{r'}(\mathcal{Y})$. Since $\rho_{\mathcal{Y}}|_{E(\mathcal{Y})} \equiv \rho_p|_{E(\mathcal{Y})}$ (for any $p \in \mathcal{Y}$), and since $\rho_{\mathcal{Y}} > r'$ on $E(\mathcal{Y}) \backslash E_{r'}(\mathcal{Y})$, we conclude that $c(\mathcal{Y})$ is the unique minimum of $\rho_{\mathcal{Y}}$ in $E(\mathcal{Y})$. Since M is compact so is M^k, for $k \in \mathbb{N}$, hence we can minimize the value of r' over all k-tuples $\mathcal{Y} \subset M$ as in the statement. From here on we will define $r_{\max}$ to be this minimum value of r'. ∎

Remark 4.3. Given $r_{\max}$ as in Lemma 4.2 we have that:

- For any $\mathcal{P}$, $\rho^2_{\mathcal{P}}$ is a Morse min-type function on $B_{r_{\max}}(\mathcal{P})$.
- For every $\mathcal{Y} \subset \mathcal{P}$ with $E_{r_{\max}}(\mathcal{Y})$ nonempty, $c(\mathcal{Y})$ is uniquely defined.
- For $r \le r_{max}$, the exponential map $\exp_p$ is always defined for $v \in T_pM$ with $|v| \le r$.

Remark 4.4. In the 1920s Cartan used a different way to associate a center to a finite set of points in a negatively curved Riemannian manifold, called the *center of mass*. This definition could be adapted to our context as follows. By the definition of $r_{\max}$, for each $p \in \mathcal{Y}$ the function ρ_p is strictly convex in $E_{r_{max}}(\mathcal{Y})$. Thus, the function $\max_{p \in \mathcal{Y}} \rho_p(\cdot)$ is strictly convex in $B^{\cap}_{r_{max}}(\mathcal{Y})$, and achieves a unique minimum there. This minimum is defined to be the center of mass $\mathrm{cm}(\mathcal{Y})$. We point out that in general, the center of mass need not be the same as a center $c(\mathcal{Y})$ we defined above, and does not serve our purposes.

[3]Two submanifolds N_1 and N_2 of M are said to intersect transverly at a point p, if $T_pN_1 + T_pN_2 = T_pM$. They are said to intersect transversely if they intersect transversely at all points $p \in N_1 \cap N_2$.

4.2 | Critical points for the distance function

In classical Morse theory, the critical points of a functions are those points where the gradient ∇f is zero. The index of a critical point is the number of negative eigenvalues of the Hessian H_f, which can be thought of as the number of independent directions we can leave the critical point along which the function values will be decreasing. Consequently, critical points of index 0 are the minima, and if f is defined over a d-dimensional manifold, then critical points of index d correspond to the maxima. We now wish to investigate the critical locus of $\rho^2_{\mathcal{P}}$ in an analog fashion. However, since $\rho^2_{\mathcal{P}}$ is non differentiable, definitions of critical points have to be adjusted. We start by noticing that $\rho^2_{\mathcal{P}}$ is nonnegative and vanishes precisely at the points $p \in \mathcal{P}$. Thus, $\mathcal{P}$ is the set of minima, or index 0 critical points, of $\rho^2_{\mathcal{P}}$. To find the critical points of higher index, we will use the results of [17], which require the following definition.

Definition 4.5. Let $\mathcal{Y} = \{y_1, \dots, y_k\} \subset M$ and $p \in M$, then we define the polytope in $\Delta(\mathcal{Y}) \subset T_pM$, given by

$$\Delta(\mathcal{Y}) := \mathrm{conv}(\{\nabla \rho^2_{y_i}(p)\}_{i=1}^k),$$

where conv$(\cdot)$ means convex hull, and $\nabla \rho^2_{y_i}$ is the gradient of $\rho^2_{y_i}$, defined as the unique vector field such that $g(\nabla \rho^2_{y_i}, v) = d\rho^2_{y_i}(v)$, for any vector field v.

The critical points of $\rho^2_{\mathcal{P}}$ are then characterized by the following Proposition.

Proposition 4.6 *Let $c \in M$ be a critical point of $\rho^2_{\mathcal{P}}$. Then, c has index k if and only if there exists a set $\mathcal{Y} \subset \mathcal{P}$ of $k+1$ points such that*:

1. $c(\mathcal{Y}) = c$, *where $c(\mathcal{Y})$ is the center of $\mathcal{Y}$* (*see Lemma* 4.2),
2. $\Delta(\mathcal{Y}) \in T_cM$ *contains the origin* $0 \in T_cM$,
3. $B(\mathcal{Y}) \cap \mathcal{P} = \mathcal{Y}$.

Figure 3 depicts the conditions in the last proposition.

Proof Let c be a critical point of the min-type function $\rho^2_{\mathcal{P}}$ and suppose that the minimal representation of $\rho^2_{\mathcal{P}}$ in a neighborhood of c is of the form

$$\rho^2_{\mathcal{P}}(\cdot) = \min_{i=1,\dots,k+1} \rho^2_{y_i}(\cdot).$$

Denoting $\mathcal{Y} = \{y_1, \dots, y_{k+1}\}$, and given that the representation above is minimal, at c we must have $\rho_{y_1}(c) = \cdots = \rho_{y_{k+1}}(c)$. In addition, using the definition of critical points of a min-type function from [17], we have that c is a critical point of each ρ_{y_i} restricted to the smooth $d-(k-1)$ dimensional submanifold:

$$E(\mathcal{Y}) \cap B_{r_{\max}}(c) = \{p \in B_{r_{\max}}(c) \mid \rho_{y_1}(p) = \cdots = \rho_{y_{k+1}}(p)\}.$$

However, by the definition of r_{max}, each distance function is strictly convex on $E_{r_{\max}}(\mathcal{Y})$, and so has a unique critical point (a minimum) which is $c(\mathcal{Y})$ by definition. Hence, by [17] the index I_c of $c = c(\mathcal{Y})$ is

$$I_c(\rho) = k + I_c(\rho|_{E(\mathcal{Y})}) = k.$$

The second and third conditions are now immediate from the definition of a critical point of a min-type function in [17]. Conversely, if the conditions in the statement hold, then it immediately follows from

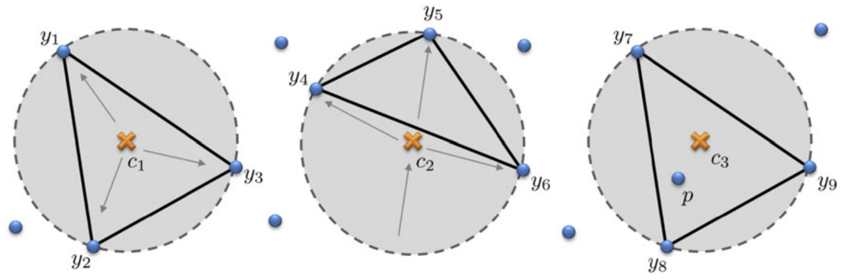

FIGURE 3 Critical points of index 2 in $\mathbb{R}^2$. The blue points are the points of $\mathcal{P}$. We examine three subsets of $\mathcal{P}$: $\mathcal{Y}_1 = \{y_1, y_2, y_3\}$, $\mathcal{Y}_2 = \{y_4, y_5, y_6\}$, and $\mathcal{Y}_3 = \{y_7, y_8, y_9\}$. The orange x's are the centers $c(\mathcal{Y}_i) = c_i$. The shaded balls are $B(\mathcal{Y}_i)$, and the interior of the triangles are $\Delta(\mathcal{Y}_i)$. The arrows represent the flow direction. For $\mathcal{Y}_1$ both conditions in Proposition 4.6 hold and therefore c_1 is a critical point. However, for $\mathcal{Y}_2$ condition (2) does not hold, and for $\mathcal{Y}_3$ condition (3) fails, therefore c_2 and c_3 are not critical points [Colour figure can be viewed at wileyonlinelibrary.com]

the definition of a critical point of a Morse min type function in [17] that c is an index k critical point. ∎

Remark 4.7. If $c \in M$ is a critical point of $\rho^2_{\mathcal{P}}$ with $\mathcal{Y} \subset \mathcal{P}$, and $c(\mathcal{Y}) = c$ as in proposition 4.6. Then, c is not only a center $c(\mathcal{Y})$ - it is also the center of mass $\mathrm{cm}(\mathcal{Y})$. However, this only true for critical points.

Remark 4.8. In [17], the gradient of $\rho^2_{\mathcal{P}}$ at a point $p \in M$ is defined to be $\Delta(\mathcal{Y})$, where $\mathcal{Y} = \{y_1, \dots, y_k\}$ forms a minimal representation of $\rho_{\mathcal{P}}$ in a neighborhood of p, ie, $\rho^2_{\mathcal{P}}(\cdot) = \min_i \rho^2_{y_i}(\cdot)$. This definition of a critical point was first used (in an implicit way) in the work of Grove and Shiohama in [20] and also in that of Gromov in [19]. One could have equally defined $\Delta^*(\mathcal{Y}) \in T^*_p M$ as the convex hull of the 1-forms $\{d\rho_{y_i}(\cdot)\}_{i=1}^{k+1}$. Then, the conditions 2 and 3 in the previous proposition have trivial analogues for $\Delta^*(\mathcal{Y})$. We have chosen to work with $\Delta(\mathcal{Y})$ rather than its dual in order to more closely follow the definitions in [17].

Now that we have defined the critical points and their index, Morse theory (and in particular the inequalities discussed in Section 4.3) for $\rho^2_{\mathcal{P}}$ follows from [17]. In particular, we can deduce the role of these critical points in generating the homology of the sublevel sets,

$$\rho^{-1}_{\mathcal{P}}(0, r) = B_r(\mathcal{P}),$$

which are homotopy equivalent to $C_r(\mathcal{P})$. While we have considered $\rho^2_{\mathcal{P}}$, since the sublevel sets of both $\rho_{\mathcal{P}}$ and $\rho^2_{\mathcal{P}}$ are the same union of balls (just at different levels), everything we discussed applies the same way to $\rho_{\mathcal{P}}$. Thus, from here own we shall refer about critical points and Morse theory for $\rho_{\mathcal{P}}$, referring to the definitions we provided for $\rho^2_{\mathcal{P}}$.

4.3 | Morse inequalities

In proving the main results of this paper, we will make use Morse inequalities. In this section we present the version of these inequalities for Morse min-type functions which we shall apply to $\rho^2_{\mathcal{P}}$. The proof uses a standard argument in algebraic topology and the reader unfamiliar with the basics of the subject is welcome to skip to remark 4.10.

Let f be a min-type Morse function, and define:

$$C_k(a,b) := \#\text{ critical points } c \text{ of index } k, \text{ such that } a < f(c) \le b$$

The following lemma provides a slightly less standard version of the Morse inequalities.

Lemma 4.9 *Let $f : M \to \mathbb{R}$ be a Morse min-type function. For $a \in \mathbb{R}$ denote $M_a := f^{-1}(-\infty, a]$. Then, for all $k \in \mathbb{N}$ the following inequalities hold*

$$\beta_k(M_a) - \beta_k(M) \le C_{k+1}(a, +\infty). \tag{4.2}$$

Proof Using the approximation results of [17] there is no loss of generality to restrict to the case when f is a standard Morse-Smale function. Then, we consider the long exact sequence in homology for the pair (M, M_a), namely

$$\cdots \to H_{k+1}(M) \to H_{k+1}(M, M_a) \xrightarrow{\delta} H_k(M_a) \xrightarrow{i} H_k(M) \to \cdots$$

The exactness of this sequence yields

$$\begin{aligned} \beta_k(M_a) &= \dim H_k(M_a) = \dim(\mathrm{Im}(i)) + \dim(\ker(i)) \\ &= \dim(\mathrm{Im}(i)) + \dim(\mathrm{Im}(\delta)) \\ &\le \dim(H_k(M)) + \dim(H_{k+1}(M, M_a)). \end{aligned}$$

This proves that $\beta_k(M_a) - \beta_k(M) \le \beta_{k+1}(M, M_a)$. The statement then follows from the standard Morse inequalities applied to $f|_{M\setminus M_a}$, stating that $\beta_{k+1}(M, M_a) \le C_{k+1}(a, +\infty)$. ■

Remark 4.10. Applying Lemma 4.9 to the function $\rho^2_{\mathcal{P}}$ we have $M_r = \rho^{-1}_{\mathcal{P}}(-\infty, r] = B_r(\mathcal{P})$. Then, using the Nerve Lemma 2.2 we have $\beta_k(M_r) = \beta_k(C_r(\mathcal{P}))$ and so

$$\beta_k(C_r(\mathcal{P})) - \beta_k(M) \le C_{k+1}(r, +\infty), \tag{4.3}$$

where $C_{k+1}(r, \infty)$ denotes the number of index $k+1$ critical points c of $\rho^2_{\mathcal{P}}$ with $\rho^2_{\mathcal{P}}(c) > r^2$.

5 | CHANGE OF VARIABLES (BLASCHKE-PETKANTSCHIN-TYPE FORMULA)

To prove the main results in this paper we will need to evaluate probabilities related to the existence of certain critical points. These computations often result in complicated integral formulae. In this section we present a change-of-variables technique that simplifies these calculations significantly. This technique can be thought of as a Riemannian generalization of the Blaschke-Petkantschin formula that appeared in Edelsbrunner and coworkers [13, 32]. We start by introducing some useful notation.

- $Gr(k, d)$ denotes the Grassmannian of k-planes in a d-dimensional real vector space. When we pick local coordinates, there is a fixed isomorphism $\mathbb{R}^d \cong T_cM$. Then, we may refer to $Gr(k, T_pM)$ when we want to emphasize the fact that we are parametrizing k-planes in the tangent space to M at p.

- Pick normal coordinates $(x^1, \dots, x^d)$ centered at $p \in M$ associated with an orthonormal frame $\{(\partial_{x^i})_c\}_{i=1}^d$ of T_pM. For $V \in Gr(k, T_pM)$ and $r > 0$, we define $S_r(V) \subset M$ to be the image under the exponential map of $(k-1)$-dimensional sphere of radius r centered at the origin and equipped with the induced metric from g. In addition, we denote by $\mathbb{S}_r(V)$ the sphere of radius r equipped with the Euclidean metric $g_E = \delta_{ij} dx^i \otimes dx^j$. Notice that as a manifolds $S_r(V) = \mathbb{S}_r(V)$ by the Gauss lemma, the goal of the notation is to distinguish the metrics with which they are equipped.
- We define $\mathbb{1}_r : 2^M \to \{0, 1\}$ as,

$$\mathbb{1}_r(\mathcal{Y}) := \mathbb{1}\left\{E_{r_{\max}}(\mathcal{Y}) \neq \emptyset \text{ and } \rho(\mathcal{Y}) \leq r\right\},$$

  where $\mathbb{1}\{\cdot\}$ stands for the indicator function.
- Subsets of M will be referred to as either $\mathcal{Y} \subset M$ or $\mathbf{y} \in M^k$, depending on the context. To keep notation simple, we will allow functions defined on 2^M (such as $c(\cdot), \rho(\cdot), \mathbb{1}_r(\cdot)$), to be applied to $\mathbf{y} \in M^k$ as well.
- Given $\mathcal{Y} = \{y_1, \dots, y_k\} \subset M$ with well defined center $c = c(\mathcal{Y})$ we let $v_i \in T_cM$ be such that $y_i = \exp_c(v_i)$, in particular $|v_i| = \rho(\mathcal{Y})$. Let $(x^1, \dots, x^d)$ be normal coordinates such that the $(\partial_{x^i})_c$ form an orthonormal frame of $(T_cM, \langle\cdot,\cdot\rangle)$. Suppose the k-vectors v_i lie in a $(k-1)$-dimensional vector subspace $V \subset T_cM$ (this will be shown to be the case in the proof of Lemma 5.1), then they generate a $(k-1)$-dimensional parallelogram. For $u \in \mathbb{R}^+$, let $\Upsilon_u(\mathbf{v})$, for $\mathbf{v} = (v_1, \dots, v_k)$, be the $(k-1)$-volume of the parallelogram generated by the $u\frac{v_i}{|v_i|}$ in T_cM. Moreover, notice that this is a homogeneous function of degree $(k-1)$ in u.

The following lemma will be useful for us in the proofs of the main results. The main idea is that instead of integrating over tuples $\mathbf{y} \in M^{k+1}$, we can perform a change of variables so that we first choose a $(k-1)$-sphere on which the points will be placed (with center c and radius u), and then we place $k+1$ points on that sphere.

Lemma 5.1 *Let $\mathcal{P}$ and $r_{\max}$ be as in the discussion above, and let $r < r_{\max}$. Then, there is an invariant measure $d\mu_{k,d}$ on $Gr(k, d)$, such that for any smooth $f : M^{k+1} \to \mathbb{R}$, the integral $\int_{M^{k+1}} f(\mathbf{y})\mathbb{1}_r(\mathbf{y}) \left|d\mathrm{vol}_g(\mathbf{y})\right|$ can be written as:*

$$\int_M \left|d\mathrm{vol}_g(c)\right| \int_0^r du\, u^{dk-1} \int_{Gr(k,T_cM)} d\mu_{k,d}(V) \\ \times \left(\prod_{i=1}^{k+1} \int_{\mathbb{S}_1(V)} \sqrt{\det(g_{\exp_c(uw_i)})} \left|d\mathrm{vol}_{\mathbb{S}_1(V)}(w_i)\right| \right) \Upsilon_1^{d-k}(\boldsymbol{w}) f(\exp_c(u\boldsymbol{w})), \tag{5.1}$$

where $\boldsymbol{w} = (w_1, \dots, w_{k+1}) \in (\mathbb{S}_1(V))^{k+1}$, $\Upsilon_1(\boldsymbol{w})$ the k-volume of the parallelogram generated by the $\boldsymbol{w}$ in T_cM, and $d\mathrm{vol}_{\mathbb{S}_1(V)}$ is the Riemannian measure of the Euclidean unit sphere in V.

Proof We start by noticing that since $r < r_{\max}$ the center $c = c(\mathbf{y})$, as in Lemma 4.2, is well-defined and the points in $\mathbf{y}$ are all at the same distance $u \leq r$ from c. Hence, the points lie in a normal sphere centered at c, and can be written as $y_i = \exp_c(v_i)$, for some $v_i \in T_cM$ with $|v_i| = u$. We will show below that these v_i-s lie in a k-dimensional subspace $V \subset T_{c(\mathbf{y})}M$. In this case, (5.1) is the result of integrating first over the center point c, then over the possible distances $u = \rho(\mathbf{y}) \in [0, r]$, then over all possible k-dimensional subspaces V in T_cM containing the vectors v_i, and finally, over the $(k-1)$-spheres in V, where the v_i-s live. Next, we will justify (5.1) by examining how the measures change under this reparametrization.

Given the center $c = c(\mathbf{y})$ and fixing normal coordinates $(x^1, \dots, x^d)$ centered in c, we have

$$\begin{aligned}\mathrm{dvol}_g(\mathbf{y}) &= \bigwedge_{i=1}^{k+1} \mathrm{dvol}_g(y_i) = \bigwedge_{i=1}^{k+1} \sqrt{\left|\det(g_{y_i})\right|}\, dx^1(y_i) \wedge \cdots \wedge dx^d(y_i) \\ &= \prod_{i=1}^{k+1} \sqrt{\left|\det(g_{y_i})\right|} \bigwedge_{i=1}^{k+1} dx^1(y_i) \wedge \cdots \wedge dx^d(y_i).\end{aligned}$$

Now let u denote the distance between y_i and c, then we have that $y_i = \exp_c(v_i) = \exp_c(u\frac{v_i}{|v_i|})$, for some $v_i \in T_cM$. In fact, in the coordinates $(x^1, \dots, x^d)$ we have $v_i = \sum_{j=1}^d x^j(y_i)(\frac{\partial}{\partial x^i})_c$. Then, using the Euclidean version of the Blaschke-Petkantschin formula [32] in the coordinates $(x^1, \dots, x^d)$ we have

$$\left|\mathrm{dvol}_g(\mathbf{y})\right| = \left(\prod_{i=1}^{k+1} \sqrt{\left|\det(g_{\exp_c(v_i)})\right|}\right) \Upsilon_u^{d-k}(\mathbf{v})\ \left|\mathrm{dvol}_g(c)\right|\ du\, d\mu_{k,d}(V) \left|\bigwedge_{i=1}^{k+1} \mathrm{dvol}_{\mathbb{S}_u(V)}(v_i)\right|.$$

Recall that $\Upsilon_u(\mathbf{v})$ denotes the k-volume of the parallelogram spanned by the v_i-s in $V \subset T_cM$, measured using the Euclidean metric on V induced by taking the ∂_{x^i} to form an orthonormal frame on T_cM. Equivalently, if we choose the coordinates $(x^1, \dots, x^d)$ so that $\{(\partial_{x^i})_c\}_{i=1}^d$ forms an orthonormal frame of $(T_cM, \langle\cdot,\cdot\rangle_c)$, then $\Upsilon_u(\mathbf{v})$ is simply the k-volume of the parallelogram spanned by the v_i-s in T_cM. In particular, Υ_u is homogeneous of degree k in u, so that $\Upsilon_u = u^k\Upsilon_1$. Similarly $\mathrm{dvol}_{\mathbb{S}_u(V)}$ is homogeneous of degree $k-1$, and therefore $\mathrm{dvol}_{\mathbb{S}_u(V)} = u^{k-1}\,\mathrm{dvol}_{\mathbb{S}_1(V)}$. Putting it all together we have

$$\left|\mathrm{dvol}_g(\mathbf{y})\right| = \left|\mathrm{dvol}_g(c)\right| u^{dk-1} du\, d\mu_{k,d}(V) \left|\bigwedge_{i=1}^{k+1} \sqrt{\left|\det(g_{\exp_c(v_i)})\right|}\ \mathrm{dvol}_{\mathbb{S}_1(V)}\left(\frac{v_i}{|v_i|}\right)\right| \Upsilon_1^{d-k}(\mathbf{v}),$$

Finally, note that $y_i = \exp_c(v_i) = \exp_c(u\frac{v_i}{|v_i|})$, and by setting $w_i = \frac{v_i}{|v_i|}$ we obtain (5.1).

To complete the proof, we need to show that $v_1, \dots, v_{k+1}$ lie in a k-dimensional subspace of T_cM. We start by showing that $v_i = -u(\nabla\rho_{y_i})_c$, where $u = \rho_{y_i}(c) = |v_i|$. This follows from the fact that by definition, v_i is such that $\rho_{y_i}(\exp_c(sv_i)) = u(1-s)$ for all $s \in [0,1]$. In particular sv_i is the point on the sphere of radius su in T_cM, such that $\exp_c(sv_i)$ minimizes the distance to y_i. Hence, for any $v' \in T_{sv_i}T_cM \cong T_cM$ with $\langle v', v_i\rangle_c = 0$

$$\langle\nabla\rho_{y_i}, (d\exp_c)_{sv_i}(v')\rangle_{\exp_c(sv_i)} = \frac{d}{dt}|_{t=0}\rho_{y_i}(\exp_c(sv_i + tv')) = 0,$$

and similarly by differentiating $\rho_{y_i}(\exp_c(sv_i)) = u(1-s)$ with respect to s

$$\langle\nabla\rho_{y_i}, (d\exp_c)_{sv_i}v_i\rangle_{\exp_c(sv_i)} = -u.$$

The Gauss Lemma guarantees that $\left|d\exp_{sv_i} v_i\right| = |v_i| = u$ and so we conclude that for $s \neq 0$, $(\nabla\rho_{y_i})_{\exp_c(sv_i)} = -u^{-1}(d\exp_c)_{sv_i}v_i$. The continuity of $\nabla\rho_{y_i}$ in a normal ball and the fact that $(d\exp_c)_0$ is the identity then implies that $v_i = -u\nabla\rho_{y_i}$ at c, as we wanted to show. Now we simply need to show that the $\nabla\rho_{y_i}$'s for $i \in \{1, \dots, k+1\}$ are linearly dependent at c. Arguing by contradiction, suppose they are linearly independent and so span a $(k+1)$-dimensional space $W \subset T_{c(\mathbf{y})}M$. Let $W = \mathrm{span}\{\nabla\rho_{y_i}\}_{i=1}^{k+1}$, for a generic set $\mathbf{y} \in M^{k+1}$, $E(\mathbf{y}) \cap \exp_c(W) \subset E(\mathbf{y})$ is a nonempty 1-dimensional manifold (it is nonempty since $c \in E(\mathbf{y}) \cap \exp_c(W)$). By definition $c \in E(\mathbf{y}) \cap \exp_c(W)$ is a critical point of all the ρ_{y_i}'s restricted to $E(\mathbf{y})$ and so the orthogonal projection of $(\nabla \rho_{y_i})_c$'s to $T_c(E(\mathbf{y}) \cap \exp_c(W))$ must vanish. However, by definition $\{\nabla \rho_{y_i}\}_{i=1}^{k+1}$ span the tangent space to W, and therefore we have obtained a contradiction with the fact that $E(\mathbf{y}) \cap \exp_c(W)$ is a nonempty one-dimensional submanifold. ■

Remark 5.2. In the previous proof, the fact that the w_i-s lie in a k-dimensional subspace of T_cM can be intuitively explained as follows. If this is not true, then the image under the exponential map of a small ball around 0 in W is a smooth $(k+1)$-dimensional manifold which intersects $E(\mathbf{y})$ along a 1-dimensional manifold. Moreover, there is a direction along this intersection such that the function $\rho_{y_i}(\cdot)$ is decreasing, and so contradicting the fact that $c(\mathbf{y})$ is the minimum.

6 | EXPECTED BETTI NUMBERS—UPPER BOUND

In this section we present an upper bound on the Betti numbers of a random Čech complex in terms of Λ (recall that $\Lambda = n\omega_d r^d$ is approximately the expected number of points inside a ball of radius r). This upper bound is interesting by itself, and also useful for finding the upper threshold in Theorem 1.1. The main result in this section is the following.

Proposition 6.1 *Let $n \to \infty$ and $r \to 0$ in such a way that $\Lambda \to \infty$ and $\Lambda r \to 0$ Then, for every $1 \le k \le d-1$ there exists a positive constant $b_k > 0$ (depending on k, d, and g) such that*

$$\mathbb{E}\{\beta_k(r))\} \le \beta_k(M) + b_k n\Lambda^k e^{-\Lambda}.$$

To prove Proposition 6.1 we will use Morse theory discussed in Section 4. The main idea is to use the Morse inequalities (4.2) and bound the number of critical points of $\rho_{\mathcal{P}_n}$. Defining,

$$C_k(r_1, r_2) := \#\text{ critical points } c \text{ of } \rho_{\mathcal{P}_n} \text{ with index } k, \text{ such that } \rho_{\mathcal{P}_n}(c) \in (r_1, r_2],$$

we prove the following lemma.

Lemma 6.2 *Let $n \to \infty$ and $r, r_0 \to 0$ such that $r = o(r_0)$, $\Lambda \to \infty$, and $\Lambda_{r_0} r_0^2 \to 0$, where $\Lambda_{r_0} := \omega_d n r_0^d$. Then for every $k \ge 1$ we have*

$$\mathbb{E}\{C_k(r, r_0)\} = O(n\Lambda^{k-1}e^{-\Lambda}).$$

Remark 6.3. The proofs will make use of various constant values. Some of them will be given a name, while the ones whose value is not relevant for the main results will be denoted by $C > 0$ (which might depend on k, d, g).

Proof Let c be a critical point of the distance function $\rho_{\mathcal{P}_n}$ with $\rho_{\mathcal{P}_n}(c) \in (r, r_0]$. Following the discussion in Section 3 and Proposition 4.6, we know that c is generated by a subset $\mathcal{Y} \subset \mathcal{P}_n$, so that $c = c(\mathcal{Y})$, and $\mathcal{Y}$ satisfies the following:

$$(1)\ 0 \in \Delta(\mathcal{Y}) \subset T_c M, \qquad (2)\ B(\mathcal{Y}) \cap \mathcal{P} = \emptyset, \qquad (3)\ r < \rho(\mathcal{Y}) \leq r_0, \tag{6.1}$$

where $\Delta(\mathcal{Y})$ is defined in 4.5, and $B(\mathcal{Y}) := B_{\rho(\mathcal{Y})}(c(\mathcal{Y}))$. Note that in this case we have $\rho_{\mathcal{P}_n}(c) = \rho(\mathcal{Y})$. Next, we define the following indicator functions:

$$\begin{aligned} h(\mathcal{Y}) &:= \mathbb{1}\{0 \in \Delta(\mathcal{Y})\}, \\ h_{r,r_0}(\mathcal{Y}) &:= h(\mathcal{Y})\mathbb{1}\{r < \rho(\mathcal{Y}) \leq r_0\}, \text{ and} \\ g_{r,r_0}(\mathcal{Y}, \mathcal{P}) &:= h_{r,r_0}(\mathcal{Y})\mathbb{1}\left\{B(\mathcal{Y}) \cap (\mathcal{P} \backslash \mathcal{Y}) = \emptyset\right\}. \end{aligned} \tag{6.2}$$

With these definitions we can now write

$$C_k(r, r_0) = \sum_{\substack{\mathcal{Y} \subset \mathcal{P}_n \\ |\mathcal{Y}| = k+1}} g_{r,r_0}(\mathcal{Y}, \mathcal{P}_n).$$

Applying Palm theory to the mean value (see Theorem .1 in the appendix) we have that

$$\mathbb{E}\{C_k(r, r_0)\} = \frac{n^{k+1}}{(k+1)!}\mathbb{E}\left\{g_{r,r_0}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\right\},$$

where $\mathcal{Y}'$ is a set of $(k+1)$ i.i.d. points, uniformly distributed on M, and independent of $\mathcal{P}_n$. Next, the properties of the Poisson process $\mathcal{P}_n$ imply that

$$\mathbb{E}\left\{\mathbb{1}\left\{B(\mathcal{Y}') \cap \mathcal{P}_n = \emptyset\right\} \mid \mathcal{Y}'\right\} = \mathbb{P}\left(\mathcal{P}_n(B(\mathcal{Y}')) = 0 \mid \mathcal{Y}'\right) = e^{-n\operatorname{Vol}(B(\mathcal{Y}'))}.$$

Therefore,

$$\mathbb{E}\{C_k(r, r_0)\} = \frac{n^{k+1}}{(k+1)!}\int_{M^{k+1}} h_{r,r_0}(\mathbf{y})e^{-n\operatorname{Vol}(B(\mathbf{y}))}\left|\mathrm{dvol}_g(\mathbf{y})\right|.$$

We can now apply Lemma 5.1 with $f(\mathbf{y}) = h(\mathbf{y})e^{-n\operatorname{Vol}(B(\mathbf{y}))}$, and have

$$\begin{aligned} \mathbb{E}\{C_k(r, r_0)\} = \frac{n^{k+1}}{(k+1)!}&\int_M \left|\mathrm{dvol}_g(c)\right| \int_r^{r_0} du\, u^{dk-1} \int_{Gr(k,T_cM)} \mathrm{d}\mu_{k,d}(V) \\ &\times \left(\prod_{i=1}^{k+1}\int_{\mathbb{S}_1(V)} \sqrt{\left|\det(g_{\exp_c(uw_i)})\right|}\left|\mathrm{dvol}_{\mathbb{S}_1(V)}(w_i)\right|\right)\Upsilon_1^{d-k}(\mathbf{w})f(\exp_c(u\mathbf{w})), \end{aligned}$$

where $c = c(\mathbf{y})$, $u = \rho(\mathbf{y})$, and $\mathbf{y} = \exp_c(u\mathbf{w})$. Replacing $f(\mathbf{y})$ with $h(\exp_c(u\mathbf{w}))e^{-n\operatorname{Vol}(B_u(c))}$, we have

$$\begin{aligned} \mathbb{E}\{C_k(r, r_0)\} = \frac{n^{k+1}}{(k+1)!}&\int_M \left|\mathrm{dvol}_g(c)\right| \int_r^{r_0} du\, u^{dk-1}e^{-n\operatorname{Vol}(B_u(c))} \int_{Gr(k,T_cM)} \mathrm{d}\mu_{k,d}(V) \\ &\times \left(\prod_{i=1}^{k+1}\int_{\mathbb{S}_1(V)} \sqrt{\left|\det(g_{\exp_c(uw_i)})\right|}\left|\mathrm{dvol}_{\mathbb{S}_1(V)}(w_i)\right|\right)\Upsilon_1^{d-k}(\mathbf{w})h(\exp_c(u\mathbf{w})). \end{aligned}$$

As the Grassmannian $Gr(k, T_cM)$ is compact, there is a subspace $V_{\max} \subset T_cM$ which maximizes the last integral over $(\mathbb{S}_1(V))^{k+1}$ and we have

$$\mathbb{E}\{C_k(r,r_0)\} \le Cn^{k+1}\int_M \left|\mathrm{dvol}_g(c)\right| \int_r^{r_0} du\, u^{dk-1}e^{-n\,\mathrm{Vol}(B_u(c))} \times \left(\prod_{i=1}^{k+1}\int_{\mathbb{S}_1(V_{max})}\sqrt{\left|\det(g_{\exp_c(uw_i)})\right|}\left|\mathrm{dvol}_{\mathbb{S}_1(V_{max})}(w_i)\right|\right)\Upsilon_1^{d-k}(\mathbf{w})h(\exp_c(u\mathbf{w})), \quad (6.3)$$

where $C > 0$ only depends on the dimension d and the index k. Using (3.2) and Lemma 3.1, for each $y_i = \exp_c(uw_i)$, we can write

$$\sqrt{\left|\det(g_{y_i})\right|} = 1 - \frac{Ric_{mn}}{3}x^m(y_i)x^n(y_i) + O(u^3),$$

The second-order term above is bounded by $\frac{1}{3}|Ric|\,r_0^2$ (in fact only by the values of Ric restricted to V_{max}). In addition, $\Upsilon_1(\mathbf{w})$ (the k-volume of the parallelogram generated by the w_i) is bounded from above, since $\mathbf{w}$ contains unit vectors. Putting it all back into (6.3) yields

$$\begin{aligned}\mathbb{E}\{C_k(r,r_0)\} &\le Cn^{k+1}(1+c_R r_0^2)^{k+1}\int_M \left|\mathrm{dvol}_g(c)\right|\int_r^{r_0} du\,u^{dk-1}e^{-n\,\mathrm{Vol}(B_u(c))} \\ &\quad\times\int_{(\mathbb{S}_1)^{k+1}}\left|\mathrm{dvol}_{\mathbb{S}_1}(\mathbf{v})\right| h(\exp_c(\mathbf{v})) \\ &= Cn^{k+1}\int_M\left|\mathrm{dvol}_g(c)\right|\int_r^{r_0} du\,u^{dk-1}e^{-n\,\mathrm{Vol}(B_u(c))},\end{aligned}$$

where c_R depends on the metric g. The constant C in the last line includes the product of $(1+c_R r_0^2)^{k+1}$ with integral of h over the $(k+1)$ spheres. Recall from (3.3) that for small u we have

$$\mathrm{Vol}(B_u(c)) = \omega_d u^d\left(1 - \frac{s(c)}{6(d+2)}u^2 + O(u^3)\right).$$

Thus, using the Taylor expansion $e^x = 1 + x + O(x^2)$, and the fact that $u \le r_0 \to 0$, we have

$$\begin{aligned}e^{-n\,\mathrm{Vol}(B_u(c))} &= e^{-n\omega_d u^d\left(1-\frac{s(c)}{6(d+2)}u^2+o(u^2)\right)} \\ &= e^{-n\omega_d u^d}\left(1+\frac{s(c)}{6(d+2)}n\omega_d u^{d+2} + o(nu^{d+2})\right) \\ &\le e^{-n\omega_d u^d}\left(1+s_{\max}n\omega_d r_0^{d+2}\right)\end{aligned} \quad (6.4)$$

where $s_{\max} = \sup_{c\in M}\frac{s(c)}{6(d+2)} + \delta$ for some $\delta > 0$. Applying the change of variables $s = \frac{u}{r}$, and recalling that $\Lambda = n\omega_d r^d$, yields

$$\begin{aligned}\mathbb{E}\{C_k(r,r_0)\} &\le Cn^{k+1}r^{dk}\left(1+s_{\max}n\omega_d r_0^{d+2}\right)\int_M\left|\mathrm{dvol}_g(c)\right|\int_1^{\frac{r_0}{r}} ds\,s^{dk-1}e^{-\Lambda s^d}, \\ &= Cn\Lambda^k\left(1+s_{\max}\Lambda_{r_0}r_0^2\right)\int_M\left|\mathrm{dvol}_g(c)\right|\int_1^{\frac{r_0}{r}} ds\,s^{dk-1}e^{-\Lambda s^d}.\end{aligned} \quad (6.5)$$

The last integral is known as the *lower incomplete gamma function* and has a closed form expression which yields,

$$\begin{aligned}\mathbb{E}\{C_k(r,r_0)\} &\le nC(1+s_{\max}\Lambda_{r_0}r_0^2)\left(\left(1-e^{-\Lambda_{r_0}}\sum_{j=0}^{k-1}\frac{\Lambda_{r_0}^j}{j!}\right)-\left(1-e^{-\Lambda}\sum_{j=0}^{k-1}\frac{\Lambda^j}{j!}\right)\right)\\ &= nC(1+s_{\max}\Lambda_{r_0}r_0^2)\left(e^{-\Lambda}\sum_{j=0}^{k-1}\frac{\Lambda^j}{j!}-e^{-\Lambda_{r_0}}\sum_{j=0}^{k-1}\frac{\Lambda_{r_0}^j}{j!}\right). \end{aligned} \tag{6.6}$$

Finally, using the assumptions that $\Lambda_{r_0}r_0^2 \to 0$, $\Lambda \to \infty$, and $r = o(r_0)$ yields $\mathbb{E}\{C_k(r,r_0)\} = O(n\Lambda^{k-1}e^{-\Lambda})$, and that completes the proof. ∎

We are now ready to prove Proposition 6.1.

Proof of Proposition 6.1 Let $\hat{\beta}_k(r) := \beta_k(r) - \beta_k(M)$, then we need to show that $\mathbb{E}\{\hat{\beta}_k(r)\} \le b_k n\Lambda^k e^{-\Lambda}$ for some $b_k > 0$. Using Morse theory, and in particular Lemma 4.9, it is enough to control the number of critical points of index $k+1$ occurring at a radius greater than r (ie, $C_{k+1}(r,+\infty)$).

Let $r < r_0 < r_{\max}$ such that $r = o(r_0)$, and $\Lambda_{r_0}r_0^2 \to 0$. Let E denote the event that $B_{r_0}(\mathcal{P}_n)$ covers M, then

$$\mathbb{E}\{\hat{\beta}_k(r)\} = \mathbb{E}\{\hat{\beta}_k(r) \mid E\}\mathbb{P}(E) + \mathbb{E}\{\hat{\beta}_k(r) \mid E^c\}\mathbb{P}(E^c). \tag{6.7}$$

The proof will now be split into two steps, dealing with evaluating each of the terms in the sum above.

Step 1: We prove that: $\mathbb{E}\{\hat{\beta}_k(r) \mid E\}\mathbb{P}(E) = O(n\Lambda^k e^{-\Lambda})$.

If the event E occurs, then $\rho_{\mathcal{P}_n}(x) \le r_0$ everywhere on M, and therefore Lemma 4.9 applied to $\rho_{\mathcal{P}_n}$ implies that any nonvanishing k-cycle in $\mathcal{C}_r$ which is mapped to a trivial cycle in M, is terminated by a critical point of index $k+1$ with value in $(r, r_0]$. Therefore, we must have $\hat{\beta}_k(r) \le C_{k+1}(r,r_0)$, and then

$$\mathbb{E}\{\hat{\beta}_k(r) \mid E\}\mathbb{P}(E) \le \mathbb{E}\{C_{k+1}(r,r_0) \mid E\}\mathbb{P}(E) \le \mathbb{E}\{C_{k+1}(r,r_0)\}. \tag{6.8}$$

Since $\Lambda \to \infty$ and $\Lambda_{r_0}r_0^2 \to 0$, using Lemma 6.2 we have that $\mathbb{E}\{C_{k+1}(r,r_0)\} = O(n\Lambda^k e^{-\Lambda})$, and that completes the first step.

Step 2: We prove that: $\mathbb{E}\{\hat{\beta}_k(r) \mid E^c\}\mathbb{P}(E^c) = o(n\Lambda^k e^{-\Lambda})$.

For any simplicial complex, the k-th Betti number is bounded by the number of k-dimensional faces (see, for example, [21]). Since the number of faces is bounded by $\binom{|\mathcal{P}_n|}{k+1}$, we have that

$$\begin{aligned}\mathbb{E}\{\beta_k(r) \mid E^c\}\mathbb{P}(E^c) &\le \mathbb{E}\left\{\binom{|\mathcal{P}_n|}{k+1} \mid E^c\right\}\mathbb{P}(E^c)\\ &= \sum_{m=k+1}^{\infty}\binom{m}{k+1}\mathbb{P}(|\mathcal{P}_n| = m \mid E^c)\mathbb{P}(E^c)\\ &= \sum_{m=k+1}^{\infty}\binom{m}{k+1}\mathbb{P}(E^c \mid |\mathcal{P}_n| = m)\mathbb{P}(|\mathcal{P}_n| = m),\end{aligned} \tag{6.9}$$

where we used Bayes' Theorem. Since $\mathcal{P}_n$ is a homogeneous Poisson process with intensity n we have that $\mathbb{P}(|\mathcal{P}_n| = m) = \frac{e^{-n}n^m}{m!}$, and also that given $|\mathcal{P}_n| = m$ we can write $\mathcal{P}_n$ as a set of m i.i.d. random variables $\mathcal{X}_m = \{X_1, \dots, X_m\}$ uniformly distributed on M. Therefore,

$$\mathbb{P}(E^c \mid |\mathcal{P}_n| = m) = \mathbb{P}\left(B_{r_0}(\mathcal{X}_m) \ne M\right).$$

Next, we will bound this coverage probability. Let $\mathcal{S}$ be a $\frac{r_0}{2}$-net of M, that is, for every $x \in M$ there is a point $s \in \mathcal{S}$ such that $\mathrm{dist}(x,s) \le \frac{r_0}{2}$. We can find such a $\frac{r_0}{2}$-net with $|\mathcal{S}| \le c_d r_0^{-d}$, where c_d is a constant that depends only on d and the metric g. Note that if for every $s \in \mathcal{S}$ there exists $X_i \in \mathcal{X}_m$ with $\rho(s, X_i) \le \frac{r_0}{2}$, then for every $x \in M$ we have

$$\mathrm{dist}(x, X_i) \le \mathrm{dist}(x,s) + \mathrm{dist}(s, X_i) \le r_0,$$

and therefore $B_{r_0}(\mathcal{X}_m) = M$. Thus, if $B_{r_0}(\mathcal{X}_m) \neq M$ then there exists $s \in \mathcal{S}$ with $\rho_{\mathcal{X}_m}(s) > \frac{r_0}{2}$, which yields

$$\mathbb{P}\left(B_{r_0}(\mathcal{X}_m) \neq M)\right) \le \sum_{s\in\mathcal{S}} \mathbb{P}\left(\rho_{\mathcal{X}_m}(s) > \frac{r_0}{2}\right) \le c_d r_0^{-d}(1-\alpha r_0^d)^m,$$

for any $\alpha \le 2^{-d}\omega_d\left(1 - s_{max} r_0^2\right)$, using Corollary 3.2. Putting it all back into (6.9) we have

$$\begin{aligned}\mathbb{E}\{\beta_k(r) \mid E^c\}\mathbb{P}(E^c) &\le \sum_{m=k+1}^{\infty} \binom{m}{k+1} c_d r_0^{-d}(1-\alpha r_0^d)^m \frac{e^{-n}n^m}{m!} \\ &= C r_0^{-d}(n(1-\alpha r_0^d))^{k+1} e^{-\alpha n r_0^d} \\ &\le C r_0^{-d} n^{k+1} e^{-\alpha n r_0^d}.\end{aligned}$$

To finish the proof, we take $r_0 = r\left(\frac{\omega_d}{\alpha}(1+|\log r|)\right)^{1/d}$, and use the assumption that $\Lambda r \to 0$. In particular, one can show that: (a) $r = o(r_0)$, (b) $\Lambda_{r_0} r_0^2 \to 0$, and (c) $r_0^{-d} n^{k+1} e^{-\alpha n r_0^d} = o(ne^{-\Lambda}\Lambda^k)$. Therefore,

$$\mathbb{E}\{\beta_k(r) \mid E^c\}\mathbb{P}(E^c) = o(ne^{-\Lambda}\Lambda^k). \tag{6.10}$$

Finally, recall that $\hat{\beta}_k(r) = \beta_k(r) - \beta_k(M)$. Note that $\mathbb{E}\{\beta_k(M) \mid E^c\} = \beta_k(M)$, and in addition, similar calculations to the ones above yield $\mathbb{P}(E^c) = O(e^{-\alpha n r_0^d}) = o(n\Lambda^k e^{-\Lambda})$. These facts together with (6.10) show that $\mathbb{E}\{\hat{\beta}_k(r) \mid E^c\}\mathbb{P}(E^c) = o(ne^{-\Lambda}\Lambda^k)$, and conclude the proof. ■

Remark 6.4. Notice that if the scalar curvature is everywhere negative and instead of requiring that $nr^{d+2} \to 0$ take $nr^{d+3} \to 0$, then the above bound gets smaller, namely:

$$\mathbb{E}\{\beta_k(r)\} \le \beta_k(M) + b_k(1 + s_{\max}\Lambda_{r_0} r_0^2) n\Lambda^k e^{-\Lambda},$$

with $s_{\max} = \delta + \sup_{c\in M} \frac{s(c)}{6(d+2)} < 0$ (ie, we choose δ so that $s_{\max} < 0$). It is therefore conceivable that the scalar curvature affects some lower-order term the phase transition of Theorem 1.1.

7 | EXPECTED BETTI NUMBERS–LOWER BOUND

In this section we compute a lower bound for the Betti numbers of C_r in terms of Λ. While in Section 6 we obtained an upper bound by making use of the Morse inequalities to simply count critical points, in order to obtain a lower bound we must proceed differently, since even if there are many critical points it is not clear which homology degree they contribute to. Thus, we shall instead consider a special type of critical points that are guaranteed to generate nontrivial cycles. These were first introduced in [7] and named Θ-cycles, after their unique structure. We start by stating the main result in this section.

Proposition 7.1 *Let $n \to \infty$ and $r \to 0$ such that $\Lambda \to \infty$ and $\Lambda r^2 \to 0$, then for every $1 \le k \le d-1$ there exists $a_k > 0$ (depending only on k, d and the metric g) such that*

$$\mathbb{E}\{\beta_k(r)\} \ge a_k n \Lambda^{k-2} e^{-\Lambda}.$$

Moreover, if (M, g) has everywhere positive scalar curvature, one just needs to assume that nr^{d+2} stays bounded.

As mentioned earlier, the proof uses the strategy of [7], and follows from combining the following lemmas. We start with some definitions. Let $\mathcal{P} \subset M$ and let $\mathcal{Y} \subset \mathcal{P}$ be a generic set. For $\alpha > 0$ define the closed annulus

$$A_\alpha(\mathcal{Y}) = B_{\rho(\mathcal{Y})}(c(\mathcal{Y})) \backslash B^\circ_{\alpha\rho(\mathcal{Y})}(c(\mathcal{Y})),$$

where $B^\circ_r(p)$ is an open ball. Notice that $\alpha > 0$ represents a scale invariant quantity. The following lemma is taken from [7] where it is proved for the torus. However, the proof remains the same for any compact Riemmanian manifold M.

Lemma 7.2 (Lemma 7.1 in [7]) *Let $\mathcal{P} \subset M$, and let $\mathcal{Y} \subset \mathcal{P}$ be a set of $k+1$ points, such that $c(\mathcal{Y})$ is a critical point of index k. Define*

$$\phi = \phi(\mathcal{Y}) := \frac{1}{2\rho(\mathcal{Y})} \min_{v \in \partial\Delta(\mathcal{Y})} |v|,$$

If $\rho(\mathcal{Y}) < r_{\max}$ and $A_\phi(\mathcal{Y}) \subset B_{\rho(\mathcal{Y})}(\mathcal{P})$, then the critical point $c(\mathcal{Y})$ generates a new nontrivial cycle in $H_k(B_{\rho(\mathcal{Y})}(\mathcal{P}))$.

The cycles created this way are called Θ-cycles, and the the idea behind the proof of Proposition 7.1 is to count them. Let $\epsilon > 0$, and define $\beta_k^\epsilon(r)$ to be the number of Θ-cycles generated by those $\mathcal{Y}$, such that

$$\text{(C1) } \rho(\mathcal{Y}) \in (r_1, r], \quad \text{(C2) } B_{r_2}(c(\mathcal{Y})) \cap \mathcal{P} = \mathcal{Y}, \quad \text{and} \quad \text{(C3) } \phi(\mathcal{Y}) \ge \epsilon,$$

where $r_2 > r > r_1 > 0$ are positive real constants (to be determined later). The next lemma shows that indeed, the Θ-cycles counted by β_k^ϵ provide a lower bound on the Betti numbers. In the statement of the next result, the constant c_g is taken from Lemma .1.

Lemma 7.3 *Let $r, r_2 \in \mathbb{R}^+$ with $r_2 > r$, then for $r_1 > r\sqrt{1 - \frac{1}{c_g^2}\left(\frac{r_2}{r} - 1\right)^2}$ and any $\epsilon \in (0, 1)$, we have $\beta_k(r) \ge \beta_k^\epsilon(r)$.*

Proof We need to show that any such Θ-cycle created prior to r still exists at r and so gives rise to a nonzero element in $H_k(\mathcal{C}_r)$, that is, it contributes to $\beta_k(r)$.

The proof of Lemma 7.2 uses the fact that every Θ-cycle introduces an uncovered k-simplex Δ in $\mathcal{C}_{r_1}$ (ie, Δ is not a face of any $(k+1)$-simplex). Since uncovered simplexes cannot be boundaries, it is thus enough to show that Δ is still uncovered at radius r. This requires that $B_r^\cap(\mathcal{Y})$, does not intersect any of the balls $B_r(p)$, for $p \in \mathcal{P}\backslash\mathcal{Y}$. Using Lemma .1, and condition (C1) above, we have that for all $x \in B_r^\cap(\mathcal{Y})$,

$$\text{dist}(c(\mathcal{Y}), x) \le c_g\sqrt{r^2 - \rho(\mathcal{Y})^2} \le c_g\sqrt{r^2 - r_1^2}.$$

Using condition (C2), for all $p \in \mathcal{P}\backslash\mathcal{Y}$ we have $\mathrm{dist}(p, c(\mathcal{Y})) \geq r_2$. Thus, using the triangle inequality we have

$$\mathrm{dist}(p, x) \geq \mathrm{dist}(p, c(\mathcal{Y})) - \mathrm{dist}(c(\mathcal{Y}), x) \geq r_2 - c_g\sqrt{r^2 - r_1^2}.$$

Hence, if we take $r_1 > r\sqrt{1 - \frac{1}{c_g^2}\left(\frac{r_2}{r} - 1\right)^2}$, we get that $\mathrm{dist}(p, x) > r$, which implies that $B_r(\mathcal{P}\backslash\mathcal{Y}) \cap B_r^{\cap}(\mathcal{Y}) = \emptyset$. Thus, Δ is still uncovered at radius r, which completes the proof. ■

Next, we define the following (related to conditions (C1)-(C3) above).

$$\begin{aligned} h_r^{\epsilon}(\mathcal{Y}) &:= h_{r_1,r}(\mathcal{Y})\mathbb{1}\left\{\phi(\mathcal{Y}) \geq \epsilon\right\}, \\ g_r^{\epsilon}(\mathcal{Y}, \mathcal{P}) &:= h_r^{\epsilon}(\mathcal{Y}, \mathcal{P})\mathbb{1}\left\{B_{r_2}(c(\mathcal{Y})) \cap (\mathcal{P}\backslash\mathcal{Y}) = \emptyset\right\}\mathbb{1}\left\{A_{\epsilon}(\mathcal{Y}) \subset B_{\rho(\mathcal{Y})}(\mathcal{P})\right\}. \end{aligned}$$

Thus, we can write

$$\beta_k^{\epsilon}(r) = \sum_{\mathcal{Y}\subset\mathcal{P}_n} g_r^{\epsilon}(\mathcal{Y}, \mathcal{P}_n). \tag{7.1}$$

The last result we need before proving Proposition 7.1 is the following lemma.

Lemma 7.4 *Let $\epsilon > 0$ be sufficiently small, and let $r > 0$ be such that $\Lambda \to \infty$, and $\Lambda r^2 \to 0$. Then there exists $a_k > 0$, and there exists a choice of r_1, r_2 with $0 < r_1 < r < r_2$, such that*

$$\mathbb{E}\{\beta_k^{\epsilon}(r)\} \geq a_k n\Lambda^{k-2}e^{-\Lambda}.$$

Proof Let $\epsilon > 0$, then the expectation of $\beta_k^{\epsilon}(r)$, can be computed in a similar way to the computation in the proof of Proposition 6.1. Suppose that $r_2 < r_{\max}$ (defined in Section 3), then using Palm theory and the properties of Poisson processes we have (Theorem .1),

$$\begin{aligned} \mathbb{E}\{\beta_k^{\epsilon}(r)\} &= \frac{n^{k+1}}{(k+1)!}\mathbb{E}\left\{g_r^{\epsilon}(\mathcal{Y}', \mathcal{Y}' \cup \mathcal{P}_n)\right\} \\ &= \frac{n^{k+1}}{(k+1)!}\int_{M^{k+1}} h_r^{\epsilon}(\mathbf{y})p_{\epsilon}(\mathbf{y})e^{-n\,\mathrm{Vol}(B_{r_2}(c(\mathbf{y})))}\left|\mathrm{dvol}_g(\mathbf{y})\right|, \end{aligned} \tag{7.2}$$

where

$$p_{\epsilon}(\mathbf{y}) := \mathbb{P}\left(A_{\epsilon}(\mathcal{Y}') \subset B_{\rho(\mathcal{Y}')}(\mathcal{P}_n \cup \mathcal{Y}') \mid \mathcal{Y}' = \mathbf{y}, \mathcal{P}_n \cap B_{r_2}(c(\mathcal{Y}')) = \emptyset\right).$$

We shall now evaluate the integral in (7.2) in two steps.

Step 1: We show that $p_{\epsilon} \to 1$ uniformly in $\mathbf{y}$ as $n \to \infty$. Denoting $\mathbb{P}_{\emptyset}(\cdot) := \mathbb{P}\left(\cdot \mid \mathcal{P}_n \cap B_{r_2}(c(\mathbf{y})) = \emptyset\right)$, we have that

$$p_{\epsilon}(\mathbf{y}) \geq \mathbb{P}_{\emptyset}\left(A_{\epsilon}(\mathbf{y}) \subset B_{\rho(\mathbf{y})}(\mathcal{P}_n)\right).$$

In the following we will use the shorthand notation:

$$\rho = \rho(\mathbf{y}), \quad A_{\epsilon} = A_{\epsilon}(\mathbf{y}), \quad B_{r_2} = B_{r_2}(c(\mathbf{y})), \quad p_{\epsilon} = p_{\epsilon}(\mathbf{y}).$$

Then, using Equation (3.4) we have

$$\mathrm{Vol}(A_\epsilon) = d\omega_d \int_{\epsilon\rho}^{\rho} u^{d-1}\left(1 - \frac{s(c(\mathbf{y}))}{6d}u^2 + O(u^3)\right) du \le \omega_d \rho^d (1-\epsilon^d)(1 + s_{\max}\rho^2),$$

where $s_{\max} = \sup_M(-\frac{s(c(\mathbf{y}))}{6d}) + \delta$, for some $\delta > 0$ as in the proof of Proposition 6.1. Let $\mathcal{S}$ be a $(\epsilon\rho/2)$-net of A_ϵ, that is, for every $x \in A_\epsilon$ there exists $s \in \mathcal{S}$ with $\mathrm{dist}(x, s) \le \epsilon\rho/2$. Since M is a compact manifold, we can find a positive constant c such that $c^{-1}u^d \le \mathrm{Vol}(B_u(p)) \le cu^d$ for all $p \in M$ and $u \le r_{\max}$. Therefore,

$$|\mathcal{S}| \le C\frac{\mathrm{Vol}(A_\epsilon)}{\inf_{p\in M}\mathrm{Vol}(B_{\frac{\epsilon\rho}{2}}(p))} \le C\frac{\rho^d(1-\epsilon^d)}{(\epsilon\rho/2)^d} = C\frac{1-\epsilon^d}{\epsilon^d}, \tag{7.3}$$

where the constant C changes in each inequality (see Remark 6.3), and depends only on the metric g. In other words, the bound for the number of points in this net only depends on ϵ and the metric g. If for all $s \in \mathcal{S}$, we have $\mathcal{P}_n \cap B_{\rho(1-\epsilon/2)}(s) \neq \emptyset$, then the triangle inequality implies that $A_\epsilon \subset B_\rho(\mathcal{P}_n)$. Notice that since we are conditioning on the event $\left\{\mathcal{P}_n \cap B_{r_2} = \emptyset\right\}$, and using spatial independence property of the Poisson process, for every $s \in \mathcal{S}$ we have

$$\mathbb{P}_\emptyset\left(\mathcal{P}_n \cap B_{\rho(1-\epsilon/2)}(s) = \emptyset\right) = e^{-n\,\mathrm{Vol}(B_{\rho(1-\epsilon/2)}(s)\setminus B_{r_2})},$$

and therefore,

$$\begin{aligned} p_\epsilon &\ge \mathbb{P}_\emptyset\left(\forall s \in \mathcal{S} : \mathcal{P}_n \cap B_{\rho(1-\epsilon/2)}(s) \neq \emptyset\right) \\ &= 1 - \mathbb{P}_\emptyset\left(\exists s \in \mathcal{S} : \mathcal{P}_n \cap B_{\rho(1-\epsilon/2)}(s) = \emptyset\right) \\ &\ge 1 - \sum_{s\in\mathcal{S}} e^{-n\,\mathrm{Vol}(B_{\rho(1-\epsilon/2)}(s)\setminus B_{r_2})} \\ &\ge 1 - C\max_{s\in\mathcal{S}} e^{-n\,\mathrm{Vol}(B_{\rho(1-\epsilon/2)}(s)\setminus B_{r_2})}, \end{aligned} \tag{7.4}$$

where for the last inequality we used (7.3), and the fact that ϵ is fixed. In order to estimate $\mathrm{Vol}(B_{\rho(1-\epsilon/2)}(s)\setminus B_{r_2})$ we look at the radial arc-length parametrized geodesic γ, from $c(\mathbf{y})$ to s (see Figure D1). This geodesic first enters $B_{\rho(1-\epsilon/2)}(s)\setminus B_{r_2}$ at a point $p_{\mathrm{in}} = \gamma(r_2)$ which is at distance r_2 from $c(\mathbf{y})$ and leaves it through $p_{\mathrm{out}} = \gamma(\rho(1-\epsilon/2) + \mathrm{dist}(s, c(\mathbf{y})))$. Note that since $s \in A_\epsilon$, we have that $\mathrm{dist}(s, c(\mathbf{y})) \ge \epsilon\rho$. Therefore,

$$\mathrm{dist}(p_{\mathrm{in}}, p_{\mathrm{out}}) = \rho(1-\epsilon/2) + \mathrm{dist}(s, c(\mathbf{y})) - r_2 \ge \rho(1+\epsilon/2) - r_2.$$

Now, take

$$r_2 = r(1+\xi), \quad \text{and} \quad r_1 = r(1 - \xi^2/2c_g^2) \tag{7.5}$$

with $\xi = o(\epsilon)$ (ξ is to be determined later). Then, these satisfy the conditions in the statement of Lemma 7.3 and

$$\mathrm{dist}(p_{\mathrm{in}}, p_{\mathrm{out}}) \ge r_1(1+\epsilon/2) - r_2 > \frac{r\epsilon}{4},$$

for $\xi = o(\epsilon)$. As a consequence, the ball of radius $\frac{r\epsilon}{8}$ centered at the midpoint p_{mid} from p_{in} to p_{out} is completely contained in $B_{\rho(1-\epsilon/2)}(s)\setminus B_{r_2}$, that is

$$B_{r\epsilon/8}(p_{\mathrm{mid}}) \subset B_{\rho(1-\epsilon/2)}(s)\setminus B_{r_2}.$$

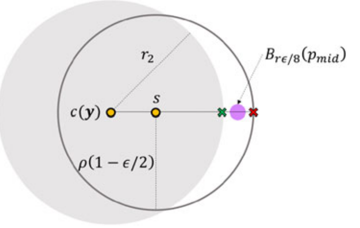

FIGURE 4 Bounding the volume of $R = B_{\rho(1-\epsilon/2)} \backslash B_{r_2}$.The solid line going through $c(\mathbf{y})$ and s represents the geodesic. This geodesic enters the region R through the green marker (p_{in}) and leaves through the red marker (p_{out}). The small shaded disc is a ball of radius $r\epsilon/8$ which we show to be contained inside R, and we use its volume as a lower bound for Vol(R) [Colour figure can be viewed at wileyonlinelibrary.com]

Hence $\text{Vol}(B_{\rho(1-\epsilon/2)}(s)\backslash B_{r_2}) \geq C\epsilon^d r^d$, for some constant $C > 0$, which depends on the metric, but can be taken to be independent of $\mathbf{y}$ and s. Putting this back into (7.4), we have

$$1 - Ce^{-Cn\epsilon^d r^d} \leq p_\epsilon(\mathbf{y}) \leq 1, \quad \forall \mathbf{y} \in M^{k+1}.$$

Since $\Lambda \to \infty$ we conclude that $p_\epsilon(\mathbf{y})$ converges uniformly to 1, as $\epsilon > \epsilon_0 > 0$, for some ϵ_0, that is, stays bounded away from zero.

Step 2: We estimate the integral in (7.2) and show that $\mathbb{E}\left\{\beta_k^\epsilon(r)\right\} = \Omega(n\Lambda^{k-2}e^{-\Lambda})$.

Combining the result of step 1 with (7.2), we have that for sufficiently large n

$$\mathbb{E}\{\beta_k^\epsilon(r)\} \geq \frac{1}{2}\frac{n^{k+1}}{(k+1)!}\int_{M^{k+1}} h_r^\epsilon(\mathbf{y})e^{-n\,\text{Vol}(B_{r_2}(c(\mathbf{y})))}\left|\text{dvol}_g(\mathbf{y})\right|,$$

Next, we use normal coordinates to estimate the volume of $B_{\rho(\mathbf{y})}(c(\mathbf{y}))$ for $\rho(\mathbf{y}) < r$, as in Corollary 3.2. This use is very similar to that used in Equation (6.4), in the proof of Lemma 6.2, yielding

$$e^{-n\,\text{Vol}(B_{r_2}(c(\mathbf{y})))} \geq e^{-\Lambda_{r_2}}(1 + s_{\min}\Lambda_{r_2}r^2),$$

where $\Lambda_{r_2} = \omega_d n r_2^d$ and $s_{\min} = \inf_{c\in M}\frac{s(c)}{6(d+2)} + \delta$, for some $\delta > 0$. We proceed as in the proof of Proposition 6.1 and then

$$\mathbb{E}\{\beta_k^\epsilon(r)\} \geq D_k^\epsilon(1 + s_{\min}\Lambda r^2)(1 - c_R r^2)n\Lambda^k e^{-\Lambda_{r_2}}\int_{\frac{r_1}{r}}^1 s^{dk-1}ds,$$

where $c_R = \sup_{V\in Gr(k,TM)}\left|-\frac{Ric^V}{3}\right| + \nu$ and

$$D_k^\epsilon := \frac{1}{2\omega_d^k(k+1)!}\int_M \left|\text{dvol}_g(c)\right|\int_{Gr(k,T_cM)} \text{d}\mu_{k,d}(V) \times$$
$$\times\left|\prod_{i=1}^{k+1}\int_{\mathbb{S}_1(V)}\inf_{s\in(r_1,r)}\sqrt{\left|\det(g_{\exp_c(sw_i)})\right|}\,h^\epsilon(\exp_c(s\mathbf{w}))\ \text{dvol}_{\mathbb{S}_1(V)}(w_i)\right|\Upsilon_1^{d-k}(\mathbf{w})$$

with $h^\epsilon(\mathbf{y}) := h(\mathbf{y})\mathbb{1}\{\phi(\mathbf{y}) \geq \epsilon\}$. First we notice that from (3.2) we have that if r and r_1 are small enough, then each of the terms $\det(g(\exp_c(uw_i)))$ is as close to 1 as we want. Secondly, we recall that

$$\phi(\exp_c(u\mathbf{w})) = \frac{1}{2}\min_{w\in\partial\Delta(\exp_c(\mathbf{w}))}|w|,$$

which is a nonnegative continuous function of $\mathbf{w}$ vanishing along a measure zero set consisting of those $\mathbf{w}$ for which $0 \in T_cM$ is contained in a face of their convex hull. This is a measure zero set and for any sufficiently small $\epsilon > 0$, the support of $\mathbb{1}\left\{\phi(\exp_c(u\mathbf{w})) \geq \epsilon\right\}$ has a nonzero measure. Also, notice that if $\phi(\exp_c(u\mathbf{w})) \geq \epsilon$ then $\Upsilon_1(\mathbf{w})$ (the k-volume of the simplex $\Delta(\mathbf{w})$) is bounded from below by a quantity of the order of ϵ^k. Putting these two facts together we conclude that for all sufficiently small $\epsilon > 0$, we do have $D_k^\epsilon > 0$. Moreover, since we set $r_1 = r(1 - \xi^2/2c_g^2)$ we have $\int_{\frac{r_1}{r}}^1 s^{dk-1}ds = \xi^2/2c_g^2 + O((\xi^2/2c_g^2)^2) \geq \xi^2/3c_g^2$, and thus

$$\mathbb{E}\left\{\beta_k^\epsilon(r)\right\} \geq \frac{D_k^\epsilon}{3c_g^2}(1 + s_{\min}\Lambda r^2)(1 - c_{_R} r^2)^{k+1} n\Lambda^k \xi^2 e^{-\Lambda_{r_2}}. \tag{7.6}$$

Note that in the exponent we have $\Lambda_{r_2} = \Lambda(1 + \xi)^d$ (since we set $r_2 = r(1 + \xi)$), and therefore $e^{-\Lambda_{r_2}} = e^{-\Lambda(1+\xi)^d} = e^{-\Lambda}e^{\Lambda(-d\xi+o(\xi))}$. However, as $\Lambda \to +\infty$ this second exponential, $e^{\Lambda(-d\xi+o(\xi))}$, is bounded from below if and only if ξ decays at least as fast as $O(\Lambda^{-1})$. Therefore, we take $\xi = \Lambda^{-1}$, in which case $e^{-\Lambda_{r_2}} \geq \frac{1}{2}e^{-d}e^{-\Lambda}$, and putting it back into (7.6) we have

$$\mathbb{E}\left\{\beta_k^\epsilon(r)\right\} \geq \frac{D_k^\epsilon}{6c_g^2}e^{-d}(1 + s_{\min}\Lambda r^2)(1 - c_{_R} r^2)^{k+1} n\Lambda^{k-2} e^{-\Lambda}.$$

Now if we take $\Lambda \to +\infty$, but $\Lambda r^2 \to 0$, then the statement follows. ∎

Remark 7.5. A more refined conclusion of the statement above would be that given a sufficiently small $\epsilon > 0$ and $r > 0$, then for $r_1 = r(1 - \frac{\Lambda^{-2}}{2c_g^2})$ and $r_2 = r(1 + \Lambda^{-1})$ the following holds.

1. If $\Lambda \to +\infty$ and $\Lambda r^2 \to 0$, then $\mathbb{E}\left\{\beta_k^\epsilon(r)\right\} \geq C(k, d, \epsilon, g)n\Lambda^{k-2}e^{-\Lambda}$, for some $C(k, d, \epsilon, g) > 0$ depending on k, d, ϵ and the metric g.
2. If (M, g) has everywhere positive scalar curvature, then as $\Lambda \to +\infty$ one does not need to assume that $\Lambda r^2 \to 0$. Instead it is enough to assume this stays bounded, in which case

$$\mathbb{E}\left\{\beta_k^\epsilon(r)\right\} \geq Cn\Lambda^{k-2}(1 + s_{\min}\Lambda r^2)e^{-\Lambda},$$

for some $C > 0$ as in the previous point.

Putting all the previous lemmas together, we can now prove the main result of this section.

Proof of Proposition 7.1 It follows from Lemma 7.3 that $\mathbb{E}\left\{\beta_k(r)\right\} \geq \mathbb{E}\left\{\beta_k^\epsilon(r)\right\}$, and using Lemma 7.4, we have

$$\mathbb{E}\left\{\beta_k(r)\right\} \geq a_k n\Lambda^{k-2}e^{-\Lambda},$$

for some a_k depending only on k, g and d. ∎

8 | SECOND MOMENT CALCULATIONS

The following proposition will be used to provide the lower threshold in Theorem 1.1. It uses a second moment argument, based on Chebyshev's inequality $\mathbb{P}\left(|X - \mathbb{E}\left\{X\right\}| \geq a\right) \leq \frac{\mathrm{Var}(X)}{a^2}$.

Proposition 8.1 *For all sufficiently small $\epsilon > 0$, and $1 \le k \le d-1$ the following holds. Fix $\delta \in (0,1)$. If $\Lambda = \log n + (k-2)\log\log n - w(n)$, with $w(n) \to \infty$, then*

$$\lim_{n\to+\infty} \mathbb{P}\left(\beta_k^\epsilon(r) > \delta\, \mathbb{E}\left\{\beta_k^\epsilon(r)\right\}\right) = 1.$$

Proof (Proof of proposition 8.1) Since $\beta_k^\epsilon(r) \ge 0$, then using Chebyshev's inequality we have

$$\mathbb{P}\left(\beta_k^\epsilon(r) \le \delta\, \mathbb{E}\left\{\beta_k^\epsilon(r)\right\}\right) \le \frac{\mathrm{Var}(\beta_k^\epsilon(r))}{(1-\delta)^2\mathbb{E}\{\beta_k^\epsilon(r)\}^2},$$

where $\mathrm{Var}(\beta_k^\epsilon(r)) = \mathbb{E}\{(\beta_k^\epsilon(r))^2\} - \mathbb{E}\{\beta_k^\epsilon(r)\}^2$. Thus, showing that the right hand side converges to zero will prove the statement. Recall from Equation (7.1) that $\beta_k^\epsilon(r) = \sum_{\mathcal{Y}\subset\mathcal{P}_n} g_r^\epsilon(\mathcal{Y},\mathcal{P}_n)$ and so we can write

$$(\beta_k^\epsilon(r))^2 = \sum_{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n} g_r^\epsilon(\mathcal{Y}_1,\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2,\mathcal{P}_n), \tag{8.1}$$

where $\mathcal{Y}_i$ $(i=1,2)$ run through all subsets of $\mathcal{P}_n$ with $(k+1)$-points. Next, defining

$$\Phi_r(\mathcal{Y}_1,\mathcal{Y}_2) := \mathbb{1}\left\{B_r(c(\mathcal{Y}_1)) \cap B_r(c(\mathcal{Y}_2)) = \emptyset\right\}.$$

we can write

$$\begin{aligned}
\mathbb{E}\{(\beta_k^\epsilon(r))^2\} &= \mathbb{E}\left\{\sum_{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n} g_r^\epsilon(\mathcal{Y}_1,\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2,\mathcal{P}_n)\right\}\\
&= \mathbb{E}\left\{\sum_{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n} g_r^\epsilon(\mathcal{Y}_1,\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2,\mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1,\mathcal{Y}_2)\right\}\\
&\quad + \mathbb{E}\left\{\sum_{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n} g_r^\epsilon(\mathcal{Y}_1,\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2,\mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1,\mathcal{Y}_2))\right\}\\
&= \mathbb{E}\{T_1\} + \mathbb{E}\{T_2\},
\end{aligned}$$

where T_1 and T_2 are the first and second sums appearing in the expectation above. Thus, we have

$$\begin{aligned}
\mathrm{Var}(\beta_k^\epsilon(r)) &= \mathbb{E}\{(\beta_k^\epsilon(r))^2\} - \mathbb{E}\{\beta_k^\epsilon(r)\}^2\\
&= (\mathbb{E}\{T_1\} - \mathbb{E}\{\beta_k^\epsilon(r)\}^2) + \mathbb{E}\{T_2\},
\end{aligned}$$

and our next step is to bound the terms $(\mathbb{E}\{T_1\} - \mathbb{E}\{\beta_k^\epsilon\}^2)$ and $\mathbb{E}\{T_2\}$ separately. Using Palm theory (Theorem .1) we can write

$$\begin{aligned}
\mathbb{E}\{\beta_k^\epsilon(r)\}^2 &= \frac{n^{2k+2}}{((k+1)!)^2}\mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1',\mathcal{Y}_1'\cup\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2',\mathcal{Y}_2'\cup\mathcal{P}_n')\}\\
\mathbb{E}\{T_1\} &= \frac{n^{2k+2}}{((k+1)!)^2}\mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1',\mathcal{Y}'\cup\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2',\mathcal{Y}'\cup\mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1,\mathcal{Y}_2)\},
\end{aligned} \tag{8.2}$$

where $\mathcal{Y}_1', \mathcal{Y}_2'$ are independent sets of $k+1$ points uniformly distributed in M, $\mathcal{Y}' = \mathcal{Y}_1' \cup \mathcal{Y}_2'$, and $\mathcal{P}_n'$ an independent copy of $\mathcal{P}_n$. Thus, using (8.2) we have

$$\begin{aligned}
\mathbb{E}\{T_1\} - \mathbb{E}\{\beta_k^\epsilon(r)\}^2 &= \frac{n^{2k+2}}{((k+1)!)^2}\Big(\mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}' \cup \mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2')\} \\
&\quad - \mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n')\Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2')\} \\
&\quad - \mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n')(1-\Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2'))\}\Big) \\
&\le \frac{n^{2k+2}}{((k+1)!)^2}\Big(\mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n)\Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2')\} \\
&\quad - \mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n')\Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2')\}\Big). \\
&= \frac{n^{2k+2}}{((k+1)!)^2}\mathbb{E}\{\Delta g_r^\epsilon\}
\end{aligned}$$

where

$$\Delta g_r^\epsilon := \left(g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n) - g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n')\right)\Phi_{2r}(\mathcal{Y}_1', \mathcal{Y}_2').$$

Similarly to [7], we can show that $\mathbb{E}\{\Delta g_r^\epsilon\} = 0$ as follows. Consider the conditional distribution of Δg_r^ϵ given $\mathcal{Y}_1', \mathcal{Y}_2'$, and denote $\mathbb{E}_{\mathcal{Y}_1', \mathcal{Y}_2'}\{\cdot\} = \mathbb{E}\left\{\cdot \mid \mathcal{Y}_1', \mathcal{Y}_2'\right\}$. If $\Delta g_r^\epsilon \neq 0$ then necessarily $B_{2r}(c(\mathcal{Y}_1')) \cap B_{2r}(c(\mathcal{Y}_2')) = \emptyset$. Using the spatial independence property of the Poisson process, together with the fact that the value of $g_r^\epsilon(\mathcal{Y}_i', \mathcal{Y}_i' \cup \mathcal{P}_n)$ only depends on the points of $\mathcal{P}_n$ lying inside $B_{2r}(c(\mathcal{Y}_i'))$, we conclude that

$$\begin{aligned}
\mathbb{E}_{\mathcal{Y}_1', \mathcal{Y}_2'}\left\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n)\right\} &= \mathbb{E}_{\mathcal{Y}_1', \mathcal{Y}_2'}\left\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)\right\}\mathbb{E}_{\mathcal{Y}_1', \mathcal{Y}_2'}\left\{g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n)\right\} \\
&= \mathbb{E}_{\mathcal{Y}_1', \mathcal{Y}_2'}\left\{g_r^\epsilon(\mathcal{Y}_1', \mathcal{Y}_1' \cup \mathcal{P}_n)\right\}\mathbb{E}_{\mathcal{Y}_1', \mathcal{Y}_2'}\left\{g_r^\epsilon(\mathcal{Y}_2', \mathcal{Y}_2' \cup \mathcal{P}_n')\right\}
\end{aligned}$$

since $\mathcal{P}_n$ and $\mathcal{P}_n'$ are independent and have the same distribution. Thus, we have that $\mathbb{E}_{\mathcal{Y}_1, \mathcal{Y}_2}\{\Delta g_r\} = 0$. Consequently, $\mathbb{E}\{\Delta g_r\} = \mathbb{E}\left\{\mathbb{E}_{\mathcal{Y}_1, \mathcal{Y}_2}\{\Delta g_r\}\right\} = 0$.

Next, we wish to bound $\mathbb{E}\{T_2\}$. We start by writing,

$$\begin{aligned}
T_2 &= \sum_{\mathcal{Y}_1, \mathcal{Y}_2 \subset \mathcal{P}_n} g_r^\epsilon(\mathcal{Y}_1, \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2, \mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)) \\
&= \sum_{j=0}^{k+1}\sum_{|\mathcal{Y}_1 \cap \mathcal{Y}_2| = j} g_r^\epsilon(\mathcal{Y}_1, \mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2, \mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1, \mathcal{Y}_2)).
\end{aligned} \tag{8.3}$$

In the following, it will be convenient to refer to the inner sum as I_j. Lemmas 8.2 and 8.3 respectively provide upper bounds on $\mathbb{E}\{I_0\}$ and $\mathbb{E}\left\{I_j\right\}$ for $j \ge 1$. The largest of these is the upper bound for $\mathbb{E}\{I_0\}$ which yields

$$\mathbb{E}\{T_2\} \le Cn\Lambda^{2k+1}\Lambda^{-4}e^{-\Lambda}\left(e^{-\omega_d\alpha\Lambda/\omega_{d-1}} + \alpha^d\right),$$

for any $\alpha < 1$. Using the fact that $\mathbb{E}\{\beta_k^\epsilon(r)\} \ge a_k n\Lambda^{k-2}e^{-\Lambda}$ and $\Lambda r \to 0$, we get

$$\frac{\mathbb{E}\{T_2\}}{\mathbb{E}\{\beta_k^\epsilon(r)\}^2} \le C\frac{n\Lambda^{2k+1}\Lambda^{-4}e^{-\Lambda}}{n^2\Lambda^{2k-4}e^{-2\Lambda}}(e^{-\omega_d\alpha\Lambda/\omega_{d-1}} + \alpha^d) \le \frac{\Lambda e^\Lambda}{n}(e^{-\omega_d\alpha\Lambda/\omega_{d-1}} + \alpha^d).$$

Taking $\Lambda = \log n + (k-2)\log\log n - w(n)$ and $\alpha = \frac{k\omega_{d-1}}{\omega_d}\frac{\log\log n}{\log n}$ we have

$$\frac{\mathbb{E}\{T_2\}}{\mathbb{E}\{\beta_k^\epsilon(r)\}^2} \leq Ce^{-w(n)}(\log n)^{k-1}\left(e^{-\frac{\omega_d}{\omega_{d-1}}\alpha\log n} + \alpha^d\right)$$
$$\leq Ce^{-w(n)}(\log n)^{k-1}\left(\frac{1}{(\log n)^k} + \frac{(\log\log n)^d}{(\log n)^d}\right)$$
$$\leq Ce^{-w(n)}\left(\frac{1}{\log n} + \frac{(\log\log n)^d}{(\log n)^{d+1-k}}\right) \to 0, \text{ as } n \to +\infty. \qquad \blacksquare$$

To conclude this section, we are left with bounding the $\mathbb{E}\left\{I_j\right\}$ terms used in the proof of the previous proposition. Recall that these are defined by

$$I_j = \sum_{|\mathcal{Y}_1\cap\mathcal{Y}_2|=j} g_r^\epsilon(\mathcal{Y}_1,\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2,\mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1,\mathcal{Y}_2)),$$

and first appeared in (8.3). The following two statements provide the needed bounds.

Lemma 8.2 (Estimate on I_0) *For any* $0<\alpha<1$ *we have*

$$\mathbb{E}\{I_0\} \leq Cn\Lambda^{2k+1}\Lambda^{-4}e^{-\Lambda}\left(e^{-\omega_d\alpha\Lambda/\omega_{d-1}} + \alpha^d\right).$$

Lemma 8.3 (Estimates on I_j, for $j\geq 1$) *For any* $0<\alpha<1$, *we have*

$$\mathbb{E}\{I_j\} \leq Cn\Lambda^{2k+1-j}r^{j(k-j)+1}\Lambda^{-4}e^{-\Lambda}\left(e^{-\omega_d\alpha\Lambda/\omega_{d-1}} + \alpha^{d-j+1}\right).$$

Proof of Lemma 8.2 We start estimating $\mathbb{E}\{I_0\}$ by using Palm theory (Corollary .2) as follows.

$$\begin{aligned}\mathbb{E}\{I_0\} &= \mathbb{E}\left\{\sum_{|\mathcal{Y}_1\cap\mathcal{Y}_2|=0} g_r^\epsilon(\mathcal{Y}_1,\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2,\mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1,\mathcal{Y}_2))\right\}\\ &= \frac{n^{2k+2}}{((k+1)!)^2}\mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1',\mathcal{Y}'\cup\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2',\mathcal{Y}'\cup\mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1',\mathcal{Y}_2'))\}\\ &\leq \frac{n^{2k+2}}{((k+1)!)^2}\int_{M^{2k+2}} h_r^\epsilon(\mathbf{y}_1)h_r^\epsilon(\mathbf{y}_2)e^{-n\operatorname{Vol}(\mathbf{y}_1,\mathbf{y}_2)}(1-\Phi_{2r}(\mathbf{y}_1,\mathbf{y}_2))\left|\mathrm{dvol}_g(\mathbf{y}_1,\mathbf{y}_2)\right|.\end{aligned} \tag{8.4}$$

where $\mathbf{y}_1,\mathbf{y}_2\in M^{k+1}$ are $(k+1)$-tuples of points, $\operatorname{Vol}(\mathbf{y}_1,\mathbf{y}_2) = \operatorname{Vol}(B_{r_2}(c_1)\cup B_{r_2}(c_2))$, and $c_i = c(\mathbf{y}_i)$.

Next, we need to compare the volume of this union with the Euclidean volume of Euclidean balls with a slightly different radius. Using Lemma 3.4 to compare the balls, and Lemma 3.3 to compare the volumes yields

$$\begin{aligned}\operatorname{Vol}(B_{r_2}(c_1)\cup B_{r_2}(c_2)) &\geq \operatorname{Vol}(B^E_{(1-\nu' r)r_2}(c_1)\cup B^E_{(1-\nu' r)r_2}(c_2))\\ &\geq (1-\nu r^2)\operatorname{Vol}^E(B^E_{(1-\nu' r)r_2}(c_1)\cup B^E_{(1-\nu' r)r_2}(c_2))\\ &\geq (1-\nu r^2)\operatorname{Vol}^E(B^E_{(1-\nu' r)r}(c_1)\cup B^E_{(1-\nu' r)r}(c_2))\\ &\geq (1-\nu r^2)(2\omega_d(1-\nu' r)^d r^d - \operatorname{Vol}^E(B^E_{(1-\nu' r)r}(c_1)\cap B^E_{(1-\nu' r)r}(c_2))\\ &\geq (1-(d\nu' r+\nu r^2)+o(r^2))\omega_d r^d\\ &\quad\times\left(1+\frac{\omega_{d-1}}{\omega_d}\frac{\operatorname{dist}(c_1,c_2)}{r}+O\left(\frac{\operatorname{dist}(c_1,c_2)^2}{r}\right)\right).\end{aligned}$$

where we used the fact that $r_2 > r$, and in the last inequality we used the following Tailor expansion for the volume of the intersection of two balls (see Appendix C in [7]),

$$\mathrm{Vol}^E(B_r^E(c_1) \cap B_r^E(c_2)) = \omega_d r^d - \omega_{d-1} r^{d-1}\,\mathrm{dist}(c_1, c_2) + O(r^{d-2}\,\mathrm{dist}^2(c_1, c_2)). \tag{8.5}$$

Next, let $\alpha > 0$, to be determined later, and separate the integration in (8.4) into two regions

$$\Omega_1 = \{(\mathbf{y}_1, \mathbf{y}_2) \in M^{2k+2} \mid \frac{\mathrm{dist}(c_1, c_2)}{r} \le \alpha\},$$
$$\Omega_2 = \{(\mathbf{y}_1, \mathbf{y}_2) \in M^{2k+2} \mid \frac{\mathrm{dist}(c_1, c_2)}{r} > \alpha\}.$$

Then, in S_1 we have $\mathrm{Vol}(\mathbf{y}_1, \mathbf{y}_2) \ge (1 - (d\nu' r + \nu r^2))\omega_d r^d$. Moreover, using a similar change of variables to that of Lemma 5.1, and taking c_2 to be in polar coordinates around c_1, we have

$$\begin{aligned}
I_0^{(1)} &:= \int_{\Omega_1} h_r^\epsilon(\mathbf{y}_1) h_r^\epsilon(\mathbf{y}_2) e^{-n\,\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2)}(1 - \Phi_{2r}(\mathbf{y}_1, \mathbf{y}_2)) \left|\mathrm{dvol}_g(\mathbf{y}_1, \mathbf{y}_2)\right| \\
&\le \int_{\Omega_1} h_r^\epsilon(\mathbf{y}_1) h_r^\epsilon(\mathbf{y}_2) e^{-(1-(d\nu' r+\nu r^2))\Lambda} \left|\mathrm{dvol}_g(\mathbf{y}_1, \mathbf{y}_2)\right| \\
&\le C e^{-(1-(d\nu' r+\nu r^2))\Lambda} \int_M \left|\mathrm{dvol}_g(c_1)\right| \int_0^{\alpha r} ds \int_{\mathbb{S}_1(T_{c_1}M)} s^{d-1}\,\mathrm{dvol}_{\mathbb{S}_1(T_{c_1}M)}(w) \\
&\quad \times \prod_{i=1}^{2} \int_{r_1}^{r} du_i \int_{Gr(k,d)} u_i^{k(d-k)} d\mu_{k,d}(V) \int_{(\mathbb{S}_1(V))^{k+1}} u_i^{(k-1)(k+1)}\,\mathrm{dvol}_{(\mathbb{S}_1(V))^{k+1}}(\mathbf{w}_i) h^\epsilon(\exp_{c_i}(u_i \mathbf{w}_i)),
\end{aligned}$$

with $c_2 = \exp_{c_1}(sw)$, $\mathbf{y}_i = \exp_{c_i}(\mathbf{w}_i)$ and $\mathbb{S}_1(V)$ denotes the unit sphere in the k-dimensional vector space V. Note that the term C includes not only an upper bound for metric-related terms, but also a bound for the Υ terms representing the parallelogram-volume. After carrying out the integration in the radial coordinates we get

$$\begin{aligned}
I_0^{(1)} &\le C e^{-(1-(d\nu' r+\nu r^2))\Lambda} (\alpha r)^d r^{2dk} \left(r^{dk} - r_1^{dk}\right)^2 \\
&\le C e^{-(1-(d\nu' r+\nu r^2))\Lambda} (\alpha r)^d r^{2dk} \left(1 - \left(\frac{r_1}{r}\right)^{dk}\right)^2,
\end{aligned}$$

for some new constant C depending on k, d and the metric g (see Remark 6.3). Then, taking $r_1 = (1 - \xi^2/2c_g^2)r$ with $\xi = \Lambda^{-1}$ (as in (7.5)) we conclude that

$$I_0^{(1)} \le C e^{-(1-(d\nu' r+\nu r^2))\Lambda} r^{d(2k+1)} \alpha^d \Lambda^{-4}.$$

At this stage we require that $\Lambda r \to 0$ as $n \to +\infty$. Then, $e^{-\Lambda+(d\nu' r+\nu r^2)\Lambda} \sim e^{-\Lambda}$, this yields

$$I_0^{(1)} \le C \alpha^d \Lambda^{-4} e^{-\Lambda} r^{d(2k+1)}.$$

We now turn to evaluate the integral over Ω_2. First, notice that $\mathrm{Vol}(\mathbf{y}_1, \mathbf{y}_2)$ is increasing with $\mathrm{dist}(c_1, c_2)$ and so attains its minimum value when $\mathrm{dist}(c_1, c_2) = \alpha r$ in this set. Secondly, notice that for the term $(1-\Phi_{2r}(\mathbf{y}_1,\mathbf{y}_2))$ to be nonzero we must have $\mathrm{dist}(c_1,c_2)\leq 4r$. Inserting $\mathrm{dist}(c_1,c_2)=\alpha r$ in the Taylor expansion (8.5) and using a similar change of variables yields

$$\begin{aligned}
I_0^{(2)} &:= \int_{\Omega_2} h_r^\epsilon(\mathbf{y}_1)h_r^\epsilon(\mathbf{y}_2)e^{-n\,\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2)}(1-\Phi_{2r}(\mathbf{y}_1,\mathbf{y}_2))\,\left|\mathrm{dvol}_g(\mathbf{y}_1,\mathbf{y}_2)\right| \\
&\leq \int_{\Omega_2} h_r^\epsilon(\mathbf{y}_1)h_r^\epsilon(\mathbf{y}_2)e^{-(1-(d\nu' r+\nu r^2))(1+\omega_d\alpha/\omega_{d-1})\Lambda}\,\left|\mathrm{dvol}_g(\mathbf{y}_1,\mathbf{y}_2)\right| \\
&\leq Ce^{-(1+\omega_d\alpha/\omega_{d-1})\Lambda}\int_M \left|\mathrm{dvol}_g(c_1)\right|\int_{\alpha r}^{4r} ds\int_{\mathbb{S}_1(T_{c_1}M)} s^{d-1}\,\mathrm{dvol}_{\mathbb{S}_1(T_{c_1}M)}(w)\times \\
&\quad\times\prod_{i=1}^{2}\int_{r_1}^{r} du_i\int_{Gr(k,d)} u_i^{k(d-k)}\,d\mu_{k,d}\int_{(\mathbb{S}_1(V))^{k+1}} u_i^{(k-1)(k+1)}\,\mathrm{dvol}_{(\mathbb{S}_1(V))^{k+1}}(\mathbf{v}_i)h_r^\epsilon(\exp_{c_i}(u_i\mathbf{w}_i)).
\end{aligned}$$

In this case the integration in the radial coordinates yields

$$\int_{\Omega_2} h_r^\epsilon(\mathbf{y}_1)h_r^\epsilon(\mathbf{y}_2)e^{-n\,\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2)}\,\left|\mathrm{dvol}_g(\mathbf{y}_1,\mathbf{y}_2)\right|\leq C\Lambda^{-4}e^{-\Lambda}e^{-\omega_d\alpha\Lambda/\omega_{d-1}}r^{d(2k+1)}.$$

Putting these all together into (8.4) we have

$$\mathbb{E}\{I_0\}\leq Cn\Lambda^{2k+1}\Lambda^{-4}e^{-\Lambda}\left(e^{-\omega_d\alpha\Lambda/\omega_{d-1}}+\alpha^d\right).$$

■

Proof of Lemma 8.3 As before we evaluate I_j using Palm theory (Corolllary .2)

$$\begin{aligned}
\mathbb{E}\{I_j\} &= \mathbb{E}\left\{\sum_{|\mathcal{Y}_1\cap\mathcal{Y}_2|=j} g_r^\epsilon(\mathcal{Y}_1,\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2,\mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1,\mathcal{Y}_2))\right\} \\
&= \frac{n^{2k+2-j}}{j!((k+1-j)!)^2}\mathbb{E}\{g_r^\epsilon(\mathcal{Y}_1',\mathcal{Y}'\cup\mathcal{P}_n)g_r^\epsilon(\mathcal{Y}_2',\mathcal{Y}'\cup\mathcal{P}_n)(1-\Phi_{2r}(\mathcal{Y}_1',\mathcal{Y}_2'))\} \\
&\leq \frac{n^{2k+2-j}}{j!((k+1-j)!)^2}\int_{M^{2k+2}} h_r^\epsilon(\mathbf{y}_1)h_r^\epsilon(\mathbf{y}_2)e^{-n\,\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2)}(1-\Phi_{2r}(\mathbf{y}_1,\mathbf{y}_2))\,\mathrm{dvol}_g(\mathbf{y}).
\end{aligned}\tag{8.6}$$

where

$$\begin{aligned}
\mathbf{y} &= (y_1,\dots,y_{2k+2-j})\subset M^{2k+2-j},\\
\mathbf{y}_1 &= (y_1,\dots,y_{k+1}),\\
\mathbf{y}_2 &= (y_1,\dots,y_j,y_{k+2},\dots,y_{2k+2-j}).
\end{aligned}$$

We can repeat the previous computations to show that

$$\begin{aligned}
\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2) &\geq (1-(d\nu' r+\nu r^2))\omega_d r^d \\
&\quad\times\left(1+\frac{\omega_{d-1}}{\omega_d}\frac{\mathrm{dist}(c_1,c_2)}{r}+O\left(\frac{\mathrm{dist}(c_1,c_2)^2}{r}\right)\right),
\end{aligned}$$

where again $c_i = c(\mathbf{y}_i)$. Proceeding using the same steps as in the previous proof, we fix $\alpha > 0$ and separate the integration in (8.6) into two regions

$$\Omega_1 = \left\{ \mathbf{y} \in M^{2k+2-j} \mid \frac{\mathrm{dist}(c_1, c_2)}{r} \leq \alpha \right\},$$
$$\Omega_2 = \left\{ \mathbf{y} \in M^{2k+2-j} \mid \frac{\mathrm{dist}(c_1, c_2)}{r} > \alpha \right\}.$$

To proceed, we need a slightly more involved version of the change of variables used than the one we used in Lemma 8.2 for I_0. The iterated integration goes as follows.

- Integrate over M to determine the first center c_1.
- Find $\mathbf{y}_1$ on a $(k-1)$-sphere centered at c_1 with radius in (r_1, r). This goes very much along the same lines as Lemma 5.1
- Find the second center c_2 so that the points $c_2 \in E = E(y_1, \dots, y_j)$ (ie, $y_1, \dots y_j$ are equidistant to c_2). Since $c_1 \in E$ as well, in order to find c_2 we integrate in E using geodesic polar coordinates around c_1. Note that E is of dimension $d - j + 1$.
- Fixing c_2, we need to choose the remaining $k + 1 - j$ points $(y_{k+1}, \dots, y_{2k+1-j})$. These points, together with $(y_1, \dots, y_j)$ span a k-dimensional vector space $V_2 \subset T_{c_2}M$. However, we must integrate over those spaces V_2 that contain the j-dimensional vector space generated by the $(y_1, \dots, y_j)$, that is, the span of the $\langle (\nabla \rho^2_{y_i})_{c_2} \rangle_{i=1}^{j}$. Hence, we integrate over the subspaces W that satisfy $V_2 = \langle (\nabla \rho^2_{p_i})_{c_2} \rangle_{i=1}^{j} \oplus W$. Note that these are $(k-j)$-dimensional subspaces of $T_{c_2}M$.
- Finally, we determine the remaining $k + 1 - j$ points, in geodesic polar coordinates, by integrating over a sphere in the vector subspace $V_2 \subset T_{c_2}M$.

In order to make the notation lighter we shall omit the reference to the points over which we are integrating in many occasions. Recalling that in S_1 we have $\mathrm{Vol}(\mathbf{y}_1, \mathbf{y}_2) \geq (1 - (dv'r + vr^2))\omega_d r^d$, we have

$$\begin{aligned}
I_j^{(1)} &:= \int_{\Omega_1} h_r^\epsilon(\mathbf{y}_1) h_r^\epsilon(\mathbf{y}_2) e^{-n\,\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2)} (1 - \Phi_{2r}(\mathbf{y}_1, \mathbf{y}_2)) \left|\mathrm{dvol}_g(\mathbf{y}_1, \mathbf{y}_2)\right| \\
&\leq \int_{\Omega_1} h_r^\epsilon(\mathbf{y}_1) h_r^\epsilon(\mathbf{y}_2) e^{-(1-(dv'r+vr^2))\Lambda} \left|\mathrm{dvol}_g(\mathbf{y}_1, \mathbf{y}_2)\right| \\
&\leq C e^{-(1-(dv'r+vr^2))\Lambda} \int_M \left|\mathrm{dvol}_g(c_1)\right| \\
&\quad \times \int_{r_1}^{r} du_1 \int_{Gr(k,d)} u_1^{k(d-k)} d\mu_{k,d}(V_1) \int_{(\mathbb{S}_1(V_1))^{k+1}} u_1^{(k-1)(k+1)} \,\mathrm{dvol}_{(\mathbb{S}_1(V_1))^{k+1}} \\
&\quad \times \int_0^{\alpha r} ds \int_{\mathbb{S}_1(T_{c_1}E)} s^{d-j} \,\mathrm{dvol}_{\mathbb{S}^{d-1}}(w) \int_{r_1}^{r} du_2 \int_{Gr(k-j,d)} u_i^{(k-j)(d-(k-j))} d\mu_{k-j,d}(W) \\
&\quad \times \int_{(\mathbb{S}_1(V_2))^{k+1-j}} u_2^{(k-1)(k+1-j)} \,\mathrm{dvol}_{(\mathbb{S}_1(V_2))^{k+1-j}}\; h_r^\epsilon(\exp_{c_1}(u_1 \mathbf{w}_1)) h_r^\epsilon(\exp_{c_2}(u_2 \mathbf{w}_2)),
\end{aligned}$$

with $c_2 = \exp_{c_1}(sw)$ and the points $\mathbf{y}_1 = \exp_{c_1}(u_1\mathbf{w}_1)$, $\mathbf{y}_2 = \exp_{c_2}(u_2\mathbf{w}_2)$ determined as in the previous discussion. As before, the constant C accounts for the metric-dependent terms and the Υ-terms representing the parallelogram-volumes. Integrating in the radial coordinates and taking $r_1 = (1 - \xi^2/2c_g^2)r$ with $\xi = \Lambda^{-1}$ we conclude that

$$I_j^{(1)} \leq C e^{-(1-(dv'r+vr^2))\Lambda} \alpha^{d-j+1} r^{d(2k+1-j)+j(k-j)+1} \left(1 - \left(\frac{r_1}{r}\right)^{dk}\right) \left(1 - \left(\frac{r_1}{r}\right)^{d(k-j)+j(k+1-j)}\right).$$

for some new C depending on j, k, d and the metric g. Given that $\Lambda r \to 0$ as $n \to +\infty$, the exponential term can be estimated as $e^{-\Lambda+(d\nu' r+\nu r^2)\Lambda} \sim e^{-\Lambda}$, which gives

$$I_j^{(1)} \le C\alpha^{d-j+1}\Lambda^{-4}e^{-\Lambda}r^{d(2k+1-j)+j(k-j)+1}.$$

We now turn to evaluate the integral over Ω_2. Firstly, notice that $\mathrm{Vol}(\mathbf{y}_1, \mathbf{y}_2)$ is increasing with $\mathrm{dist}(c_1, c_2)$ and so attains its minimum value when $\mathrm{dist}(c_1, c_2) = \alpha r$ in this set. Secondly, notice that for the term $(1 - \Phi_{2r}(\mathbf{y}_1, \mathbf{y}_2))$ to be nonzero we must have $\mathrm{dist}(c_1, c_2) \le 4r$. Inserting $\mathrm{dist}(c_1, c_2) = \alpha r$ in the Taylor expansion (8.5) and using the same kind of change of variables yields, we have

$$\begin{aligned}
I_j^{(2)} &:= \int_{\Omega_2} h_r^{\epsilon}(\mathbf{y}_1)h_r^{\epsilon}(\mathbf{y}_2)e^{-n\,\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2)}(1-\Phi_{2r}(\mathbf{y}_1,\mathbf{y}_2))\left|\mathrm{dvol}_g(\mathbf{y}_1,\mathbf{y}_2)\right| \\
&\le \int_{\Omega_2} h_r^{\epsilon}(\mathbf{y}_1)h_r^{\epsilon}(\mathbf{y}_2)e^{-(1-(d\nu' r+\nu r^2))(1+\omega_d\alpha/\omega_{d-1})\Lambda}\left|\mathrm{dvol}_g(\mathbf{y}_1,\mathbf{y}_2)\right| \\
&\le Ce^{-(1+\omega_d\alpha/\omega_{d-1})\Lambda}\int_M \left|\mathrm{dvol}_g(c_1)\right| \\
&\quad \times \int_{r_1}^{r} du_1 \int_{Gr(k,d)} u_1^{k(d-k)}\,d\mu_{k,d}(V_1)\int_{(\mathbb{S}_1(V_1))^{k+1}} u_1^{(k-1)(k+1)}\,\mathrm{dvol}_{(\mathbb{S}_1(V_1))^{k+1}} \\
&\quad \times \int_{\alpha r}^{4r} ds \int_{\mathbb{S}_1(T_{c_1}E)} s^{d-j}\,\mathrm{dvol}_{\mathbb{S}^{d-1}}(w)\int_{r_1}^{r} du_2 \int_{Gr(k-j,d)} u_i^{(k-j)(d-(k-j))}\,d\mu_{k-j,d}(W) \\
&\quad \times \int_{(\mathbb{S}_1(V_2))^{k+1-j}} u_2^{(k-1)(k+1-j)}\,\mathrm{dvol}_{(\mathbb{S}_1(V_2))^{k+1-j}}\; h_r^{\epsilon}(\exp_{c_1}(u_1\mathbf{w}_1))h_r^{\epsilon}(\exp_{c_2}(u_2\mathbf{w}_2)),
\end{aligned}$$

In this case the integration in the radial coordinates yields

$$I_j^{(2)} = \int_{\Omega_2} h_r^{\epsilon}(\mathbf{y}_1)h_r^{\epsilon}(\mathbf{y}_2)e^{-n\,\mathrm{Vol}(\mathbf{y}_1,\mathbf{y}_2)}\left|\mathrm{dvol}_g(\mathbf{y}_1,\mathbf{y}_2)\right| \le C\Lambda^{-4}e^{-\Lambda}e^{-\omega_d\alpha\Lambda/\omega_{d-1}}r^{d(2k+1-j)+j(k-j)+1}.$$

Putting these all together into (8.6) we have

$$\mathbb{E}\{I_j\} \le Cn\Lambda^{2k+1-j}r^{j(k-j)+1}\Lambda^{-4}e^{-\Lambda}\left(e^{-\omega_d\alpha\Lambda/\omega_{d-1}} + \alpha^{d-j+1}\right).$$

■

9 | PROOF OF THE MAIN THEOREM

In this section we combine the results proved in Sections 6–8 to prove the main result - Theorem 1.1. During the proof, we shall use the term “with high probability” (w.h.p.), meaning that the probability goes to 1 as $n \to \infty$.

Proof of Theorem 1.1 We divide the proof into two parts, corresponding to the upper and lower thresholds for the phase transition.

Upper threshold:
Suppose that $\Lambda = \log n + k \log\log n + w(n)$, and take r_0 to satisfy the conditions in Lemma 6.2. Using Lemma 6.2 we have that $\mathbb{E}\{C_k(r, r_0)\} \to 0$ and $\mathbb{E}\{C_{k+1}(r, r_0)\} \to 0$, which implies (using Markov’s inequality) that $\mathbb{P}(C_k(r, r_0) > 0) \to 0$ and $\mathbb{P}(C_{k+1}(r, r_0) > 0) \to 0$. Using similar arguments to the ones used in the proof of Proposition 6.1, we can conclude that $H_k(\mathcal{C}_r) \cong H_k(\mathcal{C}_{r_0})$ w.h.p. since there are no

critical points of index k and $k+1$ in $(r, r_0]$. In addition, from [16] we know that at radius r_0 the union of balls $B_{r_0}(\mathcal{P})$ covers M w.h.p. , and therefore by the Nerve Lemma 2.2 we have that $H_k(\mathcal{C}_{r_0}) \cong H_k(M)$. Thus, we conclude that $H_k(\mathcal{C}_r) \cong H_k(M)$.

Lower threshold:
From Proposition 8.1 we know that w.h.p. $\beta_k^\epsilon(r) > \frac{1}{2}\mathbb{E}\left\{\beta_k^\epsilon(r)\right\}$. If $\Lambda = \log n + (k-2)\log\log n - w(n)$, then from Lemma 7.4 we have that $\mathbb{E}\left\{\beta_k^\epsilon(r)\right\} \geq a_k e^{w(n)} \to \infty$. In addition, from Lemma 7.3 we know that $\beta_k(r) \geq \beta_k^\epsilon(r)$. Therefore, we have that w.h.p. $\beta_k(r) > \beta_k(M)$, which implies that $H_k(\mathcal{C}_r) \ncong H_k(M)$ ∎

10 | DISCUSSION

Homological connectivity is one of the fundamental properties of random simplicial complexes. In this paper we showed that the phase transition discovered in Bobrowski [7], applies to any compact Riemannian manifold. This is due to the fact that any Riemannian metric can be locally well approximated by an Euclidean one. Our results suggest that the phase transition for homological connectivity exhibited by the Čech complex should occur at a critical value of Λ which is inside the interval

$$[\log n + (k-2)\log\log n, \log n + k\log\log n].$$

We note that the first and second-order terms are identical to those in [7] describing the flat torus, while the lower order term could be different and depend on the Riemannian metric. Clearly, our work here is not done, as we are yet to have found the exact threshold for homological connectivity (for either the torus or general Riemannian manifolds). This, however, is left as future work. In particular, we propose the following conjecture which will be addressed in future work.

Conjecture 10.1 *Let $w(n) \to \infty$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(H_k(\mathcal{C}_r) \cong H_k(M)\right) = \begin{cases} 1 & \Lambda = \log n + (k-1)\log\log n + w(n) \\ 0 & \Lambda = \log n + (k-1)\log\log n - w(n). \end{cases}$$

Our intuition for this conjecture is the following. In random graph models (eg, Erdös-Rényi, and geometric), the obstruction to connectivity are isolated vertices. More specifically, in [15] it is shown that the graph becomes connected exactly at the point where the last isolated vertex connects to another vertex. A similar observation appeared later in both the Linial-Meshulam model [30] as well as the random clique complexes [25]. In both models it was shown that the obstruction to homological connectivity are "isolated" or "uncovered" k-faces (k-simplexes that are not a face of any $(k+1)$-simplexes). These results suggest that the same phenomenon should occur in the random Čech complex. Condition 2 in (6.1) implies that every critical point of index k introduces an uncovered k-face. Thus, a possible candidate for the homological connectivity threshold is the point where the last critical point of index k appears. Lemma 6.2 strongly suggests that this threshold is $\Lambda = \log n + (k-1)\log\log n$.

A final remark—in this paper we focused on extending the homological connectivity result from [7] to compact Riemannian manifolds. However, the methods we used in this paper could be used to translate any of the previous statements made for random Čech (and also Vietoris-Rips) complexes [3, 5, 24, 27, 43] from the Euclidean setup to Riemannian manifolds. For example, we could provide formulae for the expected Betti numbers in the sparse regime ($\Lambda \to 0$), as well as prove a central limit theorem (CLT) in either the sparse or the thermodynamic ($\Lambda = \lambda \in (0, \infty)$) regime. For the sake of keeping this paper at a reasonable length, we did not include these statements here.

ACKNOWLEDGEMENTS

The authors are very grateful to Robert Adler, Robert Bryant, Mark Stern and Shmuel Weinberger for helpful conversations about this and related work.

REFERENCES

1. R. J. Adler, O. Bobrowski, and S. Weinberger, *Crackle: The homology of noise*, Discrete Comput. Geom. **52** (2014), no. 4, 680–704.
2. S. Balakrishnan, A. Rinaldo, D. Sheehy, A. Singh, and L. A. Wasserman, *Minimax rates for homology inference*. AISTATS, vol. 9, 2012, pp. 206–207.
3. O. Bobrowski and R. J. Adler, *Distance functions, critical points, and the topology of random Čech complexes*, Homology Homotopy Appl. **16** (2014), no. 2, 311–344.
4. O. Bobrowski, M. Kahle, and P. Skraba, *Maximally persistent cycles in random geometric complexes*, Ann. Appl. Probab. **27** (2017), no. 4, 2032–2060.
5. O. Bobrowski and S. Mukherjee, *The topology of probability distributions on manifolds*, Probab. Theory Related Fields. **161** (2014), no. 3–4, 651–686.
6. O. Bobrowski, S. Mukherjee, and J. E. Taylor, *Topological consistency via kernel estimation*, Bernoulli. **23** (2017), no. 1, 288–328.
7. O. Bobrowski and S. Weinberger, *On the vanishing of homology in random Čech complexes*, Random Structures Algorithms. **51** (2017), no. 1, 14–51.
8. K. Borsuk, *On the imbedding of systems of compacta in simplicial complexes*, Fund. Math. **35** (1948), no. 1, 217–234.
9. G. Carlsson, *Topology and data*, Bull. Am. Math. Soc. **46** (2009), no. 2, 255–308.
10. F. Chazal, D. Cohen-Steiner, and A. Lieutier, *A sampling theory for compact sets in Euclidean space*, Discrete Comput. Geom. **41** (2009), no. 3, 461–479.
11. M. P. Do Carmo, *Riemannian Geometry. Mathematics: Theory & Applications*, Birkhuser, Boston, 1992.
12. T. K. Duy, Y. Hiraoka, and T. Shirai, Limit Theorems for Persistence Diagrams. arXiv:1612.08371 [math].
13. H. Edelsbrunner, A. Nikitenko, and M. Reitzner, *Expected sizes of Poisson-Delaunay mosaics and their discrete Morse functions*, Adv. Appl. Probab. **49** (2017), no. 3, 745–767.
14. P. Erdös and A. Rényi, *On random graphs*, Publ. Math. Debrecen. **6** (1959), 290–297.
15. P. Erdös and A. Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61.
16. L. Flatto and D. J. Newman, *Random coverings*, Acta Math. **138** (1977), no. 1, 241–264.
17. V. Gershkovich and H. Rubinstein, *Morse theory for Min-type functions*, Asian J. Math. **1** (1997), 696–715.
18. E. N. Gilbert, *Random plane networks*, J. Soc. Ind. Appl. Math. **9** (1961), no. 4, 533–543.
19. M. Gromov, *Curvature, diameter and Betti numbers*, Comment. Math. Helv. **56** (1981), no. 1, 179–195.
20. K. Grove and K. Shiohama, *A generalized sphere theorem*, Ann. Math. **106** (1977), no. 1, 201–211.
21. A. Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, 2002.
22. H. Hopf and W. Rinow, *Ueber den Begriff der vollständigen differentialgeometrischen Fläche*, Comment. Math. Helv. **3** (1931), no. 1, 209–225.
23. M. Kahle, *Topology of random clique complexes*, Discrete Math. **309** (2009), no. 6, 1658–1671.
24. M. Kahle, *Random geometric complexes*, Discrete Comput. Geom. **45** (2011), no. 3, 553–573.
25. M. Kahle, *Sharp vanishing thresholds for cohomology of random flag complexes*, Ann. Math. **179** (2014), no. 3, 1085–1107.
26. M. Kahle, *Topology of random simplicial complexes: a survey*, AMS Contemp. Math. **620** (2014), 201–222.
27. M. Kahle and E. Meckes, *Limit the theorems for Betti numbers of random simplicial complexes*, Homology Homotopy Appl. **15** (2013), no. 1, 343–374.
28. S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, vol. 1, Wiley Classics Library, New York, 1996.
29. S. Lang, *Differential and Riemannian Manifolds*, Springer, New York, 1995.
30. N. Linial and R. Meshulam, *Homological connectivity of random 2-complexes*, Combinatorica. **26** (2006), no. 4, 475–487.

31. R. Meshulam and N. Wallach, *Homological connectivity of random k-dimensional complexes*, Random Structures Algorithms. **34** (2009), no. 3, 408–417.
32. R. E. Miles, *Isotropic random simplices*, Adv. Appl. Probab. **3** (1971), no. 02, 353–382.
33. J. W. Milnor, *Morse theory*, Princeton University Press, Princeton, NJ, 1963.
34. J. R. Munkres, *Elements of Algebraic Topology*, vol. 2, Addison-Wesley, Reading, 1984.
35. P. Niyogi, S. Smale, and S. Weinberger, *Finding the homology of submanifolds with high confidence from random samples*, Discrete Comput. Geom. **39** (2008), no. 1–3, 419–441.
36. P. Niyogi, S. Smale, and S. Weinberger, *A topological view of unsupervised learning from noisy data*, SIAM J. Comput. **40** (2011), no. 3, 646–663.
37. T. Owada and R. J. Adler, *Limit theorems for point processes under geometric constraints (and topological crackle)*, Ann. Probab. **45** (2017), no. 3, 2004–2055.
38. M. Penrose, *Random Geometric Graphs*, vol. 5, Oxford University Press, Oxford, 2003.
39. P. Peterson, *Riemannian Geometry*, vol. 171, Springer, New York, 2006.
40. J. Viaclovsky, *Topics in Riemannian Geometry*. Notes of Curse Math 865, Fall 2011.
41. L. Wasserman, *Topological data analysis*, Annu. Rev. Stat. Appl. **5** (2018).
42. D. Yogeshwaran and R. J. Adler, *On the topology of random complexes built over stationary point processes*, Ann. Appl. Probab. **25** (2015), no. 6, 3338–3380.
43. D. Yogeshwaran, E. Subag, and R. J. Adler, *Random geometric complexes in the thermodynamic regime*, Probab. Theory Related Fields (2016), 1–36.
44. A. Zomorodian, *Topological data analysis*, Adv. Appl. Comput. Topol. **70** (2007), 1–39.

How to cite this article: Bobrowski O, Oliveira G. Random Čech complexes on Riemannian manifolds. *Random Struct Alg.* 2019;54:373–412. https://doi.org/10.1002/rsa.20800

APPENDIX A: PALM THEORY FOR POISSON PROCESSES

The following theorem will be very useful when computing expectations related to Poisson processes.

Theorem A.1 (Palm theory for Poisson processes) *Let (X,ρ) be a metric space, $f : X \to \mathbb{R}$ be a probability density on X, and let $\mathcal{P}_n$ be a Poisson process on X with intensity $\lambda_n = nf$. Let $h(\mathcal{Y},\mathcal{X})$ be a measurable function defined for all finite subsets $\mathcal{Y} \subset \mathcal{X} \subset X^d$ with $|\mathcal{Y}| = k$. Then*

$$\mathbb{E}\left\{\sum_{\mathcal{Y}\subset\mathcal{P}_n} h(\mathcal{Y},\mathcal{P}_n)\right\} = \frac{n^k}{k!}\mathbb{E}\left\{h(\mathcal{Y}',\mathcal{Y}'\cup\mathcal{P}_n)\right\}$$

where $\mathcal{Y}'$ is a set of k iid points in X with density f, independent of $\mathcal{P}_n$.

For a proof of Theorem .1, see for example [38]. We shall also need the following corollary, which treats second moments:

Corollary A.2 *With the notation above, assuming $|\mathcal{Y}_1| = |\mathcal{Y}_2| = k$,*

$$\mathbb{E}\left\{\sum_{\substack{\mathcal{Y}_1,\mathcal{Y}_2\subset\mathcal{P}_n\\ |\mathcal{Y}_1\cap\mathcal{Y}_2|=j}} h(\mathcal{Y}_1,\mathcal{P}_n)h(\mathcal{Y}_2,\mathcal{P}_n)\right\} = \frac{n^{2k-j}}{j!((k-j)!)^2}\mathbb{E}\left\{h(\mathcal{Y}_1',\mathcal{Y}'\cup\mathcal{P}_n)h(\mathcal{Y}_2',\mathcal{Y}'\cup\mathcal{P}_n)\right\}$$

where $\mathcal{Y}' = \mathcal{Y}_1'\cup\mathcal{Y}_2'$ is a set of $2k-j$ iid points in X with density f, independent of $\mathcal{P}_n$, and $\left|\mathcal{Y}_1'\cap\mathcal{Y}_2'\right| = j$.

For a proof of this corollary, see for example [3].

APPENDIX B: PROOFS FOR SECTION 3.3

This appendix contains the proof of Lemmas 3.1, 3.4 and corollary 3.2.

Proof of Lemma 3.1 Using normal coordinates $(x^1, \dots, x^d)$ in a neighborhood of the point $p \in M$ we have $\mathrm{dvol}_g = \sqrt{\det(g_{ij})}\ \mathrm{dvol}_{g_E} = dr \wedge r^{d-1}\sqrt{\det(g_{ij})}\,\mathrm{dvol}_{\mathbb{S}^{d-1}}$, where $\mathbb{S}^{d-1}$ denotes the unit $(d-1)$-sphere in the Euclidean metric $g_E = \delta_{ij}dx^i \otimes dx^j$, and $\mathrm{dvol}_{\mathbb{S}^{d-1}}$ the volume form of the induced by the round metric on $\mathbb{S}^{d-1}$. Thus, we have that $\mathrm{dvol}_{S_r(p)} = \sqrt{\det(g_{ij})}r^{d-1}\ \mathrm{dvol}_{\mathbb{S}^{d-1}}$, and using (3.2) we have

$$\mathrm{dvol}_{S_r(p)} = r^{d-1}\left(1 - \frac{Ric_{ij}}{3}x^ix^j + O(|x|^3)\right)\mathrm{dvol}_{\mathbb{S}^{d-1}}\ .$$

For any $\nu > 0$ we can find $r_\nu(p) > 0$, such that for all $r \le r_\nu(p)$

$$\left|\mathrm{dvol}_{S_r(p)}\right| \le r^{d-1}\left(1 + \frac{|Ric| + \nu}{3}r^2\right)|\mathrm{dvol}_{\mathbb{S}^{d-1}}|\,,$$

and similarly for the lower bound. Moreover, as the metric g is smooth (ie, C^∞) we can take $r_\nu(p)$ to depend smoothly on $p \in M$. Then, we take $r_\nu = \min_{p\in M} r_\nu(p)$, which is achieved and positive as M is compact and one can chose $r_\nu(p)$ to vary continuously with $p \in M$, see remark .1 below. ■

Remark B.1. Around any point $p \in M$ we can pick geodesic normal coordinates $(x^1, \dots, x^d)$, valid in $B_r(p)$, for $r < \mathrm{inj}(p)$, the injectivity radius at p[4]. Then, for $r < \mathrm{inj}(p)$ we may write $\mathrm{dvol}_{S_r(p)} = r^{d-1}f(x^1, \dots, x^d)\,\mathrm{dvol}_{\mathbb{S}^{d-1}}$. Moreover, for small r the Taylor expansion of f is such that

$$f(x^1, \dots, x^d) = 1 - \frac{Ric_{ij}}{3}x^ix^j - \frac{1}{9}\nabla_k Ric_{ij}x^ix^jx^k + \dots .$$

Then for any $\nu > 0$, we have

$$f(x^1, \dots, x^d) \le 1 + \frac{\left|Ric_p\right| + \nu}{3}r^2,$$

provided that

$$\begin{aligned} 0 &\le \left(1 + \frac{\left|Ric_p\right| + \nu}{3}r^2\right) - \left(1 - \frac{Ric_{ij}}{3}x^ix^j - \frac{1}{9}\nabla_k Ric_{ij}x^ix^jx^k + \dots\right) \\ &\le \frac{\left|Ric_p\right| r^2 - Ric_{ij}x^ix^j + \nu r^2}{3} + c_3(p)r^3 \\ &\le \frac{\nu r^2}{3} + (c_3(p) + 1)r^3 \end{aligned}$$

where $c_3(p) = \sup_{|v|=1}\left(\frac{1}{9}(\nabla_v Ric)(v, v)\right)$. Solving this inequality for r we find that it is enough that

$$r \le r_\nu(p) = \min\left\{\frac{\nu}{3(|c_3(p)| + 1)}, \mathrm{inj}(p)\right\}.$$

[4]The injectivity radius at p is the supremum of the values of r such that $\exp_p : B_r(0) \subset T_pM \to M$ is injective.

As g is a smooth metric, both $c_3(\cdot)$ and $\mathrm{inj}(\cdot)$ are continuous and so the minimum of the two is continuous. It will be important for the proof of the previous result that one can choose this $r_\nu(p)$ to vary continuously with p.

Proof of Corollary 3.2 For every point $p \in M$, we have from (3.2) that $\sqrt{\det(g_{ij})} = 1 - \frac{Ric_{ij}}{3}x^i x^j + O(|x|^3)$. Thus, if we take $\nu(p) = |Ric(p)| + \delta$ for some fixed $\delta > 0$, then we can find $r_\nu(p) > 0$ such that Lemma 3.3 holds. To complete the proof we take $\nu = \min_{p\in M} \nu(p)$, and $r_\nu = \max_{p\in M} r_\nu(p)$. ■

Proof of Lemma 3.4 Let p_1, p_2 be at a sufficiently small distance from each other, and let p be the midpoint in the minimizing geodesic connecting p_1 and p_2. For $\mathrm{dist}(p_1, p_2) < 2s$ we have $B_s(p_1) \cup B_s(p_2) \subset B_{2s}(p)$. Pick normal coordinates $(x^1, \dots, x^d)$ centered at p and valid in a ball of radius $2s$ centered around it. In these coordinates, the metric g is within $O(s^2)$ of the Euclidean metric and its distance function dist_g is within $O(s)$ of the Euclidean one dist_E, that is, $\mathrm{dist}_g / \mathrm{dist}_E = 1 + O(s)$. Hence, there exists $\nu_p > 0$ such that

$$(1 - \nu_p s)\rho^E_{p_i} \le \rho_{p_i} \le (1 + \nu_p s)\rho^E_{p_i}, \quad i = 1, 2.$$

From this, it immediately follows that for any $r \le s$

$$\left(B^E_{(1-\nu_p s)r}(p_1) \cup B^E_{(1-\nu_p s)r}(p_2)\right) \subset (B_r(p_1) \cup B_r(p_2)) \subset \left(B^E_{(1+\nu_p s)r}(p_1) \cup B^E_{C(1+\nu_p s)r}(p_2)\right).$$

The result then follows from using the compactness of M, and maximizing ν_p over all $p_1, p_2 \in M$ within distance $2s$ from each other. ■

APPENDIX C: A CONVEXITY RESULT

In this section we present a result regarding equidistant hypersurfaces, that is, set of points which are equidistant to a fixed pair points. In the Euclidean case this is simply a hyperplane and thus a totally geodesic submanifold. In the following we prove that in a Riemannian manifolds, if two points are sufficiently close, then their equidistant hypersurface is approximately totally geodesic. This result is used in the proof of Lemma 4.2, where we restrict a strictly convex function to a small open set in an equidistant hypersurface. Our result here guarantees that the restricted function is locally convex and thus admits a unique minimum.

Lemma C.1 *Let $p \in M$, then there exists $r_p > 0$ that satisfies the following. For $p_1, p_2 \in B_{r_p}(p)$ let*

$$E_{p_1,p_2} = \{x \in M \mid \mathrm{dist}(p_1, x) = \mathrm{dist}(p_2, x)\}$$

be the equidistant hypersurface between p_1 and p_2. Then, $\mathrm{dist}^2(p, \cdot)$ restricted to $E_{12} \cap B_{r_p}(p)$ is strictly convex.

Proof Fix normal coordinates $x = (x^1, \dots, x^d)$ in a δ-neighborhood of p with $x(p) = 0$ and let

$$s_\delta(x) = \delta x,$$

for $(x^1)^2 + \dots (x^d)^2 \leq 1$. Using the pullback $s_\delta^* g$ of the metric g to the Euclidean unit ball, we define the metric $g_\delta := \delta^{-2} s_\delta^* g$. In the coordinates $(x^1, \dots, x^d)$ we can write $g_\delta = g_{ij}^\delta dx^i \otimes dx^j$ with

$$g_{ij}^\delta(x) = g_{ij}(\delta x) = \delta_{ij} + \frac{\delta^2}{3} R_{iklj} x^k x^l + O(\delta^3). \tag{C.1}$$

Moreover, given $p_1 \in \mathrm{B}_\delta(p)$, we can write the distance function with respect to g_δ in the coordinates $(x^1, \dots, x^d)$ as

$$\delta \operatorname{dist}_\delta(\delta^{-1} p_1, \delta^{-1} x) = \operatorname{dist}(p_1, x), \tag{C.2}$$

where $\operatorname{dist}(p_1, x)$ denotes the distance function of g in the same coordinates. The reason for using this rescaling is twofold:

- The metric g_δ converges uniformly to the Euclidean metric, in the sense that $g_{ij}^\delta \to \delta_{ij}$ as $\delta \to 0$. This is a direct consequence of Equation (C.1). Moreover, this way the coordinates x are fixed and their range is not shrinking.
- Let $p_1, p_2 \in B_\delta(p)$. The equidistant hypersurfaces E_{p_1,p_2} and $E^\delta_{\delta^{-1}p_1, \delta^{-1}p_2}$ of the metrics g and g_δ inside $B_1(p)$ are related by $E^\delta_{\delta^{-1}p_1, \delta^{-1}p_2} = \delta^{-1} E_{p_1,p_2}$. This claim is immediate from Equation (C.2).

We wish prove that, restricted to $E_{p_1,p_2} \cap B_\delta(p)$, the function $\operatorname{dist}(p, \cdot)$ is convex. Recall that $p = 0$ in this coordinates. Then, given the relations above and the fact that $\operatorname{dist}_\delta(p, \cdot) = \delta^{-1} \operatorname{dist}(p, \delta \cdot)$ it will be enough to show that $\operatorname{dist}_\delta(p, \cdot)$ restricted to $E^\delta_{\delta^{-1}p_1, \delta^{-1}p_2} \cap B_1(p)$ is strictly convex.
From the first bullet above it follows that $\operatorname{dist}_\delta(p, \cdot)$ varies smoothly with δ and agrees with the Euclidean distance function $\sqrt{(x^1)^2 + \dots + (x^d)^2}$ when $\delta = 0$. Hence, using a Taylor expansion in δ we have

$$\operatorname{dist}^2_\delta(p, x) = (x^1)^2 + \dots (x^d)^2 + \delta f(x, \delta),$$

where $f(x, \delta)$ is smooth in both variables, and uniformly bounded in a neighborhood of $\delta = 0$. Similarly, the equidistant hypersurfaces vary smoothly with δ and at $\delta = 0$ must coincide with the Euclidean equidistant hyperplane between p_1 and p_2. Using a Taylor expansion we have

$$\operatorname{dist}^2_\delta(p_1, x) - \operatorname{dist}^2_\delta(p_2, x) = H_{12}(x) + \delta h(x, \delta),$$

where H_{12} is an affine function and $h(x, \delta)$ is a smooth function, uniformly bounded around $\delta = 0$. Restricting $\operatorname{dist}^2_\delta(p, x)$ to $E^\delta_{\delta^{-1}p_1, \delta^{-1}p_2} = E^\delta_{12}$ and using the induced metric on E^δ_{12}, we can regard the Hessian H_δ of $\operatorname{dist}^2_\delta$ as an endomorphism of TE^δ_{12}. For $\delta = 0$ we have a restriction of a strictly convex function to a totally geodesic submanifold (an hyperplane), and therefore H_0 is positive definite. Thus, for small δ we have

$$H_\delta = H_0 + \delta H_1 + O(\delta^2),$$

and there exists $r_p > 0$ such that for $\delta < r_p$ we have that H_δ is still positive definite, which implies that the restriction of $\operatorname{dist}^2_\delta(p, x)$ to E^δ_{12} is strictly convex. ∎

APPENDIX D: AN EXCESS INEQUALITY

In this section we prove an excess inequality for certain sets of points $\mathcal{Y}$ in a sufficiently small normal ball. This inequality is used in the proof of Lemma 7.3. Recall that using normal coordinates (3.1) centered at a point $p \in M$, the metric is approximately Euclidean to first order. Hence, ρ_p^2 approaches the squared Euclidean distance in a small enough neighborhood of p. In the Euclidean space, the Hessian matrix satisfies $\nabla^2 \rho_p^2 = 2I$, where I is the identity matrix. Therefore, for a Riemannian distance function, if $r > 0$ is sufficiently small then there exists $A(r) > 0$ such that on $B_r(p)$ we have

$$\nabla^2 \rho_p^2 \geq 2A(r)I. \tag{D.3}$$

Moreover, the constant $A(r)$ converges to 1 as $r \to 0$. We note that the constant A can be taken to be greater than 1 if g has negative sectional curvatures at p, and it must be taken to be less than 1 if one of the sectional curvatures is positive.

Lemma D.1 *There exists a continuous function $c_g : (0, r_{\max}] \to \mathbb{R}$, that depends only the metric g, and satisfies the following. Let $\mathcal{Y} \subset \mathcal{P}$ be such that $E_{r_{\max}}(\mathcal{Y}) \neq \emptyset$ (and $c(\mathcal{Y})$ is well defined) and such that $0 \in \Delta(\mathcal{Y})$. Then for every $r \in (\rho(\mathcal{Y}), r_{\max})$ and $x \in B_r^{\cap}(\mathcal{Y})$, we have*

$$\operatorname{dist}(c(\mathcal{Y}), x) \leq c_g(r)\sqrt{r^2 - \rho^2(\mathcal{Y})}.$$

In addition, we have that $\lim_{r \to 0} c_g(r) = 1$.

Proof Let $\mathcal{Y} = \{y_1, \dots, y_k\}$, $x \in B_r^{\cap}(\mathcal{Y})$, and $\ell = \operatorname{dist}(c(\mathcal{Y}), x)$. In addition, let $\gamma : [0, \ell] \to M$ be the minimizing geodesic from $c(\mathcal{Y})$ to x, using the arc-length parametrization. For each $y_i \in \mathcal{Y}$ consider the squared distance from y_i to a point $\gamma(t)$ on that geodesic (see Figure D1). Using the Hessian inequality (D.3), and recalling that γ is parametrized with respect to arc-length, we have

$$\frac{1}{2}\frac{d^2}{dt^2}(\rho_{y_i}^2 \circ \gamma)(t) = \frac{1}{2}(\nabla^2_{\dot{\gamma},\dot{\gamma}} \rho_{y_i}^2)(\gamma(t)) \geq A(r)\,|\dot{\gamma}(t)|^2 = A(r), \quad 1 \leq i \leq k,$$

Integrating this inequality from 0 to t, using the fundamental theorem of calculus and multiplying by 2 yields $\frac{d}{dt}(\rho_{y_i}^2 \circ \gamma)(t) \geq \langle \nabla \rho_{y_i}^2, \dot{\gamma}(0)\rangle + 2At$. Applying the fundamental theorem of calculus again we have

$$\rho_{y_i}^2(\gamma(t)) - \rho_{y_i}^2(\gamma(0)) \geq \langle \nabla \rho_{y_i}^2, \dot{\gamma}(0)\rangle t + A(r)t^2.$$

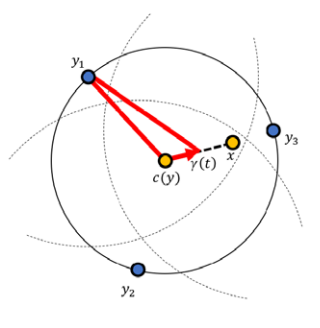

FIGURE D1 The temporary construction of geodesics used in the proof of Lemma .1. Here $\mathcal{Y} = \{y_1, y_2, y_3\}$, with $c(\mathcal{Y})$ as its center. $\gamma(t)$ is a geodesic connecting $c(\mathcal{Y})$ and $x \in B_r^{\cap}(\mathcal{Y})$ [Colour figure can be viewed at wileyonlinelibrary.com]

Recall that $\gamma(\ell) = x$, $\gamma(0) = c(\mathcal{Y})$, $\rho_{y_i}(x) \leq r$, $\rho_{y_i}(c(\mathcal{Y})) = \rho(\mathcal{Y})$, and $\ell = \operatorname{dist}(c(\mathcal{Y}), x)$, then putting everything into the last inequality we have

$$A(r)\,\operatorname{dist}^2(c(\mathcal{Y}), x) \leq r^2 - \rho^2(\mathcal{Y}) - \langle \nabla \rho^2_{y_i}, \dot{\gamma}(0)\rangle l, \quad 1 \leq i \leq k. \tag{D.4}$$

Finally, we use the assumption that $0 \in \Delta(\mathcal{Y}) \subset T_{c(\mathcal{Y})}M$, which implies that there exist $\{\alpha_i\}_{i=1}^k$ such that $\alpha_i \in [0,1]$, $\sum_i \alpha_i = 1$, and $\sum_{i=1}^k \alpha_i \nabla \rho^2_{y_i} = 0$. Multiplying each of the k inequalities (D.4) by the corresponding α_i and summing them up yields

$$\operatorname{dist}(c(\mathcal{Y}), x)^2 \leq (r^2 - \rho^2(\mathcal{Y}))/A^2(r).$$

To complete the proof, we set $c_g(r) = 1/\sqrt{A(r)}$. The continuity of $c_g(r)$ and the fact that is approaches 1 as $r \to 0$ follow from the properties of $A(r)$. ∎