Maximum Persistent Betti Numbers of Čech Complexes

Herbert Edelsbrunner
ISTA (Institute of Science and Technology Austria), Klosterneuburg, Austria

Matthew Kahle
Department of Mathematics, Ohio State University, Columbus, Ohio

Shu Kanazawa
Institute for Advanced Study, Kyoto University, Kyoto, Japan

Abstract

This note proves that only a linear number of holes in a Čech complex of n points in $\mathbb{R}^d$ can persist over an interval of constant length. The proof uses a packing argument supported by relating the Čech complexes with corresponding snap complexes over the cells in a partition of space. The bound also applies to Alpha complexes and Vietoris–Rips complexes.

2012 ACM Subject Classification Theory of computation → Computational geometry

Keywords and phrases Čech complexes, Betti numbers, persistent homology.

Funding The three authors are supported by the Wittgenstein Prize, Austrian Science Fund (FWF), grant no. Z 342-N31, by the DFG Collaborative Research Center TRR 109, Austrian Science Fund (FWF), grant no. I 02979-N35, the U.S. National Science Foundation (NFS-DMS), grant no. 2005630, and a JSPS Grant-in-Aid for Transformative Research Areas (A) (22H05107, Y.H.), respectively.

1 Introduction

What is the maximum number of holes created by n possibly overlapping closed unit balls in a Euclidean space, and how big can they be? To move toward a more precise formulation of this question, let $p < d$ be two positive integers, and consider the asymptotic behavior of

$$M_{p,d}(n) = \max \left\{ \beta_p \left(\bigcup\nolimits_{x \in A} B(x,1) \right) | A \subseteq \mathbb{R}^d, \mathrm{card}\, A = n \right\} \tag{1}$$

as $n \to \infty$. Here, $B(x,1)$ is the closed ball of radius 1 centered at x, and β_p of the union of such balls is the p-th Betti number of this space. The *Čech complex* of A for radius r, denoted by $\check{C}_r = \check{C}_r(A)$, is the simplicial complex whose vertices are the points in A, and whose simplices are the subsets of points such that the closed balls of radius r centered at these points have a non-empty common intersection. By the Nerve Theorem, the Čech complex for radius $r = 1$ has the homotopy type of the union of the unit balls in (1). Therefore, $M_{p,d}(n)$ is also the maximum p-th Betti number of $\check{C}_1$, for any set of n points in $\mathbb{R}^d$.

Recently, Edelsbrunner and Pach [4] proved that $M_{p,d}(n) = \Theta(n^{\min\{p+1, \lceil d/2 \rceil\}})$. For example, $M_{1,2}(n)$ grows linearly in n, but $M_{1,3}(n)$ and $M_{2,3}(n)$ grow quadratically in n. The upper bound is easily derived from the Upper Bound Theorem for convex polytopes, see e.g. [9], and their main contribution is the actual construction that proves that this upper bound is asymptotically tight. However, most holes in their construction appear to be small: they vanish when the balls are slightly enlarged. In the language of persistent homology, they have short lifetimes. This motivates us to fix $\varepsilon > 0$ and to look at

$$M_{p,d,\varepsilon}(n) = \max\{\beta_p(\check{C}_1, \check{C}_{1+\varepsilon}) \mid A \subseteq \mathbb{R}^d, \mathrm{card}\, A = n\}, \tag{2}$$

where $\beta_p(\check{C}_1, \check{C}_{1+\varepsilon})$ is the p-th persistent Betti number of the inclusion of $\check{C}_1$ in $\check{C}_{1+\varepsilon}$. In other words, this is the number of p-dimensional holes in $\check{C}_1$ that are still holes in $\check{C}_{1+\varepsilon}$.

When we fix the parameter $\varepsilon > 0$, most of the holes in Edelsbrunner and Pach's construction are no longer counted since they do not persist even to radius $1 + \varepsilon$, so it is not surprising that the asymptotic behavior of $M_{p,d,\varepsilon}(n)$ is different from that of $M_{p,d}(n)$, and we show that $M_{p,d,\varepsilon}(n)$ grows only linearly in n, for every fixed p and d. In words, the number of holes that cover a fixed interval of positive length in the Čech filtration is at most some constant times the number of points.

The Čech complex has the same homotopy type as the Alpha complex for the same points and the same radius [3, Section III.4], which is a subcomplex of the Delaunay mosaic of the points [2]. Hence, $M_{p,d}(n)$ is also the maximum p-th Betti number of the Alpha complex of n points in $\mathbb{R}^d$, and $M_{p,d,\varepsilon}(n)$ is also the maximum p-th persistent Betti number of the Alpha complex for unit radius included in that of radius $1 + \varepsilon$. Less obviously, our linear upper bound for $M_{p,d,\varepsilon}(n)$ extends to the inclusion of Vietoris–Rips complexes for unit radius in that of radius $1 + \varepsilon$. While we do not know of a direct connection, every step in our proof of the bound for Čech complexes extends to Vietoris–Rips complexes. The resulting linear upper bound for the persistent Betti numbers should be compared to the bounds of Goff [5] on the (non-persistent) Betti numbers of Vietoris–Rips complexes. For $p = 1$, he shows a linear upper bound in all dimensions, which is stronger than our linear upper bound for the persistent Betti numbers. However, for $p > 1$, his bound is only $o(n^p)$, which is much higher than our linear upper bound for the persistent Betti numbers.

The outline of this paper is as follows. Section 2 proves technical lemmas about how cycles are maintained if we glue vertices in a simplicial complex. Section 3 introduces the snap complex of a Čech complex and relates its Betti numbers to the persistent Betti numbers of the Čech complex. Section 4 combines these preparations to prove the linear upper bound on the persistent Betti number of Čech complexes. Section 5 concludes the paper.

2 Gluing Vertices

Let $A \subseteq \mathbb{R}^d$ be finite. By definition, the Čech complex of A for radius $r \geq 0$ consists of all subsets $B \subseteq A$ such that the closed balls of radius r centered at the points in B have a non-empty common intersection. Writing $r(B)$ for the radius of the smallest enclosing sphere of B, we have $B \in \check{C}_r(A)$ iff $r(B) \leq r$. Letting $B' \subseteq \mathbb{R}^d$ be another finite set, the *Hausdorff distance* between B and B' is

$$H(B, B') = \max \left\{ \max_{b \in B} \min_{b' \in B'} \|b - b'\|, \max_{b' \in B'} \min_{b \in B} \|b' - b\| \right\}. \tag{3}$$

We show that the difference between $r(B)$ and $r(B')$ is at most the Hausdorff distance between the two sets.

▶ **Lemma 2.1.** *$|r(B) - r(B')| \leq H(B, B')$.*

Proof. Let x be the center of the smallest enclosing sphere of B, whose radius is $r(B)$. By definition of Hausdorff distance, B' is contained in the union of balls of radius $\varepsilon = H(B, B')$ centered at the points in B. Hence, the sphere with center x and radius $r(B) + \varepsilon$ encloses B', so $r(B') \leq r(B) + \varepsilon$. Symmetrically, $r(B) \leq r(B') + \varepsilon$, which implies the claim. ◀

We employ Lemma 2.1 to reduce the size of a cycle. To explain how, we use standard terminology for homology with $\mathbb{Z}/2\mathbb{Z}$ coefficients, which we briefly review. A *p-chain*, γ, is a collection of p-simplices, and the *vertices* of γ are the vertices of its p-simplices. To *add* two p-chains means taking their symmetric difference: $\gamma + \gamma' = \gamma \oplus \gamma'$. The *boundary* of a p-simplex are its $(p-1)$-dimensional faces, and the *boundary* of γ is the sum of the boundaries of its p-simplices. A *p-cycle* is a p-chain with empty boundary, denoted by $\partial\gamma = 0$. Two p-cycles are *homologous*, denoted by $\gamma \sim \gamma'$, if there is a $(p+1)$-chain, Γ, that satisfies $\partial\Gamma = \gamma + \gamma'$. In this case, we say Γ is a *filling* of $\gamma + \gamma'$. Finally, γ is *trivial* if $\gamma \sim 0$.

▶ **Definition 2.2.** *Let γ be a p-cycle and $x \neq y$ vertices of γ. To glue x and y, we substitute a new vertex, z, for x and y in all p-simplices, and write $\gamma|_{x\sim y}$ for the resulting p-chain.*

After the substitution, a p-simplex that contains both, x and y, is a $(p-1)$-simplex and thus implicitly removed from the p-cycle by the gluing operation.

To reason about the gluing of x and y, we write $\mathrm{St}_\gamma(x)$ for the p-simplices in γ that share x, and note that $\mathrm{St}_\gamma(x,y) = \mathrm{St}_\gamma(x) \cap \mathrm{St}_\gamma(y)$ are the p-simplices in γ that share the edge connecting x and y. As illustrated in Figure 1, it is possible that after substituting z for x

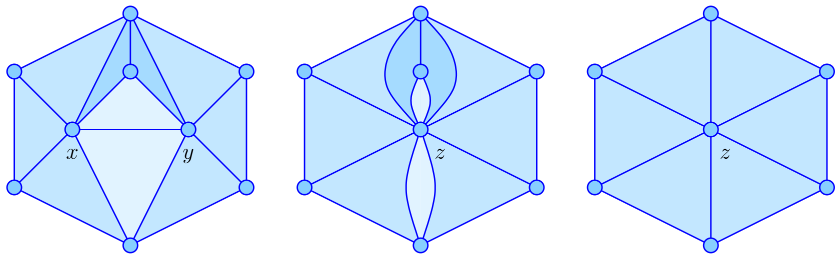

Figure 1: A portion of a 2-cycle on the *left*, in which x and y belong to two triangles that share an edge different from the edge connecting x and y. For better visibility, we shade the triangles in the stars of x and y depending on whether they share such an edge, they belong to both stars, or neither. The contraction of the edge connecting x and y produces two bi-gons and two triangles with the same three vertices in the *middle*, which are removed to get the portion of a 2-cycle on the *right*.

and y, we have two p-simplices with the same $p+1$ vertices, which, by definition, are two copies of the same p-simplex. By the logic of modulo-2 arithmetic, the two copies cancel each other. We prove that $\gamma|_{x\sim y}$ is a p-cycle.

▶ **Lemma 2.3.** *Let γ be a p-cycle and $x \neq y$ two vertices of γ. Then $\gamma|_{x\sim y}$ is a p-cycle.*

Proof. By construction, $\gamma' = \gamma|_{x\sim y}$ is a p-chain, so it suffices to show that every $(p-1)$-simplex belongs to an even number of p-simplices in γ'. Before the gluing, every $(p-1)$-simplex, σ, belongs to an even number of p-simplices in γ, by assumption.

Consider first the case in which σ is a $(p-1)$-face of a p-simplex $\tau \in \mathrm{St}_\gamma(x,y)$. We may assume that σ contains x but not y, else it becomes a $(p-2)$-simplex after gluing, which is no longer relevant. But then there is a second $(p-1)$-simplex, σ_2, that substitutes y for x and shares the other $p-1$ vertices with σ. After gluing, σ and σ_2 become one $(p-1)$-simplex, σ', which is shared by all p-simplices that share σ or σ_2. That number of such p-simplices is the sum of two even numbers minus 2, which is even, as required.

A configuration of two $(p-1)$-simplices that are glued to one is also possible even if they are not faces of a p-simplex in $\mathrm{St}_\gamma(x,y)$; see Figure 1 for an example. The argument is the same except there is no -2, at least not at first. The only remaining possibility for the number of p-simplices that share a $(p-1)$-simplex to change is if p-simplices cancel. But they cancel in pairs, which again preserves the parity of the number. ◀

We are interested in situations when γ and $\gamma' = \gamma|_{x\sim y}$ are homologous, which motivates us to introduce a $(p+1)$-chain that will be helpful in proving they are. Denote by $\gamma|^{x\sim y}$ the $(p+1)$-chain that is swept out by the p-simplices in $\mathrm{St}_\gamma(x) \cup \mathrm{St}_\gamma(y)$ as we move x and

y on straight lines to the new vertex, z. In the configuration displayed in Figure 1, this would be the 3-chain whose boundary consists of the 10 and 6 triangles in the left and right drawings, respectively. More formally, $\gamma|^{x \sim y} = z \cdot [\mathrm{St}_\gamma(x) \cup \mathrm{St}_\gamma(y)]$, in which multiplication means taking the cone over the p-simplices in the union of two stars or, in the combinatorial notation, adding z to the sets of $p+1$ vertices each. It is not difficult to see that the boundary of this $(p+1)$-chain is the sum of the two p-cycles.

▶ **Lemma 2.4.** *Let γ be a p-cycle, $x \neq y$ two vertices of γ, $\gamma' = \gamma|_{x \sim y}$, and $\Gamma = \gamma|^{x \sim y}$. Then $\partial\Gamma = \gamma + \gamma'$.*

Proof. The p-simplices in γ that neither belong to $\mathrm{St}_\gamma(x)$ nor to $\mathrm{St}_\gamma(y)$ also belong to γ'. Indeed, they exhaust $\gamma' \setminus \mathrm{St}_{\gamma'}(z)$, which implies $\gamma + \gamma' = \mathrm{St}_\gamma(x) \cup \mathrm{St}_\gamma(y) \cup \mathrm{St}_{\gamma'}(z)$. By construction of $\Gamma = \gamma|^{x \sim y}$, and the assumption that γ is a p-cycle, the right-hand side of this equation is the boundary of Γ. ◀

3 Snap Complex

For values $0 \leq s \leq t$, we write $\check{C}_s = \check{C}_s(A)$ for the Čech complex, $\beta_p(\check{C}_s)$ for the rank of its p-th (reduced) homology group, $\mathsf{H}_p(\check{C}_s)$, and $\beta_p(\check{C}_s, \check{C}_t)$ for the rank of the image of $\mathsf{H}_p(\check{C}_s)$ in $\mathsf{H}_p(\check{C}_t)$ induced by the inclusion $\check{C}_s \subseteq \check{C}_t$.

▶ **Definition 3.1.** *Let Ψ be a partition of $\mathbb{R}^d$ into cells, and call the supremum diameter of the sets in Ψ the* mesh *of the partition, denoted by* $\mathrm{mesh}(\Psi)$*. Let $q \colon A \to \Psi$ be defined by inclusion, and call $Q_s = q(\check{C}_s)$ the* snap complex *of the Čech complex $\check{C}_s$ along Ψ. In other words, the vertices of Q_s are the cells that contain at least one point of A, and a set $\{\psi_0, \psi_1, \ldots, \psi_p\} \subseteq \Psi$ of distinct cells is a p-simplex in Q_s iff there exist vertices $x_i \in \psi_i$, for $0 \leq i \leq p$, such that $\{x_0, x_1, \ldots, x_p\} \in \check{C}_s$.*

Note that q applied to $\check{C}_1$ is a surjective simplicial map to Q_1. Letting γ be a p-cycle in $\check{C}_1$, we call $\alpha = q(\gamma)$ its *image*, and γ a *preimage p-cycle* of α. Given two preimage p-cycles of α, they may or may not be homologous in $\check{C}_1$. The next lemma, however, guarantees that they are homologous in $\check{C}_{1+\mathrm{mesh}(\Psi)}$.

▶ **Lemma 3.2.** *Let $\varepsilon = \mathrm{mesh}(\Psi)$, α a p-cycle in Q_1, and γ, γ_3 two preimage p-cycles of α in $\check{C}_1$. Then $\gamma \sim \gamma_3$ in $\check{C}_{1+\varepsilon}$.*

Proof. We construct a sequence of homologous p-cycles that interpolates between γ and γ_3 in $\check{C}_{1+\varepsilon}$. There is an initial sequence interpolating between γ and γ_1, a middle sequence interpolating between γ_1 and γ_2, and a terminal sequence interpolating between γ_2 and γ_3.

The initial sequence reduces the number of vertices until the preimage p-cycle contains at most one vertex per cell in Ψ. Indeed, if γ has vertices $x \neq y$ in the same cell, then we can glue x and y, as described in Section 3, which produces a new p-chain, $\gamma' = \gamma|_{x \sim y}$. This operation introduces a new vertex, z. We are free to choose its location, and to facilitate the repetition of this argument, we choose it where y used to be. By Lemma 2.3, γ' is a p-cycle with one fewer vertices than γ. Let $\Gamma = \gamma|^{x \sim y}$ be the $(p+1)$-chain from Lemma 2.4. If all simplices in Γ belong to $\check{C}_{1+\varepsilon}$, then γ and γ' are homologous in $\check{C}_{1+\varepsilon}$, so assume that at least one simplex $\upsilon \in \Gamma$ does not belong to $\check{C}_{1+\varepsilon}$. Then $r(\upsilon) > 1 + \varepsilon$. By construction of Γ, and the choice of z's location in the cell that also contains x and y, there is a p-simplex τ in γ that has a vertex in every cell that contains a vertex of υ. Since the points in the same cell have distance at most ε from each other, Lemma 2.1 implies $r(\tau) > r(\upsilon) - \varepsilon > 1$. But then τ is not in $\check{C}_1$ and neither is γ, which contradicts the assumptions. This implies $\gamma \sim \gamma'$

in $\check{C}_{1+\varepsilon}$. By repeating the argument, all p-cycles in the initial sequence are homologous in $\check{C}_{1+\varepsilon}$ and, in particular, $\gamma \sim \gamma_1$ in $\check{C}_{1+\varepsilon}$.

The terminal sequence of p-cycles reduces the number of vertices of γ_3 if read from the end forward. Appealing again to Lemmas 2.1, 2.3, and 2.4, all p-cycles in the terminal sequence are homologous in $\check{C}_{1+\varepsilon}$ and, in particular, $\gamma_2 \sim \gamma_3$ in $\check{C}_{1+\varepsilon}$.

Since γ and γ_3 are preimage p-cycles of the same p-cycle, α, their vertices lie in the same cells of Ψ, and so do the vertices of γ_1 and γ_2. The middle sequence interpolates between the latter two p-cycles by changing one vertex at a time to a possibly different point in the same cell. Appealing to Lemmas 2.1, 2.3, and 2.4, the p-cycles in the middle sequence are again homologous in $\check{C}_{1+\varepsilon}$ and, in particular, $\gamma_1 \sim \gamma_2$ in $\check{C}_{1+\varepsilon}$. But now we have $\gamma \sim \gamma_1 \sim \gamma_2 \sim \gamma_3$ in $\check{C}_{1+\varepsilon}$ and therefore $\gamma \sim \gamma_3$ in $\check{C}_{1+\varepsilon}$, as claimed. ◀

Remark on Vietoris–Rips complexes. Lemma 3.2 generalizes to Vietoris–Rips complexes. Indeed, its proof generalizes provided we adapt Lemma 2.1, which in its current formulation is specific to Čech complexes. For the Vietoris–Rips complexes, we read the radius $r(B)$ as half of the maximum distance between any two vertices in B. Then it is still true that the difference in radii is bounded from above by the Hausdorff distance between the two sets of vertices. In other words, Lemma 2.1 also applies to Vietoris–Rips complexes.

By Lemma 3.2, any two preimage cycles of a cycle in Q_1 that already exist in $\check{C}_1$ are homologous in $\check{C}_{1+\varepsilon}$. We can therefore bound the persistent Betti numbers by the Betti number of the snap complex.

▶ **Corollary 3.3.** *With* $\varepsilon = \mathrm{mesh}(\Psi)$*, we have* $\beta_p(\check{C}_1, \check{C}_{1+\varepsilon}) \leq \beta_p(Q_1)$ *for every* p*.*

Proof. Recall that $\beta_p(\check{C}_1, \check{C}_{1+\varepsilon})$ is the rank of the persistent homology group that captures all p-cycles in $\check{C}_{1+\varepsilon}$ that already exist in $\check{C}_1$. It suffices to prove that the images of non-trivial p-cycles in $\check{C}_{1+\varepsilon}$ that already exist in $\check{C}_1$ are non-trivial p-cycles in Q_1. To derive a contradiction, let γ be a p-cycle in $\check{C}_1$ that is non-trivial in $\check{C}_{1+\varepsilon}$, and assume that $\alpha = q(\gamma)$ is trivial in Q_1. Hence, there exists a $(p+1)$-chain, A, in Q_1 with $\partial \mathrm{A} = \alpha$, and because q is surjective, there also exists a $(p+1)$-chain Γ in $\check{C}_1$ with $q(\Gamma) = \mathrm{A}$. Noting that γ and $\partial\Gamma$ are both p-cycles in $\check{C}_1$ whose images under q are equal to α, we obtain $\gamma \sim \partial\Gamma \sim 0$ in $\check{C}_{1+\varepsilon}$ by Lemma 3.2. This contradicts that γ is non-trivial in $\check{C}_{1+\varepsilon}$ and implies the claimed inequality. ◀

Remark about the right-hand side of the inequality. It is easy to see that the bound in Corollary 3.3 is not tight. Take for example three points equally spaced on a circle of radius strictly between 1 and $1+\varepsilon$. Then $\check{C}_1$ has a 1-cycle of three edges, while in $\check{C}_{1+\varepsilon}$ this cycle is filled by the triangle. Hence, $\beta_1(\check{C}_1, \check{C}_{1+\varepsilon}) = 0$, which is strictly smaller than $\beta_1(Q_1) = 1$. Note also that we cannot replace the upper bound $\beta_p(Q_1)$ in Corollary 3.3 with $\beta_p(Q_{1+\varepsilon})$ in general. The reason is that the image of a homologically non-trivial cycle in $\check{C}_{1+\varepsilon}$ may be homologically trivial in $Q_{1+\varepsilon}$, even if it already exists in $\check{C}_1$; see Figure 2 for an example. Indeed, we have $1 = \beta_1(\check{C}_1, \check{C}_{1+\varepsilon}) > \beta_1(Q_{1+\varepsilon}) = 0$ in this example.

Remark on Vietoris–Rips complexes. As mentioned earlier, Lemma 3.2 generalizes to Vietoris–Rips complexes. With this, the proof of Corollary 3.3 generalizes to Vietoris–Rips complexes.

4 The Upper Bound

We use a packing argument together with Corollary 3.3 to prove that for every $\varepsilon > 0$, the number of homology classes born before or at 1 and dying after $1+\varepsilon$ in the Čech filtration

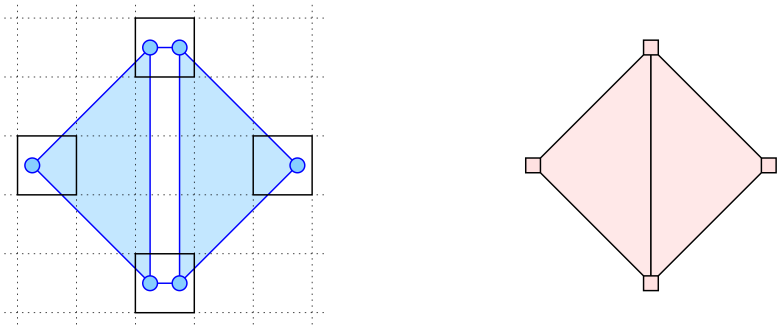

Figure 2: *Left:* the Čech complex for six points inside four cells in the partition of the plane. Assuming the two triangles are isosceles, right-angled, and have smallest enclosing circles of radius $1+\varepsilon$, the narrow rectangle between the two triangles has a smallest enclosing circle with radius strictly larger than $1+\varepsilon$. Hence, the boundary of the convex hexagon that passes through the six points is a non-trivial 1-cycle in $\check{C}_{1+\varepsilon}$, and for $\varepsilon \leq \sqrt{2}-1$, it already exists in $\check{C}_1$. *Right:* the image of the hexagon is a quadrangle in the snap complex. Its boundary is a trivial 1-cycle in $Q_{1+\varepsilon}$ because the rectangle collapses to a single edge shared by the images of the two triangles.

of n points in $\mathbb{R}^d$ is bounded from above by a constant times n. This constant depends on ε and d but not on n.

▶ **Theorem 4.1.** *For every $\varepsilon > 0$, there exists $c = c(\varepsilon, d)$ such that $\beta_p(\check{C}_1, \check{C}_{1+\varepsilon}) \leq c \cdot n$.*

Proof. We partition $\mathbb{R}^d$ into translates of $[0, \varepsilon/\sqrt{d})^d$. The diameter of every cell is ε, so we call this partition Ψ and apply Corollary 3.3. Fixing $\psi_0 = [0, \varepsilon/\sqrt{d})^d \in \Psi$, the cells that are connected to ψ_0 by an edge in Q_1 must contain a point at distance at most 2 from a point in ψ_0. Therefore, such cells lie inside the hypercube $[-2-\varepsilon/\sqrt{d}, 2+2\varepsilon/\sqrt{d})^d$. Its volume is $(4+3\varepsilon/\sqrt{d})^d$. Comparing this with the volume of a single cell, which is $(\varepsilon/\sqrt{d})^d$, the number of such cells is at most

$$C(\varepsilon, d) = \frac{(4+3\varepsilon/\sqrt{d})^d}{(\varepsilon/\sqrt{d})^d} = \left(3 + \frac{4\sqrt{d}}{\varepsilon}\right)^d. \tag{4}$$

To span a p-simplex in Q_1, we pick the fixed cell and add p from the at most $C = C(\varepsilon, d)$ cells within the mentioned distance. We thus have at most $\binom{C}{p}n$ p-simplices in Q_1, where n is the number of ways we can fix the first cell. The number of p-simplices is an upper bound on the p-th Betti number. By Corollary 3.3, the same upper bound applies to the number of p-cycles born before or at 1 and dying after $1+\varepsilon$. We have non-zero persistent Betti numbers only for $p < d$, so $c = \binom{C}{p} < 2^C$ is a constant for which the claimed inequality holds. ◀

Remark on Vietoris–Rips complexes. Since Corollary 3.3 generalizes to Vietoris–Rips complexes, so does Theorem 4.1.

5 Discussion

The main result of this paper is a linear upper bound on the number of holes in the Čech complex of n points in $\mathbb{R}^d$ that persist from radius $r = 1$ to $r = 1+\varepsilon$, in which ε is a fixed constant strictly larger than 0. The upper bound generalizes to the Alpha complex and the Vietoris–Rips complex and thus holds for three of the classic types of complexes used in topological data analysis [1, 3]. The work reported in this short note raises a number of questions, and we mention two.

The first natural question is how small we can make $\varepsilon > 0$ in our linear upper bound on the persistent Betti numbers when we think of ε as a function of n that tends to 0, rather than a fixed positive constant. The construction in [4] shows that the maximum p-th Betti number of a Čech complex with n vertices in $\mathbb{R}^d$ is $\Theta(n^m)$, with $m = \min\{p+1, \lceil d/2 \rceil\}$. A detailed look at the analysis shows that for even d, the persistence of the counted cycles is proportional to $1/n^2$, and for odd d, it is proportional to $1/n^4$, in which we simplify by assuming that d is a constant. In other words, the lower bound extends to the persistent Betti numbers, $\beta_p(\check{C}_1, \check{C}_{1+\varepsilon})$, provided $\varepsilon = o(1/n^2)$ and $\varepsilon = o(1/n^4)$ for even and odd d, respectively. The upper bound for the constant of proportionality in Theorem 4.1 depends on ε and d in a way that suggests it grossly over-estimates the number of holes that persist. Can this upper bound be improved to showing that the polynomially many holes in the lower bound construction of Edelsbrunner and Pach [4] are asymptotically as persistent as possible? Alternatively, can this lower bound construction be improved to increase the persistence of the holes, which currently is $\Theta(1/n^2)$ in even and $\Theta(1/n^4)$ in odd dimensions?

The second question is motivated by the third author's quest to prove the large deviation principle for persistent Betti numbers and persistence diagrams of random Čech filtrations (cf. [7, 8]). Let A be a finite set of points in a large d-dimensional cube, partitioned into points L and R to the left and right of a vertical hyperplane, respectively. Our goal is to approximate the persistent Betti numbers of the Čech filtration of A by the sum of those for the Čech filtrations of L and R. More specifically, we desire a bound on the absolute difference between $\beta_p(\check{C}_1(A), \check{C}_{1+\varepsilon}(A))$ and $\beta_p(\check{C}_1(L), \check{C}_{1+\varepsilon}(L)) + \beta_p(\check{C}_1(R), \check{C}_{1+\varepsilon}(R))$. Letting $M \subseteq A$ contain the points at distance at most $2(1+\varepsilon)$ from the hyperplane, the vertices of every simplex in $\check{C}_{1+\varepsilon}(A) \setminus (\check{C}_{1+\varepsilon}(L) \cup \check{C}_{1+\varepsilon}(R))$ must belong to M, so it is natural to estimate the absolute difference in terms of M: assuming $\varepsilon > 0$ is a fixed constant and $0 \leq p \leq d$, is it true that there exists a constant such that

$$|\beta_p(\check{C}_1(A), \check{C}_{1+\varepsilon}(A)) - \beta_p(\check{C}_1(L), \check{C}_{1+\varepsilon}(L)) - \beta_p(\check{C}_1(R), \check{C}_{1+\varepsilon}(R))| \leq \text{const} \cdot \text{card}\, M? \quad (5)$$

In other words, is the absolute difference between these persistent Betti numbers bounded from above by a constant times the number of points in the narrow strip next to the hyperplane? The absolute difference is of course bounded by the number of simplices spanned by these points (see, e.g., [6, Lemma 2.11] and [8, Proposition 16]), but this only implies that the left-hand side of (5) is bounded from above by $f_p(\check{C}_1(M)) + f_{p+1}(\check{C}_{1+\varepsilon}(M))$, and thus by $2(\text{card}\, M)^{p+2}$. Therefore, the significance of the inequality (5) lies in the linear bound with respect to the number of points in the narrow strip.

References

1. G. Carlsson. Topology and data. *Bull. Amer. Math. Soc.* **46** (2009), 255–308.
2. B. Delaunay. Sur la sphère vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennykh Nauk* **7** (1934), 793–800.
3. H. Edelsbrunner and J.L. Harer. *Computational Topology. An Introduction.* Amer. Math. Soc., Providence, Rhode Island, 2010.
4. H. Edelsbrunner and J. Pach. Maximum Betti numbers of Čech complexes. *In* "Proc. 40th Intl. Sympos. Comput. Geom., 2024", 53:1–53:14.
5. M. Goff. Extremal Betti numbers of Vietoris–Rips complexes. *Discrete Comput. Geom.* **46** (2011), 132-155.
6. Y. Hiraoka, T. Shirai and K.D. Trinh. Limit theorems for persistence diagrams. *Ann. Appl. Probab.* **28** (2018), 2740-2780.
7. C. Hirsch and T. Owada. Large deviation principle for geometric and topological functionals and associated point processes. *Ann. Appl. Probab.* **33** (2023), 4008-4043.
8. S. Kanazawa, Y. Hiraoka, J. Miyanaga and K. Tsunoda. Large deviation principle for persistence diagrams of random cubical filtrations. *J Appl. Comput. Topol.* (2024).
9. G.M. Ziegler. *Lectures on Polytopes.* Grad. Texts in Math. **152**, Springer, Berlin, Germany, 1995.