Strange Random Topology of the Circle

Uzu Lim
Mathematical Institute, University of Oxford, UK

Abstract

A paradigm in topological data analysis asserts that persistent homology should be computed to recover the homology of a data manifold. But could there be more to persistent homology? In this paper I bound probabilities that a random Čech complex built on a circle attains high-dimensional topology. This builds on the known result that any nerve complex of circular arcs has the homotopy type of a bouquet of spheres. We observe a phase transition going from one 1-sphere, bouquet of 2-spheres, one 3-sphere, bouquet of 4-spheres, and so on. Furthermore, the even-dimensional Betti numbers become arbitrarily large over shrinking intervals. Our main tool is an exact computation of the expected Euler characteristic, combined with constraints on homotopy types. The systematic behaviour we observe cannot be regarded as a "topological noise", and calls for deeper investigations from the TDA community.

2012 ACM Subject Classification Mathematics of computing → Algebraic topology

Keywords and phrases Topological data analysis, persistent homology, stochastic topology

Digital Object Identifier 10.4230/LIPIcs.SoCG.2024.70

Funding Uzu Lim is supported by BMS Centre for Innovation and Translational Research Europe and Korea Foundation for Advanced Studies.

Acknowledgements The author would like to thank Vidit Nanda and Harald Oberhauser for their contributions during initial stages of this research. The author is grateful to Henry Adams and Tadas Temčinas for valuable discussions.

1 Introduction

A conventional wisdom in topological data analysis says the following: if we construct a simplicial complex from a random sample drawn from a manifold, then the topology of the simplicial complex approximates the topology of the manifold. Indeed this is true if we scale down the connectivity radius smaller as the sample size grows larger, but what happens when the connectivity radius stays the same?

We study the random topology of the circle and characterise its unexpected high-dimensional topology. Building on the previous discovery in [4] that a nerve complex of circular arcs must be homotopy equivalent to a bouquet of spheres[1] $\vee^a \mathbb{S}^b$, we constrain probabilities that these homotopy types are achieved. We achieve this by computing expected Euler characteristic precisely and pair this with constraints on homotopy type. In [4], it was also shown that if a nerve complex of circular arcs is homotopy equivalent to $\vee^a \mathbb{S}^b$, then only $a = 1$ is allowed for odd b, while all $a \geq 1$ are allowed for even b. Indeed, this corresponds to our discovery that there is a phase transition going from one 1-sphere, bouquet of 2-spheres, one 3-sphere, bouquet of 4-spheres, and so on. Now let's describe the setup and Theorem A.

[1] We take the convention that for a topological space K, we define the 0-th wedge sum $\vee^0 K = *$, the singleton point set, and the 1st wedge sum $\vee^1 K = K$ itself.

40th International Symposium on Computational Geometry (SoCG 2024).
Editors: Wolfgang Mulzer and Jeff M. Phillips; Article No. 70; pp. 70:1–70:17
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Setup. Define the circle $\mathbb{S}^1$ as the quotient space $\mathbb{S}^1 = [0,1]/\sim$, which is an interval of length 1 glued along endpoints: $0 \sim 1$. A bouquet of spheres $\vee^a \mathbb{S}^k$ is defined as the wedge sum of a copies of $\mathbb{S}^k$.
For a positive integer n, let $\mathbf{X}_n$ be the i.i.d. (independently and identically distributed) sample of size n, drawn uniformly from $\mathbb{S}^1$. The Čech complex of filtration radius $\leq r$ is denoted by $\check{\mathrm{C}}(\mathbf{X}_n, r)$. In constructing a Čech complex, we always use the intrinsic topology of the circle, i.e. *the Čech complex is a nerve complex constructed by taking arcs as open sets.*
We denote the expected Euler characteristic and expected Betti number as follows:

$$\bar{\chi}(n,r) = \mathbb{E}\Big[\chi(\check{\mathrm{C}}(\mathbf{X}_n, r))\Big], \quad \bar{b}_k(n,r) = \mathbb{E}\Big[\dim H_k(\check{\mathrm{C}}(\mathbf{X}_n, r))\Big]$$

▶ **Theorem A** (Expected Euler Characteristic)**.** The following are true.

(1) $\bar{\chi}(n,r)$ is a continuous piecewise-polynomial function in r, given explicitly as follows for $n > 0$ and $r \in (0,1)$:

$$\bar{\chi}\left(n, \frac{1-r}{2}\right) = \sum_{k=1}^{\lfloor 1/r \rfloor} \binom{n}{k} (1-kr)^{k-1} (kr)^{n-k}$$

(2) Normalised Euler charactreristics have the following maximum values for all $k \geq 0$:

$$\lim_{n\to\infty} \max_{r \in I_k} \frac{\bar{\chi}(n,r)}{n} = \frac{(k/e)^k}{(k+1)!}, \quad I_k = \left(\frac{k}{2k+2}, \frac{k+1}{2k+4}\right)$$

(3) Let $k \geq 1$. Given $\epsilon > 0$, the following uniform bounds hold for all $r \in \left[\frac{k}{2k+2}, \frac{k+1}{2k+4}\right)$ when n is sufficiently large:

$$\frac{\bar{\chi}(n,r)}{n} - \epsilon \leq \frac{\bar{b}_{2k}(n,r)}{n} \leq \frac{\bar{\chi}(n,r)}{n}$$

Theorem A allows us to plot *exact values* of the expected Euler characteristic curves. The left side of Figure 1 shows graphs of $f_n(r) = n^{-1} \cdot \bar{\chi}(n,r)$, which are *normalised* versions of $\bar{\chi}$. We stress that these curves are *exact* values from the formula in Theorem A1. As n becomes larger, $f_n(r)$ shows peaks that converge to a sequence of narrow spikes, as Theorem A2 predicts. The right side of the figure shows the non-normalised graphs of $\bar{\chi}(n,r)$, where we see that the peaks of Theorem A2 will go to infinity as $n \to \infty$.

Meanwhile we observe that $\chi(\mathbb{S}^{2k+1}) = 0$ and $\chi(\vee^a \mathbb{S}^{2k}) = a+1$, so that only bouquets of even-dimensional spheres contribute to the Euler characteristic. Therefore in Figure 1, limiting spikes indicate contribution from $\vee^a \mathbb{S}^{2k}$ with large a, and the plateaus indicate contribution from the odd-dimensional spheres. Recalling that $\mathbf{X}_n$ is an i.i.d. sample of size n drawn from $\mathbb{S}^1$, these observations are encoded into Theorems B and C:

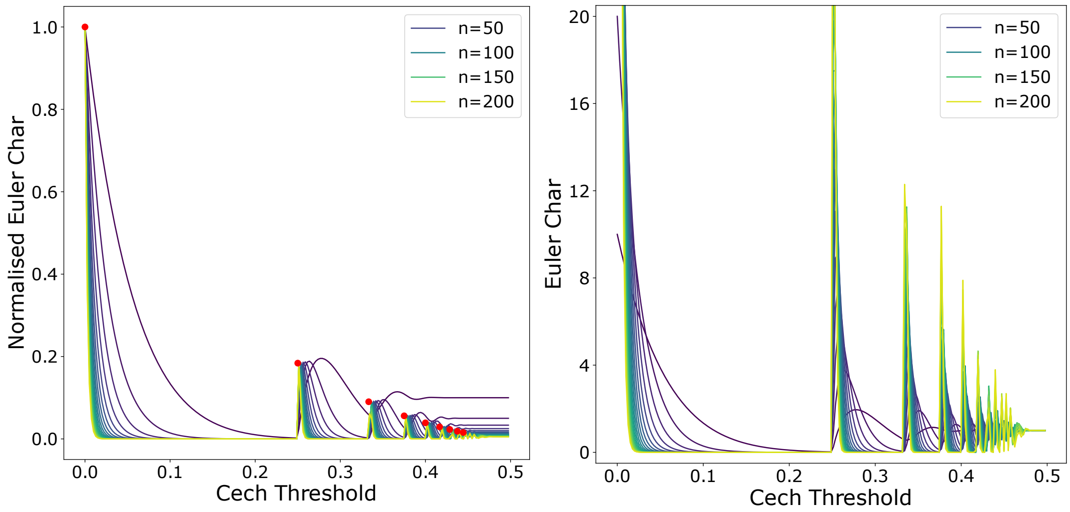

Figure 1 Left: Graphs of *normalised* expected Euler characteristics, $y = n^{-1} \cdot \bar{\chi}(n, x)$, for $n \in \{10, 20, \ldots 200\}$ and $x \in [0, 1]$. Right: Same as left, but we plot $y = \bar{\chi}(n, x)$, which are (un-normalised) expected Euler characteristics. Yellow curves correspond to larger n. Red circles are peaks of the limiting spikes, given by $\left(\frac{k}{2k+2}, \frac{(k/e)^k}{(k+1)!}\right)$ for all $k \geq 0$. Note that $\bar{\chi}$ converges to 1 at threshold 0.5 (Right), because the simplicial complex is contractible at connectivity radius $= 0.5$.

▶ **Theorem B** (Odd Spheres). Let $k \geq 0$ be an integer, and also let $\epsilon, \delta > 0$. Suppose that $|r - \nu_k| \leq \tau_k - \epsilon$. Then for sufficiently large n, the following homotopy equivalence holds with probability at least $1 - \delta$:

$$\check{C}(\mathbf{X}_n, r) \simeq \mathbb{S}^{2k+1}$$

where

$$\nu_k = \frac{2k^2 + 4k + 1}{4(k+1)(k+2)}, \quad \tau_k = \frac{1}{4(k+1)(k+2)}$$

▶ **Theorem C** (Even Spheres). Let $k \geq 2$, $\eta \in (0, 1)$. If n is sufficiently large and r satisfies $|r - \rho_{k,n}| \leq \sigma_{k,\eta}/n$, then the following homotopy equivalence holds with probability at least $\eta \cdot k\omega_k$:

$$\check{C}(\mathbf{X}_n, r) \simeq \vee^a \mathbb{S}^{2k-2}, \quad \text{for some } \frac{(1-\eta)\omega_k \cdot n}{2} \leq a + 1 \leq \frac{n}{k}$$

where

$$\rho_{k,n} = \frac{n(k+1)}{2k(n-1)}, \quad \sigma_{k,\eta} = \frac{(1-\eta)^3 (k\omega_k)^3}{320\sqrt{k+2}}, \quad \omega_k = \frac{(k-1)^{k-1}}{k! e^{k-1}}$$

▶ **Remark 1.** In Theorem B, we note that $\nu_k = \frac{1}{2}(\frac{k+1}{2k+4} + \frac{k}{2k+2})$ and $\tau_k = \frac{1}{2}(\frac{k+1}{2k+4} - \frac{k}{2k+2})$, so that Theorem B covers most of each interval $r \in [\frac{k}{2k+2}, \frac{k+1}{2k+4}]$. In Theorem C, note that the number a appearing in $\vee^a \mathbb{S}^{2k-2}$ is random, and that the theorem constrains the probability that a lies on a certain shrinking interval. To see Theorem C in action, one may simply set $\eta = 1/2$ to obtain results.

▶ **Remark 2.** Although all of the above results are proven for the Čech complex of circular arcs on the circle, similar behaviour is observed in the Rips complex constructed on the circle as well. Indeed, modifying Theorem B for the Rips complex immediately yields the following: for the Rips complex constructed from a finite random sample on a circle, all odd-dimensional spheres appear with positive persistence and probability approaching 1. Analogues of Theorems A and C for the Rips complex could not be immediately obtained with methods in this paper.

▶ **Remark 3.** An analogue of Theorem B was proven in Theorem 6.1 of [1], although the setup is slightly different - the authors consider a random process and calculate the expected number of points required for the random homotopy type to be an odd-dimensional sphere.

Structure of the paper. In Section 2 we calculate the expected Euler characteristic and prove its limit behaviours, which yield Theorems A1 and A2. In Section 3 this calculation is paired with constraints on homotopy types, which yields Theorems A3 and Theorem C. In Section 4 we prove Theorem B, by using the classical method of stability of persistence diagram; this section works separately and doesn't use the Euler characteristic method.

Theorem C takes the most work to prove. It is a simplified version of Theorem 3.5, which has a few more parameters that can be tweaked to obtain similar variants of Theorem C. Theorem 3.5 is obtained by combining three ingredients: Propositions 2.5, 3.3, and 3.4.

Related works. The classical result of Hausmann shows that the Vietoris-Rips complex constructed from the manifold with a small scale parameter recovers the homotopy type of the manifold [13]. Another classical result of Niyogi, Smale, Weinberger shows that if a Čech complex of small filtration radius is constructed from a finite random sample of a Euclidean submanifold, then the homotopy type of the manifold is recovered with high confidence [16].

An early work on understanding statistical behaviour of persistent homology was pioneered in [11], where in particular low-dimensional random topology of the circle was studied. After this, much work has been done for recovering topology of a manifold from its finite sample, when connectivity radius is scaled down with the sample size at a specific rate [9, 12, 14, 10]. A central theme of this body of work is the existence of phase transitions when parameters controlling the scaling of connectivity radius are changed. For a survey, see [18] and [8].

In comparison, the setting when connectivity radius is not scaled down with sample size is studied much less. Results on convergence of the topological quantities have been studied [17, 20], but not much attention has been devoted to analysing specific manifolds.

This paper builds on two important works that characterised the Vietoris-Rips and Čech complexes of subsets of the circle: [4] and [1]. Several variants of these ideas were studied, for ellipse [6], regular polygon [7], and hypercube graph [2]. Randomness in these systems were studied using dynamical systems in [5]. One key tool to further study the topology of Vietoris-Rips and Čech complexes arising from a manifold is metric thickening [3]. Using this tool, the Vietoris-Rips complex of the higher-dimensional sphere has been characterised up to small filtration radii [15].

2 Expected Euler characteristic

In this section we compute the expected Euler characteristic precisely. We start with a simple calculation that also briefly considers the Vietoris-Rips complex, but soon after we only work with the Čech complex. Let $\mathrm{VR}(\mathbf{X}_n, r)$ denote the Vietoris-Rips complex of threshold r. The following proposition reduces computation of expected values to the quantities T_k and Q_k, defined below:

▶ **Proposition 2.1.** *For each $n > 0$, let $\mathbf{X}_n$ be the i.i.d. sample drawn uniformly from $\mathbb{S}^1$. Then we have that:*

$$\mathbb{E}[\chi(\mathrm{VR}(\mathbf{X}_n, r))] = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} T_k(r)$$

$$\mathbb{E}[\chi(\check{C}(\mathbf{X}_n, r))] = 1 + \sum_{k=1}^{n} (-1)^{k} \binom{n}{k} Q_k(1-2r)$$

where $T_k(r)$ is the probability that every pair of points in $\mathbf{X}_k$ are within distance r, and $Q_k(r)$ is the probability that open arcs of length r centered at points of $\mathbf{X}_k$ cover $\mathbb{S}^1$. Expectation is taken over the i.i.d. sample $\mathbf{X}_n$.

Proof. Denoting by $s_k(K)$ the number of $(k-1)$-simplices in a simplicial complex K, we have that:

$$\mathbb{E}[s_k(\mathrm{VR}(\mathbf{X}_n, r))] = \binom{n}{k} T_k(r)$$

and thus

$$\mathbb{E}[\chi(\mathrm{VR}(\mathbf{X}_n, r))] = \sum_{k=0}^{n-1} (-1)^k \mathbb{E}[s_k(\mathrm{VR}(\mathbf{X}_n, r))] = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} T_k(r)$$

The relation for the Čech complex is derived in the same way, except we note the following: the probability that arcs of radius r centered at points of $\mathbf{X}_k$ intersects nontrivially is equal to $1 - Q_k(1-2r)$. This is by De Morgan's Law: for any collection of sets $\{U_j \subseteq \mathbb{S}^1\}_{j \in J}$, we have $\cap_{j\in J} U_j = \emptyset$ iff $\cup_{j \in J} U_j^c = \mathbb{S}^1$. In the case of circle (of circumference 1), the complement of a closed arc of radius r is an open arc of length $1-2r$. Applying this logic, we obtain:

$$\mathbb{E}[\chi(\check{\mathrm{C}}(\mathbf{X}_n, r))] = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (1 - Q_k(1-2r))$$

which is easily seen to be the same as the asserted expression (note that $\sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} = 1$.) ◀

The Q_k were computed by Stevens in 1939 [19]. Its proof is reproduced in Section 5 for completeness.

▶ **Theorem 2.2** (Stevens). *If k arcs of fixed length a are independently, identically and uniformly sampled from the circle of circumference 1, then the probability that these arcs cover the circle is equal to the following:*

$$Q_k(a) = \sum_{l=0}^{\lfloor 1/a \rfloor} (-1)^l \binom{k}{l} (1 - la)^{k-1}$$

We then get the following by switching the order of summation:

▶ **Theorem 2.3** (Theorem A1). *Expected Euler characteristic of random Čech complex on a circle of unit circumference obtained from n points and filtration radius* $(1-r)/2$ *is:*

$$\bar{\chi}\left(n, \frac{1-r}{2}\right) = \sum_{k=1}^{\lfloor 1/r \rfloor} \binom{n}{k} (1-kr)^{k-1}(kr)^{n-k}$$

In particular, $\bar{\chi}(n,r)$ *is a continuous piecewise-polynomial function in* r*.*

Proof. Substituting the Q_k expression in, we get:

$$\begin{aligned}
\bar{\chi}\left(n, \frac{1-r}{2}\right) =& 1 + \sum_{k=1}^{n} (-1)^k \binom{n}{k} Q_k(r) \\
=& 1 + \sum_{l=0}^{\lfloor 1/r \rfloor} \sum_{k=1}^{n} (-1)^{k+l} \binom{n}{k} \binom{k}{l} (1-rl)^{k-1} \\
=& \sum_{l=1}^{\lfloor 1/r \rfloor} \sum_{k=1}^{n} (-1)^{k+l} \binom{n}{k} \binom{k}{l} (1-rl)^{k-1}
\end{aligned}$$

where we switched the order of summation in the second equality, and isolating the $l=0$ part cancels out the 1 in the third equality. Noting that $\binom{n}{k}\binom{k}{l} = \binom{n}{l}\binom{n-l}{k-l}$, we further get:

$$\begin{aligned}
\bar{\chi}\left(n, \frac{1-r}{2}\right) &= \sum_{l=1}^{\lfloor 1/r \rfloor} (-1)^l \binom{n}{l} (1-rl)^{-1} \sum_{k=l}^{n} \binom{n-l}{k-l} (rl-1)^k \\
&= \sum_{l=1}^{\lfloor 1/r \rfloor} \binom{n}{l} (1-rl)^{l-1} \sum_{k=0}^{n-l} \binom{n-l}{k} (rl-1)^k \\
&= \sum_{l=1}^{\lfloor 1/r \rfloor} \binom{n}{l} (1-rl)^{l-1} (rl)^{n-l}
\end{aligned}$$

◀

We now characterise limit behaviour of the formula obtained above, seen as spikes in Figure 1. The main idea is that only one summand in the expected Euler characteristic contributes mainly to the spike, and this is a polynomial term that can be studied with calculus. Indeed, the main idea is simply to study the calculus of functions of the form $f(t) = t^a(1-t)^b$ for some integers (a,b). The (a) and (b) of the next Proposition directly implies Theorem A2, and (c) is used for proving Theorem C. Its proof is postponed to Section 5.

▶ **Proposition 2.4.** *Suppose that* m, n *are integers with* $2 \le m < \sqrt{n}$*. The following holds for* $\bar{\chi}(n,r)$*.*

(a) *The following bounds hold:*

$$a_{m,n} \le \frac{\bar{\chi}(n, s_{m,n})}{n} \le M \le a_{m,n} + b_{m,n}$$

where

$$\begin{aligned}
M &= \max\left\{\frac{1}{n}\bar{\chi}\left(n, \frac{1-r}{2}\right) \,\middle|\, r \in \left(\frac{1}{m+1}, \frac{1}{m}\right)\right\} \\
s_{m,n} &= \frac{(m-1)n}{2(n-1)m} \\
a_{m,n} &= \binom{n}{m}\frac{(m-1)^{m-1}(n-m)^{n-m}}{n(n-1)^{n-1}} \\
b_{m,n} &= en^{m-1}\left(1 - \frac{1}{m+1}\right)^{n-1}
\end{aligned}$$

(b) *We have the following limits:*

$$\lim_{n\to\infty} a_{m,n} = \frac{(m-1)^{m-1}}{m!e^{m-1}}, \qquad \lim_{n\to\infty} b_{m,n} = 0$$

(c) *Suppose additionally that $n > 2m^2$. Then for each $\lambda \in [0,1]$, we have that:*

$$\left| r - \frac{n-m}{(n-1)m} \right| < \frac{(1-\lambda)\sqrt{(m-1)(n-m)}}{m(n-1)^{3/2}} \implies \frac{1}{n}\bar{\chi}\left(n, \frac{1-r}{2}\right) > \lambda a_{m,n}$$

This condition for r in particular satisfies $r \in \left(\frac{1}{m+1}, \frac{1}{m}\right]$.

The following is a modification of (c) above, and this will be used for Theorem C.

▶ **Proposition 2.5.** *Let $m \geq 2$, $\epsilon > 0$. The following holds for sufficiently large n:*

$$r \in [\alpha^-, \alpha^+] \implies \frac{1}{n}\bar{\chi}\left(n, \frac{1-r}{2}\right) \in \Big[(1-\epsilon)\omega_m, (1+\epsilon)\omega_m\Big]$$

where

$$\alpha^{\pm} = \frac{n-m}{(n-1)m}\left(1 \pm \frac{\epsilon\sqrt{m-1}}{n}\right), \qquad \omega_m = \frac{(m-1)^{m-1}}{m!e^{m-1}}$$

Proof. This follows directly from the previous Proposition. $\alpha^{\pm}$ are slight relaxations of the interval in (c), where we set $\lambda = 1 - \epsilon$:

$$\left[\frac{n-m}{(n-1)m} - \epsilon R_1, \frac{n-m}{(n-1)m} + \epsilon R_1\right] \supseteq \left[\frac{n-m}{(n-1)m}(1-\epsilon R_2), \frac{n-m}{(n-1)m}(1+\epsilon R_2)\right]$$

where $R_1 = \dfrac{\epsilon\sqrt{(m-1)(n-m)}}{m(n-1)^{3/2}}, \quad R_2 = \dfrac{\sqrt{m-1}}{n}$ ◀

3 Random homotopy types

In this section we prove Theorems A3 and C, by pairing the calculations from the previous section with constraints on homotopy types. First we define $\mathbf{U}_n = \{i/n | i = 0, 1, \ldots n-1\} \subset \mathbb{S}^1$ to be the set of n regularly spaced points. Let $\mathcal{N}(n,k)$ be nerve complex of closed intervals $[i/n, (i+k)/n]$. The homotopy types of $\mathcal{N}(n,k)$ were fully calculated in [4], and this is crucially important for this paper:

▶ **Proposition 3.1** (Theorem 3.5 and 5.4, [4]). *Any nerve complex of n circular arcs is homotopy equivalent to $\mathcal{N}(n', k)$ for some $n' \leq n$. Furthermore, the homotopy types of $\mathcal{N}(n,k)$ are given by the following:*

$$\mathcal{N}(n,k) \simeq \begin{cases} \bigvee^{n-k-1}\mathbb{S}^{2l} & \text{if } \frac{k}{n} = \frac{l}{l+1} \\ \mathbb{S}^{2l+1} & \text{if } \frac{k}{n} \in \left(\frac{l}{l+1}, \frac{l+1}{l+2}\right) \end{cases}$$

In particular, if $(k,n) = (jl, j(l+1))$, then $n-k-1 = j-1$ and thus $\bigvee^{n-k-1}\mathbb{S}^{2l} = \bigvee^{j-1}\mathbb{S}^{2l}$.

This easily implies the following, whose proof is postponed to Section 5. The important thing to note below is that a is bounded regardless of n if $b < k-1$, but is allowed to be unbounded if $b = k-1$.

▶ **Proposition 3.2.** *Let $r \in (0, 1/2)$ and $n > 0$ be given. Let $k = \lfloor (1-2r)^{-1} \rfloor$. Then the following inclusion holds for subsets of $\mathbb{Z}^2$:*

$$\left\{(a,b) \,\middle|\, \check{C}(\mathbf{Y}, r) \simeq \vee^a \mathbb{S}^{2b},\ \mathbf{Y} \subset \mathbb{S}^1,\ \#\mathbf{Y} = n\right\}$$
$$\subseteq \left\{(a,b) \,\middle|\, b+1 \leq k-1,\ a+1 \leq \frac{k}{k-b-1}\right\} \cup \left\{(a, k-1) \,\middle|\, a+1 \leq \frac{n}{k}\right\}$$

where we take $k/0 = \infty$ by convention.

Now we proceed to apply the above constraints to the probabilistic setting. For a topological space K, we define the following notation for probability:

$$p(K, n, r) = \mathbb{P}[\check{C}(\mathbf{X}_n, r) \simeq K]$$

We have the following:

$$\bar{\chi}(n,r) = \mathbb{E}[\chi(\check{C}(\mathbf{X}_n, r))] = \sum_K \chi(K) \cdot p(K, n, r)$$

where the sum is well-defined because there are only finitely many combinatorial structures that $\check{C}(\mathbf{X}_n, r)$ can take.

The following Proposition directly implies Theorem A3.

▶ **Proposition 3.3.** *Let $r \in (0, 1/2)$ and $n > 0$ be given. Let $k = \lfloor (1-2r)^{-1} \rfloor$. Then the following holds:*

$$A_k \leq \bar{\chi}(n,r) \leq k + A_k, \text{ where } A_k = \sum_{1 < a+1 \leq n/k} (a+1) \cdot p(\vee^a \mathbb{S}^{2k-2}, n, r)$$

Proof. By Proposition 3.2, we have that:

$$\left\{\vee^a \mathbb{S}^{2b} \,\middle|\, p(\vee^a \mathbb{S}^{2b}, n, r) > 0\right\} \subseteq \left\{\vee^a \mathbb{S}^{2b} \,\middle|\, b+1 \leq k-1, a+1 \leq \frac{k}{k-b-1}\right\} \cup \left\{\vee^a \mathbb{S}^{2k-2} \,\middle|\, a+1 \leq \frac{n}{k}\right\}$$

Therefore we can break down the expression $\bar{\chi}(n,r)$ into two parts[2]:

$$\bar{\chi}(n,r) = A_{<k} + A_k, \text{ where } A_{<k} = \sum_{\substack{0 \leq b \leq k-2 \\ (a+1)(k-b-1) \leq k}} (a+1) \cdot p(\vee^a \mathbb{S}^{2b}, n, r)$$

Here we used $\chi(\vee^a \mathbb{S}^{2b}) = a+1$. Since sum of probabilities is 1, applying the constraint $(a+1)(k-b-1) \leq k$ implies that $A_{<k} \leq k$. ◀

[2] In $A_{<k}$, we only consider $a \leq 0$ when $b = 0$ and instead consider $a > 0$ when $b > 0$. This is so that the singleton set $\vee^a \mathbb{S}^{2b} = *$ is counted only once.

To prove Theorem C, we control probabilities that $\vee^a\mathbb{S}^{2k-2}$ appear.

▶ **Proposition 3.4.** *Let $n \in \mathbb{Z}^+$, $\delta \in (0,1)$, $r \in (0,1/2)$ be given, and let $k = \lfloor (1-2r)^{-1} \rfloor$. The following holds:*

$$\frac{kA_k - \delta n}{(1-\delta)n + k} \leq B_{k,\delta} \leq \frac{kA_k}{\delta n}$$

where

$$B_{k,\delta} = \sum_{\tilde{\delta} \leq a \leq l} p_a, \text{ and } l = \lfloor n/k \rfloor - 1, \quad \tilde{\delta} = \lceil \delta n/k \rceil - 1$$

Proof. We split A_k into two parts:

$$A_k = \left(2p_1 + 3p_2 + \cdots + \tilde{\delta} p_{\tilde{\delta}-1}\right) + \left((\tilde{\delta}+1)p_{\tilde{\delta}} + \cdots + (l+1)p_l\right)$$

where we abbreviated $p_a = p(\vee^a\mathbb{S}^{2k-2}, n, r)$. From the above it directly follows that:

$$(\tilde{\delta}+1)B_{k,\delta} \leq A_k \leq \tilde{\delta}(1 - B_{k,\delta}) + (l+1)B_{k,\delta}$$

and therefore

$$\begin{aligned}
&\Longrightarrow (\tilde{\delta}+1)B_{k,\delta} \leq A_k \leq \tilde{\delta} + (l+1-\tilde{\delta})B_{k,\delta} \\
&\Longrightarrow \frac{A_k - \tilde{\delta}}{l+1-\tilde{\delta}} \leq B_{k,\delta} \leq \frac{A_k}{\tilde{\delta}+1} \\
&\Longrightarrow \frac{A_k - \lceil \delta n/k \rceil + 1}{\lfloor n/k \rfloor - \lceil \delta n/k \rceil + 1} \leq B_{k,\delta} \leq \frac{A_k}{\lceil \delta n/k \rceil} \\
&\Longrightarrow \frac{A_k - \delta n/k}{(1-\delta)(n/k) + 1} \leq B_{k,\delta} \leq \frac{A_k}{\delta n/k}
\end{aligned}$$ ◀

Now Propositions 2.5, 3.3, 3.4 imply the following, which is a more general version of Theorem C. Indeed Theorem C is implied by setting $\epsilon = \delta = (1-\alpha)k\omega_k/2$ below, and also replacing the gap $\alpha^+ - \alpha^-$ by a smaller but simpler quantity.

▶ **Theorem 3.5.** *Let $r \in [\frac{1}{4}, \frac{1}{2})$ and let $k = \lfloor (1-2r)^{-1} \rfloor$. Given $\epsilon, \delta \in (0,1)$, the following implication holds for large enough n:*

$$1 - 2r \in [\alpha^-, \alpha^+] \implies B_{k,\delta} \in [\beta^- - \epsilon, \beta^+ + \epsilon]$$

where

$$\begin{aligned}
\alpha^\pm &= \frac{1}{k}\frac{n-k}{n-1}\left(1 \pm \frac{\sqrt{k-1}}{n} \cdot \frac{\delta(1-\delta)}{5} \cdot \epsilon\right), \\
\beta^- &= \frac{k\omega_k - \delta}{1-\delta}, \quad \beta^+ = \frac{k\omega_k}{\delta} \\
\omega_k &= \frac{(k-1)^{k-1}}{k!e^{k-1}}
\end{aligned}$$

The bounds $\beta^\pm$ satisfy $\beta^- \leq k\omega_k \leq \beta^+$. Also $\beta^- > 0$ iff $\delta < k\omega_k$ and $\beta^+ < 1$ iff $\delta > k\omega_k$.

Proof. We first describe the heuristic reasoning for the bounds, which is rather simple. Proposition 3.4 gives us:

$$\frac{kA_k - \delta n}{(1-\delta)n + k} \leq B \leq \frac{kA_k}{\delta n}$$

By Proposition 2.5 and 3.3, the upper bound has the following approximations:

$$\frac{kA_k}{\delta n} \approx \frac{k\bar{\chi}}{\delta n} \approx \frac{k\omega_k}{\delta}$$

and similarly the lower bound has the following approximations:

$$\frac{kA_k - \delta n}{(1-\delta)n + k} \approx \frac{kA_k - \delta n}{(1-\delta)n} \approx \frac{k\bar{\chi} - \delta}{1-\delta} \approx \frac{k\omega_k - \delta}{1-\delta}$$

The actual proof becomes more complicated due to using a different choice of ϵ in applying Proposition 2.5.

Let $\epsilon' = \delta(1-\delta) \cdot \epsilon/5$. We apply Proposition 2.5 with ϵ' taking the role of ϵ, and this gives the choice of $\alpha^{\pm}$ in the theorem. Therefore $r \in [\alpha^-, \alpha^+]$ implies the following:

$$(1-\epsilon')\omega_k \le \frac{\bar{\chi}}{n} \le (1+\epsilon')\omega_k \tag{3.1}$$

Before going further, we note the following inequalities for ϵ', which we will use later:

$$\begin{aligned}
\epsilon' = \frac{\delta(1-\delta)\epsilon}{4+1} &\le \frac{\delta(1-\delta)\epsilon}{4+\delta(1-\delta)\epsilon} \\
\implies \frac{\epsilon'}{1-\epsilon'} &\le \frac{\delta(1-\delta)\epsilon}{4} \\
\implies \frac{\epsilon'}{1-\epsilon'} &\le \min\left(4\delta, \delta^{-1} - 1, 1\right) \cdot \frac{\epsilon}{4}
\end{aligned} \tag{3.2}$$

Upper bound. By Equation (3.1) and Proposition 3.3, we have:

$$\frac{k\omega_k}{\delta} \ge \frac{1}{1+\epsilon'}\frac{k\bar{\chi}}{\delta n} \ge \frac{1}{1+\epsilon'}\frac{kA_k}{\delta n}$$

By Equation (3.2), we have that:

$$\frac{1}{1+\epsilon'}\frac{kA_k}{\delta n} \ge \frac{kA_k}{\delta n} - \epsilon$$

Then Proposition 3.4 applies and we have the upper bound.

Lower bound. By Equation (3.1) and Proposition 3.3, we have:

$$\frac{k\omega_k - \delta}{1-\delta} \le \frac{1}{1-\delta}\left(\frac{1}{1-\epsilon'}\frac{k\bar{\chi}}{n} - \delta\right) \le \frac{1}{1-\delta}\left(\frac{1}{1-\epsilon'}\frac{k^2 + kA_k}{n} - \delta\right)$$

Let L_0 be the right hand side. We rewrite it as follows:

$$L_0 = L_1 + E_1 = L_2 + E_1 + E_2$$

where

$$\begin{aligned}
L_1 &= \frac{kA_k - \delta n}{(1-\delta)(1-\epsilon')n}, \; E_1 = \frac{\delta\epsilon' + k^2/n}{(1-\delta)(1-\epsilon')} \\
L_2 &= \frac{kA_k - \delta n}{(1-\delta)n + k}, \; E_2 = \frac{kA_k - \delta n}{(1-\delta)(1-\epsilon')n} \cdot \frac{k + (1-\delta)n\epsilon'}{(1-\delta)n + k}
\end{aligned}$$

By Equation (3.2), the relation $kA_k \le n$ and by taking n large enough, we see that

$$E_1, E_2 \le \epsilon/2$$

This implies that:

$$\frac{k\omega_k - \delta}{1-\delta} - \epsilon \le L_0 - \epsilon = L_2 + E_1 + E_2 - \epsilon \le L_2$$

Then again Proposition 3.4 applies and we have the lower bound. ◀

4 Odd spheres

We prove Theorem B by a simple argument using the stability of persistence diagram. In this case, we will be using the Čech complex constructed from the full set of the circle, and then bound the Gromov-Hausdorff distance between the full circle and a finite sample of it. We use the following result from [1]:

▶ **Theorem 4.1.** *The homotopy types of the Rips and Čech complexes on the circle of unit circumference are as follows:*

$$\mathrm{VR}(\mathbb{S}^1, r) \simeq \begin{cases} \mathbb{S}^{2l+1} & \text{if } \frac{l}{2l+1} < r < \frac{l+1}{2l+3} \\ \bigvee^{\mathfrak{c}} \mathbb{S}^{2l} & \text{if } r = \frac{l}{2l+1} \end{cases}$$
$$\check{C}(\mathbb{S}^1, r) \simeq \begin{cases} \mathbb{S}^{2l+1} & \text{if } \frac{l}{2l+2} < r < \frac{l+1}{2l+4} \\ \bigvee^{\mathfrak{c}} \mathbb{S}^{2l} & \text{if } r = \frac{l}{2l+2} \end{cases}$$

where $\mathfrak{c}$ is the cardinality of the continuum.

We also note the stability of persistence:

▶ **Theorem 4.2** (Stability of Persistence)**.** *If X, Y are metric spaces and $\mathcal{D}_k M$ is the k-dimensional persistence diagram of persistence module M, then*

$$\mathrm{d}_B(\mathcal{D}_k \mathbf{VR}(X), \mathcal{D}_k \mathbf{VR}(Y)) \leq \mathrm{d}_{GH}(X, Y)$$
$$\mathrm{d}_B(\mathcal{D}_k \check{C}(X), \mathcal{D}_k \check{C}(Y)) \leq \mathrm{d}_{GH}(X, Y)$$

where d_{GH} denotes the Gromov-Hausdorff distance.

We conclude the main portion of this paper with the following, which is a more detailed version of Theorem B.

▶ **Proposition 4.3.** *For each $l \geq 0$ and $t \in (\frac{l}{2l+2}, \frac{l+1}{2l+4})$, the following holds with probability at least $Q_n(r'/2)$:*

$$\check{C}(\mathbf{X}_n, t) \simeq \mathbb{S}^{2l+1}$$

where r' is:

$$r' = \frac{1}{4(l+1)(l+2)} - \left| t - \frac{2l^2 + 4l + 1}{4(l+1)(l+2)} \right|$$

Proof. Consider a random sample $\mathbf{X}_n = (X_1, \ldots X_n)$. Then with probability $Q_n(r/2)$, arcs of radius r centered at $\mathbf{X}_n$ covers $\mathbb{S}^1$, so that $\mathrm{d}_{GH}(\mathbf{X}_n, \mathbb{S}^1) \leq \mathrm{d}_H(\mathbf{X}_n, \mathbb{S}^1) \leq r$. This implies:

$$\mathrm{d}_B(\mathcal{D}_k \check{\mathrm{C}}(\mathbf{X}_n), \mathcal{D}_k \check{\mathrm{C}}(\mathbb{S}^1)) \leq \mathrm{d}_{GH}(\mathbf{X}_n, \mathbb{S}^1) \leq r$$

For each $l \geq 0$, we have that:

$$\mathcal{D}_{2l+1} \check{\mathrm{C}}(\mathbb{S}^1) = \left\{ \left(\frac{l}{2l+2}, \frac{l+1}{2l+4} \right) \right\}$$

so that the definition of the bottleneck distance implies that

$$\exists (u, v) \in \mathcal{D}_{2l+1} \check{\mathrm{C}}(\mathbf{X}_n)$$
$$\text{with } \frac{l}{2l+2} - r \leq u \leq \frac{l}{2l+2} + r$$
$$\frac{l+1}{2l+4} - r \leq v \leq \frac{l+1}{2l+4} + r$$

This implies that whenever $\frac{l}{2l+2} + r \leq t \leq \frac{l+1}{2l+4} - r$, we have:

$$1 \leq \dim H_{2l+1}\check{\mathrm{C}}(\mathbf{X}_n, t)$$

and due to the enumeration of possible homotopy types, we have that:

$$\check{\mathrm{C}}(\mathbf{X}_n, t) \simeq \mathbb{S}^{2l+1}$$

The condition translates to $\left|t - \frac{1}{2}\left(\frac{l}{2l+2} + \frac{l+1}{2l+4}\right)\right| < \frac{1}{2}\left(\frac{l+1}{2l+4} - \frac{l}{2l+2}\right) - r$, or equivalently

$$\left|t - \frac{2l^2+4l+1}{4(l+1)(l+2)}\right| < \frac{1}{4(l+1)(l+2)} - r$$

and thus we obtain the proof. ◀

5 Technical tools

In this section we produce proofs of technical results that were postponed from the main portion of the paper.

▶ **Theorem 5.1** (Stevens). *If k arcs of fixed length a are independently, identically and uniformly sampled from the circle of circumference 1, then the probability that these arcs cover the circle is equal to the following:*

$$Q_k(a) = \sum_{l=0}^{\lfloor 1/a \rfloor} (-1)^l \binom{k}{l} (1 - la)^{k-1}$$

Proof. The proof is an application of inclusion-exclusion principle. Consider the set $E = \{(x_1, \dots x_k) | 0 \leq x_1 < \dots < x_k < 1\}$. For each collection of indices $J \subseteq \{1, \dots k\}$, define $\bar{E}_J$ and E_J as the following subsets of E:

$$\begin{aligned} E_J =& \{(x_1, \dots x_k) \in E | j \in J \iff x_{j+1} - x_j > a\} \\ \bar{E}_J =& \{(x_1, \dots x_k) \in E | j \in J \implies x_{j+1} - x_j > a\} = \bigsqcup_{J' \supseteq J} E_{J'} \end{aligned}$$

By definition, we have $\mathrm{Vol}(E_\emptyset) = Q_k(a)$. To compute it, we apply the inclusion-exclusion principle for the membership of each E_J over $\bar{E}_{J'}$ whenever $J' \supseteq J$. Noting the relation $\sum_{l=1}^{k} (-1)^{l+1} \binom{k}{l} = 1$, we see that:

$$1 = \sum_{J \subseteq \{1, \dots k\}} \mathrm{Vol}(E_J) = \mathrm{Vol}(E_\emptyset) - \sum_{\emptyset \neq J \subseteq \{1, \dots k\}} (-1)^{\#J} \mathrm{Vol}(\bar{E}_J)$$

Finally, if $l = \#J$ and $l \leq \lfloor 1/a \rfloor$, then $\mathrm{Vol}(\bar{E}_J) = (1 - la)^{n-1}$. This is because demanding gap conditions $x_{i+1} - x_i > a$ at l places is equivalent to sampling $n - 1$ points from an interval of length $1 - la$[3]. Meanwhile if $l > \lfloor 1/a \rfloor$, then we always have $\mathrm{Vol}(\bar{E}_J) = 0$. Plugging these into the above equation, we get:

$$\mathrm{Vol}(E_\emptyset) = 1 + \sum_{l=1}^{\lfloor 1/a \rfloor} (-1)^l \binom{k}{l} (1 - la)^{n-1}$$

as desired. ◀

[3] This can be seen more precisely by considering the collection E' of $(y_1, \dots y_{k-1})$ defined by $y_i = x_{i+1} - x_i > 0$ and $\sum y_i \leq 1$, and then considering the subset E'_J defined by $y_i > a$ for $i \in J$. The quantity of interest is $\mathrm{Vol}(E'_J)/\mathrm{Vol}(E')$. Furthermore, the map $(y_1, \dots y_{k-1}) \mapsto (y_1 - \mathbf{1}_{1 \in J}, \dots y_{k-1} - \mathbf{1}_{k-1 \in J})$ isometrically maps E'_J to $(1 - la) \cdot E'$, so that $\mathrm{Vol}(E'_J) = (1 - la)^{k-1} \mathrm{Vol}(E')$ due to the $(k-1)$-dimensional volume scaling. This is exactly the original claim.

The following two lemmas are essentially exercises in calculus, and they are needed to prove Proposition 2.4.

▶ **Lemma 5.2.** *For $a, b \geq 1$, the function $f(t) = t^a(1-t)^b$ satisfies the following:*

(a) *In the range $0 \leq t \leq 1$, $f(t)$ achieves the unique maximum value at $t = a/(a+b)$:*

$$\max_{0 \leq t \leq 1} f(t) = f\left(\frac{a}{a+b}\right) = \frac{a^a b^b}{(a+b)^{a+b}}$$

Also, $f(t)$ is increasing on $t \in (0, a/(a+b))$ and decreasing on $t \in (a/(a+b), 1)$.

(b) *The following linear lower bounds hold:*

$$f(t) \geq u\Big((a+b)vt - av + 1\Big), \text{ when } 0 < t < \frac{a}{a+b}$$
$$f(t) \geq u\Big(-(a+b)vt + av + 1\Big), \text{ when } \frac{a}{a+b} < t < 1$$

where

$$u = \frac{a^a b^b}{(a+b)^{a+b}}, \quad v = \sqrt{\frac{a+b}{ab}}$$

(c) *For each $\lambda \in [0,1]$, we have that:*

$$\left|t - \frac{a}{a+b}\right| < \frac{(1-\lambda)\sqrt{ab}}{(a+b)^{3/2}} \implies t^a(1-t)^b > \lambda u$$

Proof. The first two derivatives are:

$$f'(t) = \Big(a - (a+b)t\Big)t^{a-1}(1-t)^{b-1}$$
$$f''(t) = \Big((a+b)(a+b-1)t^2 + 2a(1-a-b)t + a(a-1)\Big)t^{a-2}(1-t)^{b-2}$$

The first derivative vanishes at $t \in \{a/(a+b), 0, 1\}$ and the second derivative vanishes at $t \in \{t_0 \pm \eta_0, 0, 1\}$ where

$$t_0 = \frac{a}{a+b}, \quad \eta_0 = \frac{1}{a+b}\sqrt{\frac{ab}{a+b-1}} > \frac{\sqrt{ab}}{(a+b)^{3/2}} = \eta_1$$

The first derivative is positive at $(0, a/(a+b))$ and negative at $(a/(a+b), 1)$. Thus the maximum at $t \in [0,1]$ is given by:

$$f(t_0) = \frac{a^a b^b}{(a+b)^{a+b}}$$

It is then easy to see the following:

$$f(t) \geq \frac{f(t_0)}{\eta_1}(t - t_0) + f(t_0), \text{ when } 0 < t < t_0$$
$$f(t) \geq \frac{-f(t_0)}{\eta_1}(t - t_0) + f(t_0), \text{ when } t_0 < t < 1$$

and

$$\pm\frac{f(t_0)}{\eta_1}(t - t_0) + f(t_0) = \frac{a^a b^b}{(a+b)^{a+b}}\left(\pm\frac{(a+b)^{3/2}}{\sqrt{ab}}\left(t - \frac{a}{a+b}\right) + 1\right)$$

(c) follows from the linear bound of (b). ◀

▶ **Lemma 5.3.** *Let $m, n \geq 1$ be integers and define:*

$$f_{m,n}(t) = \binom{n}{m}(mt)^{m-1}(1-mt)^{n-m}$$

Then $f_{m,n}$ satisfies the following:

(a) *$f_{m,n}(t)$ is increasing when $0 < t < t_0$ and decreasing when $t_0 < t < 1/m$ where $t_0 = \frac{1}{n-1}(1-\frac{1}{m})$.*

(b) *The maximum over $0 < t < 1/m$ is given by:*

$$\max_{0<mt<1} f_{m,n}(t) = f_{m,n}(t_0) = \binom{n}{m}\frac{(m-1)^{m-1}(n-m)^{n-m}}{(n-1)^{n-1}}$$

(c) *For each $\lambda \in [0,1]$, we have that:*

$$|t-t_0| < \frac{(1-\lambda)\sqrt{(m-1)(n-m)}}{m(n-1)^{3/2}} \implies f_{m,n}(t) > \lambda f_{m,n}(t_0)$$

(d) *The normalised limit of maximum as $n \to \infty$ is given by:*

$$\lim_{n\to\infty} \frac{\max_{0<t<1/m} f_{m,n}(t)}{n} = \frac{(m-1)^{m-1}}{m!e^{m-1}}$$

Proof. (a)-(c) follow from the previous lemma. For (d), we compute:

$$\begin{aligned}\lim_{n\to\infty} \frac{\max_{0<t<1/m} f_{m,n}(t)}{n} =& \frac{(m-1)^{m-1}}{m!} \lim_{n\to\infty} (n-1)(n-2)\cdots(n-m+1)\frac{(n-m)^{n-m}}{(n-1)^{n-1}} \\ =& \frac{(m-1)^{m-1}}{m!} \lim_{n\to\infty} \frac{(n-m)^{n-m}}{(n-1)^{n-m}}\end{aligned}$$

and also

$$\lim_{n\to\infty} \frac{(n-m)^{n-m}}{(n-1)^{n-m}} = \lim_{n\to\infty}\left(1-\frac{m-1}{n-1}\right)^{n-m} = \lim_{n\to\infty}\left(1-\frac{m-1}{n-1}\right)^{n-1} = \frac{1}{e^{m-1}}$$

which gives the desired expression. ◀

We now prove Proposition 2.4.

▶ **Proposition 5.4.** *Suppose that m, n are integers with $2 \leq m < \sqrt{n}$. The following holds for $\bar{\chi}(n,r)$.*

(a) *The following bounds hold:*

$$a_{m,n} \leq \frac{\bar{\chi}(n, s_{m,n})}{n} \leq M \leq a_{m,n} + b_{m,n}$$

where

$$\begin{aligned} M =& \max\left\{\frac{1}{n}\bar{\chi}\left(n, \frac{1-r}{2}\right) \,\middle|\, r \in \left(\frac{1}{m+1}, \frac{1}{m}\right)\right\} \\ s_{m,n} =& \frac{(m-1)n}{2(n-1)m} \\ a_{m,n} =& \binom{n}{m}\frac{(m-1)^{m-1}(n-m)^{n-m}}{n(n-1)^{n-1}} \\ b_{m,n} =& en^{m-1}\left(1-\frac{1}{m+1}\right)^{n-1}\end{aligned}$$

(b) *We have the following limits:*

$$\lim_{n\to\infty} a_{m,n} = \frac{(m-1)^{m-1}}{m!e^{m-1}}, \quad \lim_{n\to\infty} b_{m,n} = 0$$

(c) *Suppose additionally that $n > 2m^2$. Then for each $\lambda \in [0,1]$, we have that:*

$$\left| r - \frac{n-m}{(n-1)m} \right| < \frac{(1-\lambda)\sqrt{(m-1)(n-m)}}{m(n-1)^{3/2}} \implies \frac{1}{n}\bar{\chi}\left(n, \frac{1-r}{2}\right) > \lambda a_{m,n}$$

This condition for r in particular satisfies $r \in \left(\frac{1}{m+1}, \frac{1}{m}\right]$.

Proof. Let $r \in \left(\frac{1}{m+1}, \frac{1}{m}\right]$ and also write $r = \frac{1}{m} - t$, with $t \in \left[0, \frac{1}{m(m+1)}\right]$. Then we may rewrite the normalised expected Euler characteristic as follows:

$$\begin{aligned}
\bar{\chi}\left(n, \frac{1-r}{2}\right) &= \sum_{k=1}^{m} \binom{n}{k} (1-kr)^{k-1}(kr)^{n-k} \\
&= \sum_{k=1}^{m} \binom{n}{k} \left(1 - \frac{k}{m} + kt\right)^{k-1} \left(\frac{k}{m} - kt\right)^{n-k}
\end{aligned}$$

We now claim that the $k = m$ term is the dominant one among the above summands. As such, we split the above sum as:

$$\bar{\chi}\left(n, \frac{1-r}{2}\right) = f_{m,n}(t) + E$$

where

$$\begin{aligned}
f_{m,n}(t) &= \binom{n}{m} (mt)^{m-1}(1-mt)^{n-m}, \\
E &= \sum_{k=1}^{m-1} \binom{n}{k} \left(1 - \frac{k}{m} + kt\right)^{k-1} \left(\frac{k}{m} - kt\right)^{n-k}
\end{aligned}$$

Since $m < \sqrt{n}$, we have $s_{m,n} = \frac{1}{n-1}(1 - \frac{1}{m}) < \frac{1}{m(m+1)}$. Therefore, the previous Lemma tells us that $f_{m,n}(t)$ achieves (global) maximum at $\tilde{s} \in \left(0, \frac{1}{m(m+1)}\right]$, with the maximum value given by:

$$f_{m,n}(\tilde{s}) = n \cdot a_{m,n}, \text{ where } a_{m,n} = \binom{n}{m} \frac{(m-1)^{m-1}(n-m)^{n-m}}{n(n-1)^{n-1}}$$

We also bound E as follows, using the inequality $\frac{m}{m+1} < 1 - mt \le 1$:

$$\begin{aligned}
E &= \sum_{k=1}^{m-1} \binom{n}{k} \left(1 - \frac{k}{m}(1-mt)\right)^{k-1} \left(\frac{k}{m}(1-mt)\right)^{n-k} \\
&\le \sum_{k=1}^{m-1} \binom{n}{k} \left(1 - \frac{1}{m+1}\right)^{k-1} \left(1 - \frac{1}{m}\right)^{n-k} \\
&\le \sum_{k=1}^{m-1} \frac{n^k}{k!} \left(1 - \frac{1}{m+1}\right)^{n-1} \\
&\le e n^{m-1} \left(1 - \frac{1}{m+1}\right)^{n-1}
\end{aligned}$$

This shows (a). Now (b) follows from the previous Lemma and the fact that $(1 - \frac{1}{m+1})^n$ term causes exponential decay for $b_{m,n}$.

(c) follows from (c) of the previous Lemma. We additionally impose the condition $n > 2m^2$, so that the endpoints of t satisfying the condition fall in the interval $t \in \left[0, \frac{1}{m(m+1)}\right)$. ◀

We prove Proposition 3.2. Let $\mathbf{U}_n = \{i/n \mid i = 0, 1, \dots n-1\} \subset \mathbb{S}^1$ be the set of n equally spaced points. We note briefly that the following follows from definition:

$$\check{\mathrm{C}}(\mathbf{U}_n, r) = \check{\mathrm{C}}\left(\mathbf{U}_n, \frac{\lfloor 2rn \rfloor}{2n}\right) = \mathcal{N}(n, \lfloor 2rn \rfloor)$$

▶ **Proposition 5.5.** *Let $r \in (0, 1/2)$ and n be given; define $\tilde{r} = 1 - 2r$ and let $k = \lfloor \tilde{r}^{-1} \rfloor$. Then we have the following.*

(1) *The following equality holds for subsets of $\mathbb{Z}^3$:*

$$\left\{(n, a, b) \,\middle|\, \check{C}(\mathbf{U}_n, r) \simeq \vee^a \mathbb{S}^{2b}\right\} = \left\{((a+1)(b+1), a, b) \,\middle|\, b + 1 \le \tilde{r}^{-1},\ a + 1 \le \frac{1}{1 - (b+1)\tilde{r}}\right\}$$

(2) *The following inclusion holds for subsets of $\mathbb{Z}^2$:*

$$\begin{aligned}&\left\{(a, b) \,\middle|\, \check{C}(\mathbf{Y}, r) \simeq \vee^a \mathbb{S}^{2b},\ \mathbf{Y} \subset \mathbb{S}^1,\ \#\mathbf{Y} = n\right\}\\ \subseteq &\left\{(a, b) \,\middle|\, b + 1 \le k - 1,\ a + 1 \le \frac{k}{k - b - 1}\right\} \cup \left\{(a, k-1) \,\middle|\, a + 1 \le \frac{n}{k}\right\}\end{aligned}$$

where in the final expression, $k/0 = \infty$ by convention.

Proof.

(1) To have $\mathcal{N}(n, \lfloor 2rn \rfloor) = \check{\mathrm{C}}(\mathbf{U}_n, r) \simeq \vee^a \mathbb{S}^{2b}$, we see from Proposition 3.1 that the condition is given by $(\lfloor 2rn \rfloor, n) = ((a+1)b, (a+1)(b+1))$. This determines n from (a, b). The condition on $\lfloor 2rn \rfloor$ is then:

$$\begin{aligned}&(a+1)b \le 2r(a+1)(b+1) < (a+1)b + 1\\ \iff &\tilde{r}(b+1) \le 1,\ a < \tilde{r}(a+1)(b+1)\\ \iff &(b+1) \le \tilde{r}^{-1},\ (a+1) < (1 - \tilde{r}(b+1))^{-1}\end{aligned}$$

as desired.

(2) We claim the following:

$$\begin{aligned}&\left\{(a, b) \,\middle|\, \check{\mathrm{C}}(\mathbf{Y}, r) \simeq \vee^a \mathbb{S}^{2b},\ \mathbf{Y} \subset \mathbb{S}^1,\ \#\mathbf{Y} = n\right\}\\ = &\left\{(a, b) \,\middle|\, \check{\mathrm{C}}(\mathbf{U}_m, r) \simeq \vee^a \mathbb{S}^{2b},\ m \le n\right\}\\ \subseteq &\left\{(a, b) \,\middle|\, b + 1 \le k,\ a + 1 \le \min\left(\frac{n}{b+1}, \frac{1}{1 - (b+1)\tilde{r}}\right)\right\}\\ \subseteq &\left\{(a, b) \,\middle|\, b + 1 \le k - 1,\ a + 1 \le \frac{k}{k - b - 1}\right\} \cup \left\{(a, k-1) \,\middle|\, a + 1 \le \frac{n}{k}\right\}\end{aligned}$$

The first equality holds due to Proposition 3.1. The first inclusion follows from (1). The second inclusion follows from separating the two cases $b + 1 < k$ and $b + 1 = k$. ◀

References

1 Michał Adamaszek and Henry Adams. The vietoris–rips complexes of a circle. *Pacific Journal of Mathematics*, 290(1):1–40, 2017.

2 Michał Adamaszek and Henry Adams. On vietoris–rips complexes of hypercube graphs. *Journal of Applied and Computational Topology*, 6(2):177–192, 2022.

3 Michał Adamaszek, Henry Adams, and Florian Frick. Metric reconstruction via optimal transport. *SIAM Journal on Applied Algebra and Geometry*, 2(4):597–619, 2018.

4 Michał Adamaszek, Henry Adams, Florian Frick, Chris Peterson, and Corrine Previte-Johnson. Nerve complexes of circular arcs. *Discrete & Computational Geometry*, 56:251–273, 2016.

5 Michał Adamaszek, Henry Adams, and Francis Motta. Random cyclic dynamical systems. *Advances in Applied Mathematics*, 83:1–23, 2017.

6 Michał Adamaszek, Henry Adams, and Samadwara Reddy. On vietoris–rips complexes of ellipses. *Journal of Topology and Analysis*, 11(03):661–690, 2019.

7 Henry Adams, Samir Chowdhury, Adam Quinn Jaffe, and Bonginkosi Sibanda. Vietoris-rips complexes of regular polygons. *arXiv preprint*, 2018. `arXiv:1807.10971`.

8 Omer Bobrowski and Matthew Kahle. Topology of random geometric complexes: a survey. *Journal of applied and Computational Topology*, 1:331–364, 2018.

9 Omer Bobrowski and Goncalo Oliveira. Random čech complexes on riemannian manifolds. *Random Structures & Algorithms*, 54(3):373–412, 2019.

10 Omer Bobrowski and Shmuel Weinberger. On the vanishing of homology in random čech complexes. *Random Structures & Algorithms*, 51(1):14–51, 2017.

11 Peter Bubenik and Peter T Kim. A statistical approach to persistent homology. *Homology, homotopy and Applications*, 9(2):337–362, 2007.

12 Henry-Louis de Kergorlay, Ulrike Tillmann, and Oliver Vipond. Random čech complexes on manifolds with boundary. *Random Structures & Algorithms*, 61(2):309–352, 2022.

13 Jean-Claude Hausmann et al. On the vietoris-rips complexes and a cohomology theory for metric spaces. *Annals of Mathematics Studies*, 138:175–188, 1995.

14 Matthew Kahle. Random geometric complexes. *Discrete & Computational Geometry*, 45:553–573, 2011.

15 Sunhyuk Lim, Facundo Memoli, and Osman Berat Okutan. Vietoris-rips persistent homology, injective metric spaces, and the filling radius. *arXiv preprint*, 2020. `arXiv:2001.07588`.

16 Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39:419–441, 2008.

17 Taejin Paik and Otto van Koert. Expected invariants of simplicial complexes obtained from random point samples. *Archiv der Mathematik*, 120(4):417–429, 2023.

18 Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.

19 WL Stevens. Solution to a geometrical problem in probability. *Annals of Eugenics*, 9(4):315–320, 1939.

20 Andrew M Thomas and Takashi Owada. Functional limit theorems for the euler characteristic process in the critical regime. *Advances in Applied Probability*, 53(1):57–80, 2021.