



# Extremal lifetimes of persistent cycles

Nicolas Chenavier<sup>1</sup> · Christian Hirsch<sup>2</sup>

Received: 10 October 2020 / Revised: 23 August 2021 / Accepted: 20 September 2021 /

Published online: 30 October 2021

© The Author(s) 2021

## Abstract

Persistent homology captures the appearances and disappearances of topological features such as loops and cavities when growing disks centered at a Poisson point process. We study extreme values for the lifetimes of features dying in bounded components and with birth resp. death time bounded away from the threshold for continuum percolation and the coexistence region. First, we describe the scaling of the minimal lifetimes for general feature dimensions, and of the maximal lifetimes for cavities in the Čech filtration. Then, we proceed to a more refined analysis and establish Poisson approximation for large lifetimes of cavities and for small lifetimes of loops. Finally, we also study the scaling of minimal lifetimes in the Vietoris-Rips setting and point to a surprising difference to the Čech filtration.

**Keywords** Topological data analysis · Persistent Betti numbers · Poisson approximation

**AMS 2000 Subject Classifications** 60K35 · 82C22

## 1 Introduction

One of the key challenges in statistics is to provide scientifically rigorous methods to detect structure among noise. While for low-dimensional datasets, we do indeed have a versatile toolkit at our disposal, the situation is radically different in high dimensions. In this context, *topological data analysis (TDA)* is emerging

---

✉ Christian Hirsch  
hirsch@math.au.dk

Nicolas Chenavier  
nicolas.chenavier@univ-littoral.fr

<sup>1</sup> Université du Littoral Côte d'Opale, EA 2597, LMPA, 50 rue Ferdinand Buisson, Calais 62228, France

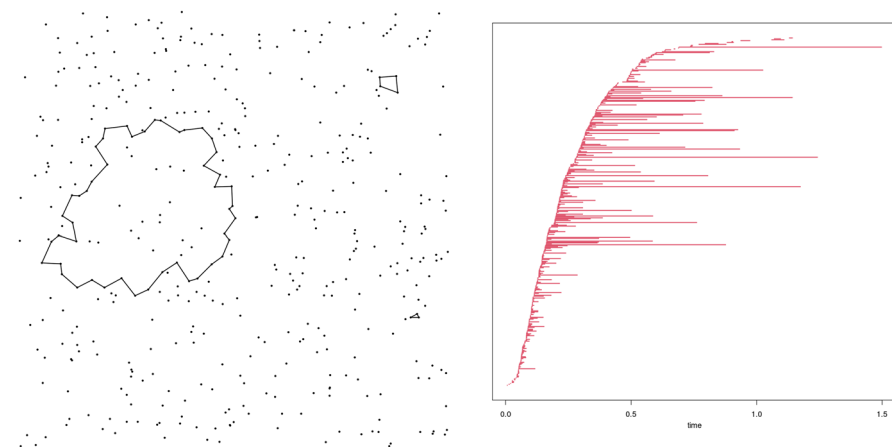
<sup>2</sup> Department of Mathematics, Aarhus University, Ny Munkegade 118, 8000 Aarhus C, Denmark

as an exciting novel approach. The simple and effective idea is to leverage invariants from algebraic topology to extract insights from data. While initially TDA started off as an eccentric mathematical idea, it is now applied in a wide variety of disciplines, ranging from astronomy and biology to finance and materials science Gidea and Katz (2018); Pranav et al. (2016); Rocks et al. (2020); Saadatfar et al. (2017).

A core component of TDA is *persistent homology*. Here, we start growing balls centered at the points of a dataset. At the beginning, all points are in isolated components, but as the balls grow, new topological features such as loops or cavities may appear. As the radii increase, these features disappear again, until finally everything is contained in a single connected component without loops or cavities. Through this mechanism, we can associate a birth time and a death time with each such feature. The representation of such features through their birth and death time can be visualized succinctly through barcodes such as the one shown in Fig. 1 (right).

Despite the spectacular successes of TDA in the application domains, major parts of it are still lacking a rigorous statistical foundation. For instance, when analyzing persistent homology, practitioners often look for features living for exceptionally long periods of time, and then draw conclusions if they do occur. However, how can we decide whether the observed long life times come from genuinely interesting phenomena and are not a mere incarnation of chance?

To answer this question, we need to understand how long life times behave under the null model of complete spatial randomness, i.e., a Poisson point process. Figure 1 illustrates cycles of a maximal, typical and minimal life times in a bounded sampling window. In this regard, a breakthrough was achieved in Bobrowski et al. (2017), where the order of the longest life time occurring in a large window was determined on a multiplicative scale. Despite the significance of this discovery, it is still difficult to build a rigorous statistical test, knowing



**Fig. 1** Left panel: Čech-features of a maximal (left), typical (right top) and minimal (right bottom) life times. Right panel: associated barcode

only the order of the life time. A further interesting cross-connection is that for 0-dimensional features, the longest life time corresponds to the longest edge in the minimum spanning tree whose asymptotic was established in Penrose (1997).

In the present paper, we move one step closer to statistical applications and establish Poisson approximation results for extremal life times of loops and cavities in large sampling windows. More precisely, in Theorem 1 we determine the scaling of the maximal life time of  $(d - 1)$ -dimensional features, i.e., cavities, in the Čech filtration. Moreover, we show through the Chen-Stein method that after normalization, the locations of exceedances over suitable thresholds converge to a homogeneous Poisson process. Thus, the investigation is in a similar spirit as the extreme value analysis of geometric characteristics of inradii in tessellations Calka and Chenavier (2014); Chenavier (2014). However, after proceeding from the general set-up of the Chen-Stein method to more concrete tasks, the geometric analyses of the two problems are entirely different. Furthermore, the Chen-Stein argument is also used in Owada and Adler (2017) to analyze the spatial distribution of topological cavities in the extremes.

Another vigorous research stream in spatial extremes concerns limit results for exceptionally small structures Schulte and Thäle (2012). Our second and third main results, concern this setting. More precisely, we determine the minimal life time of features in any dimension both in the Vietoris-Rips (VR) and in the Čech (Č) filtration. Our analysis reveals a striking difference in the scaling of minimal features between the VR- and the Č-filtration. Often the VR-filtration is chosen as an approximation to the Č-filtration, when the latter becomes prohibitively time-consuming to compute. Hence, our result provides an example for the limitations of this approximation. However, since the minimal lifetimes of features tend to 0 in increasing sampling windows, there is no contradiction with classical approximation results such as (Ghrist 2008, Lemma 2.1.)

The rest of the manuscript is organized as follows. First, in Sect. 2, we fix the notation and state the main results. Section 3 illustrates in broad strokes how the Chen-Stein method can be used to tackle the assertion. This involves two critical steps, namely identifying the correct scaling and excluding multiple exceedances, which will be established in Sects. 4 and 5, respectively. To ease the flow of reading, we have postponed overly technical volume computations into an appendix.

## 2 Model and main results

Let  $\mathcal{P} := \{X_i\}_{i \geq 1}$  be a Poisson point process in  $\mathbb{R}^d$  with intensity 1. We let

$$O_r(\mathcal{P}) := \bigcup_{i \geq 1} B_r(X_i)$$

denote the Boolean model at a radius  $r$ , and write

$$V_r(\mathcal{P}) := \mathbb{R}^d \setminus O_r(\mathcal{P})$$

for the corresponding vacant set. Our first main result concerns the lifetimes of *cavities*, i.e., of bounded connected components of  $V_r(\mathcal{P})$ . When the growing radii create a new cavity at a radius  $r$ , we say that  $r$  is the *birth time* of the new cavity. Moreover, the *death time* of a cavity is the smallest radius  $r > 0$  when it is covered completely by the Boolean model  $O_r(\mathcal{P})$ . So far, this definition is still ambiguous: when a new cavity is born by splitting an existing one into two, it is not clear which of the newly created vacant components should determine its death time. Therefore, it is a common convention to impose that cavities born first die last.

Now, we enumerate the cavities as  $\{J_i^*\}_{i \geq 1}$  and associate with each such cavity its lifetime  $L_i^* > 0$  and the point  $Z_i^* \in \mathbb{R}^d$  as the last point that is covered at the death time. In this way,  $\{(Z_i^*, L_i^*)\}_{i \geq 1}$  defines a stationary marked point process of cavities together with their lifetimes.

Long-range interactions coming from percolation effects make it difficult to study this marked point process in full generality. For cavities born in large vacant components, the seemingly innocent convention that features born first die last may induce long-range interactions in order to decide which birth times correspond to which death times. This makes analyzing extremes prohibitively challenging: we are not aware of any form of Poisson approximation result reflecting critical phenomena in percolation contexts.

Since concepts from continuum percolation are used throughout the paper, we now discuss the most important ideas. Loosely speaking, continuum percolation concerns the analysis of unbounded connected components in the occupied set  $O_r(\mathcal{P})$ . One of the central findings of percolation theory is the occurrence of a phase transition. More precisely, we say that  $O_r(\mathcal{P})$  *percolates* if it contains an unbounded connected component, and then introduce

$$r_c^O := \inf\{r \geq 0 : \mathbb{P}(O_r(\mathcal{P}) \text{ percolates}) > 0\}$$

as the *critical radius for occupied percolation*. The phase transition is non-trivial in the sense that  $0 < r_c^O < \infty$ , see Meester and Roy (1996). Above  $r_c^O$ , the occupied set intersects a large bounded sampling window  $W_n := [-n^{1/d}/2, n^{1/d}/2]^d$  in a unique giant connected component with diameter of order  $n^{1/d}$ ; whereas the other components are of logarithmic order, see (Penrose 2003, Chapter 10.6). Moreover, also the percolation theory for the vacant set  $V_r(\mathcal{P})$  has been developed, and we write

$$r_c^V := \sup\{r \geq 0 : \mathbb{P}(V_r(\mathcal{P}) \text{ percolates}) > 0\}$$

for the *critical radius for vacant percolation*. Most of percolation theory has been developed in parallel for the occupied, and the vacant set. However, the logarithmic order of the non-giant vacant connected components in  $W_n$  is not yet available from literature in the regime  $[r_c^O, r_c^V]$  of coexistence of unbounded occupied and vacant components.

To state a clean Poisson-approximation theorem that applies to the bulk of the considered cavities, we restrict  $\{J_i^*\}_{i \geq 1}$  to the family of cavities  $\{J_i\}_{i \geq 1}$  with birth time outside  $[r_c^O - \varepsilon_c, r_c^V + \varepsilon_c]$  for some arbitrarily small but positive  $\varepsilon_c > 0$ . As made precise in Lemma 1 below, the striking advantage of this restriction is that the bounded connected components are of logarithmic size. However, as a consequence,

our analysis does not cover the giant cycles considered in Bobrowski and Skraba (2020), and it would be interesting to extend the Poisson approximation results in that direction. To state our main results, we first give some notation. For each  $n \geq 1$ , we write

$$\gamma_n := (\kappa_d^{-1} \log n)^{1/d},$$

where  $\kappa_d$  is the volume of the unit ball in  $\mathbb{R}^d$ . Moreover, for fixed  $\tau > 0$ , we denote by  $\Phi_{n,\tau}^{\max}$  the rescaled *point process of exceedances above a threshold*  $u_{n,\tau}^{\max}$ , i.e.

$$\Phi_{n,\tau}^{\max} := \{n^{-1/d} Z_i : L_i > u_{n,\tau}^{\max}\}_{Z_i \in W_n} \subseteq W_1,$$

where the threshold  $u_{n,\tau}^{\max}$  is determined by

$$\tau = \mathbb{E}[\#\{Z_i \in W_n : L_i > u_{n,\tau}^{\max}\}]. \tag{1}$$

In Lemma 2 in Appendix 1, we check that  $u_{n,\tau}^{\max}$  is well-defined in the sense that for every  $\tau > 0$  there exists  $n_0(\tau)$  such that for all  $n \geq n_0(\tau)$  there is a unique  $u_{n,\tau}^{\max}$  satisfying Eq. (1). Considering this point process is classical in extreme value theory since it captures the location of the cavities with large lifetimes. Our first result claims that  $u_{n,\tau}^{\max}$  is asymptotically equivalent to  $\gamma_n$  (in the sense that the ratio of these quantities converges to 1) and that  $\Phi_{n,\tau}^{\max}$  is approximately a homogeneous Poisson point process  $\Phi_\tau$  with intensity  $\tau$  in  $W_1$ .

We write  $\Rightarrow$  and  $\xrightarrow{p}$  for convergence in distribution and in probability, respectively.

**Theorem 1 (Poisson approximation for maximal lifetimes of cavities)** *For every  $\tau > 0$ ,*

$$\lim_{n \rightarrow \infty} \frac{u_{n,\tau}^{\max}}{\gamma_n} = 1 \quad \text{and} \quad \Phi_{n,\tau}^{\max} \Rightarrow \Phi_\tau.$$

As a corollary, we obtain the scaling of the maximal lifetime.

**Corollary 1 (Scaling of maximal lifetimes of cavities)** *It holds that*

$$\gamma_n^{-1} \max_{Z_i \in W_n} L_i \xrightarrow{p} 1.$$

*The proofs of Theorem 1 and Corollary 1 are given in Sect. 3.*

**Remark 1** Theorem 1 can be slightly improved by varying the threshold  $\tau$  so that  $\Phi_{n,\tau}^{\max}$  will be enhanced to a marked point process describing both the location of the exceedance as well as the associated lifetime. After an appropriate rescaling, this enhanced point process converges to a homogeneous Poisson point process with intensity 1 in  $W_1 \times (0, \infty)$ . This type of result gives more information than Theorem 1 since it describes the joint distribution of lifetimes exceeding different thresholds. For readability of the manuscript, we refrain from discussing here the more technical statement and proof.

**Remark 2** Theorem 1 reveals a striking difference to the multiplicative persistence approach of measuring lifetimes in Bobrowski et al. (2017), where the scaling is  $(\log n / \log \log n)^{1/(d-1)} \geq \gamma_n$ . That article considers the ratio between the death time and the birth time, whereas we work with the difference. We stress that the two quantities are related to a similar phenomenon, but they measure it in a different way:  $\gamma_n$  is measured in distance units, while the left-hand-side is a ratio, which has no units. The two scalings are different as the multiplicative approach is sensitive with respect to cycles with a very small birth time. In certain application contexts this may not be desirable. For instance, technical limitations of measuring instruments may prevent the precise determination of very small birth times with a sufficiently high accuracy. Hence, it is also valuable to know the scaling of the difference between the death time and the birth time.

Next, we proceed to minimal lifetimes, where we also derive Poisson approximation results. However, here we do not restrict to the cavities, but consider more general topological features. To make this precise, we now discuss briefly the most fundamental concepts behind persistent homology, referring the reader to Boissonnat et al. (2018) for an excellent textbook treatment on this topic.

We first give the precise definition of Čech (Č-) and Vietoris-Rips (VR-) complexes, which will be of central importance throughout the paper. Both the VR- and the Č-complex describe collections of simplices in  $\mathbb{R}^d$ . A  $k$ -simplex  $\{x_0, \dots, x_k\} \subseteq \mathbb{R}^d$  is contained in the VR-complex at a level  $r > 0$  if the distance between any pair of points is smaller than  $2r$ , i.e., if  $\max_{i,j \leq k} |x_i - x_j| \leq 2r$ . The simplex is included in the Č-complex if  $\bigcap_{i \leq k} B_r(x_i) \neq \emptyset$ , where  $B_r(x_i)$  denotes the Euclidean ball with radius  $r$  centered at  $x_i$ .

Based on these complexes one can compute homology groups whose ranks reveal subtle topological properties of the underlying set. For instance, the rank of the homology group in dimension  $d - 1$  of the Č-complex at level  $r$  on a finite dataset  $\varphi \subseteq \mathbb{R}^d$  counts the number of cavities, i.e., the number of bounded connected components in the vacant space  $V_r(\varphi)$ . Similarly, the ranks of the homology groups in dimensions 0 and 1 correspond to the numbers of connected components and loops in the occupied set  $O_r(\varphi)$ . In a general dimension  $k \leq d$ , we also speak of  $k$ -dimensional features. It is a consequence of the nerve lemma that the homology groups of the Č-complex at level  $r$  agree with the homology groups of the occupied set  $O_r(\varphi)$  (Björner 1995, Theorem 10.7). Although the homology groups of the VR-complex are not as nicely related to the topology of  $O_r(\varphi)$  or  $V_r(\varphi)$ , they are still often preferred on computational grounds.

A drawback of the methodology described in the preceding paragraph is that it is tied to a single value of  $r$ . Persistent homology removes this shortcoming by tracking at which level of the parameter  $r$  the topological features appear (the birth time), and when they disappear again (the death time). More precisely, when thinking of the Č- and VR-complexes as families indexed by the parameter  $r$ , we also speak of the Č- and VR-filtrations. Conveniently, there is a precise algorithmic description of when a birth occurs and when a death occurs. Each time we add a simplex to the filtration that actually changes the homology, it either increases the homology in the dimension of the simplex or it decreases the homology in one dimension lower

(Boissonnat et al. 2018, Algorithm 9). In the first case, we add a *positive simplex*, which corresponds to the *birth* of a *persistent  $k$ -cycle*. In the second case, we add a *negative simplex* corresponding to the *death* of the persistent  $k$ -cycle. Defining the lifetime  $L_{i,k}$  of the  $i$ th persistent  $k$ -cycle as the difference between the death and the birth time, this algorithm gives rise to a family of birth-and-death times  $\{L_{i,k}\}_{i \geq 1}$ . This collection is frequently visualized through *barcodes*, where the birth and death times mark the starting points and the end points of the bars, see Fig. 1.

Similar to the setting of cavities, also for general feature dimensions, the long-range correlations close to the critical radius  $r_c^O$  would pose a formidable obstacle for proving Poisson approximation. Therefore, we henceforth fix again a time window  $[r_c^O - \epsilon_c, r_c^O + \epsilon_c]$  and consider only persistent  $k$ -cycles with death time outside  $[r_c^O - \epsilon_c, r_c^O + \epsilon_c]$  and that die in a bounded connected component. By restricting to such persistent  $k$ -cycles, we avoid the need of having to define persistent homology for the entire Euclidean space  $\mathbb{R}^d$ , which would require a form of infinite-dimensional persistent homology. Furthermore, for instance by taking the center point of the  $(k + 1)$ -simplex whose insertion kills the persistent  $k$ -cycle, we attach a center point  $Z_{i,k} \in \mathbb{R}^d$  to any such persistent  $k$ -cycle with lifetime  $L_{i,k}$ . Hence, we again obtain a stationary marked point process  $\{(Z_{i,k}, L_{i,k})\}_{i \geq 1}$ . In cases where the underlying filtration is not clear from the context, we also write  $\{(Z_{i,k,\check{C}}, L_{i,k,\check{C}})\}_{i \geq 1}$  or  $\{(Z_{i,k,VR}, L_{i,k,VR})\}_{i \geq 1}$ .

Now, similar to the setting of the cavities, fixing a feature dimension  $1 \leq k \leq d - 1$ , and given  $\tau > 0$ , we consider the rescaled *point process of undershoots*

$$\Phi_{n,k,\tau,\check{C}}^{\min} := \{n^{-1/d}Z_{i,k,\check{C}} : L_{i,k,\check{C}} < u_{n,k,\tau,\check{C}}^{\min}\}_{Z_{i,k,\check{C}} \in W_n} \subseteq W_1,$$

where the threshold  $u_{n,k,\tau,\check{C}}^{\min}$  is determined by

$$\tau = \mathbb{E}[\#\{Z_{i,k,\check{C}} \in W_n : L_{i,k,\check{C}} < u_{n,k,\tau,\check{C}}^{\min}\}]. \tag{2}$$

Next, we show that the minimal lifetime of persistent  $k$ -cycles is of order  $n^{-2}$ , independent of  $k$ . Moreover, the rescaled point process of undershoots is approximately a homogeneous Poisson point process  $\Phi_\tau$  with intensity  $\tau$  in  $W_1$ . However, due to involved geometrical configurations occurring in the  $\check{C}$ -filtration, we prove the point-process convergence only for a specific feature dimension, namely  $k = 1$  (loops).

**Theorem 2 (Minimal lifetimes,  $\check{C}$ -filtration)** *Let  $1 \leq k \leq d - 1$ . Then, for every  $\tau > 0$ ,*

(i) *See below*

$$\lim_{n \rightarrow \infty} \frac{\log u_{n,k,\tau,\check{C}}^{\min}}{\log n} = -2 \quad \text{and} \quad (\log n)^{-1} \min_{Z_{i,k} \in W_n} \log L_{i,k} \xrightarrow{P} -2.$$

(ii) *Moreover,  $\Phi_{n,k,\tau,\check{C}}^{\min} \Rightarrow \Phi_\tau$  for  $k = 1$ .*

**Remark 3** On a technical level, the reason for the restriction to  $k = 1$  in case (ii) is due to the need to bound the probability of multiple undershoots. For  $k = 1$ , two persistent  $k$ -cycles with short lifetimes are characterized by two 1-simplices causing the births of the cycles, and two 2-simplices causing the deaths of the cycles. Already among these 10 points several could coincide and conspire to create potentially intricate geometric configuration favoring the occurrence of multiple undershoots. For  $k = 1$ , it is still possible to go through the different cases by hand, but for larger  $k$  a different approach would be needed.

The construction of  $\Phi_{n,k,\tau,\check{C}}^{\min}$  outlined above does not only work for the  $\check{C}$ - but also for the VR-filtration  $\Phi_{n,k,\tau,VR}^{\min}$ . The difference in filtration manifests itself prominently in a different scaling for the threshold. Again, we consider a threshold  $u_{n,k,\tau,VR}^{\min}$  such that

$$\tau = \mathbb{E}[\#\{Z_{i,k,VR} \in W_n : L_{i,k,VR} < u_{n,k,\tau,VR}^{\min}\}]. \tag{3}$$

**Theorem 3 (Minimal lifetimes – VR-filtration)** *Let  $1 \leq k \leq d - 1$ . Then, for every  $\tau > 0$ ,*

$$\lim_{n \rightarrow \infty} \frac{\log u_{n,k,\tau,VR}^{\min}}{\log n} = -1 \quad \text{and} \quad \Phi_{n,k,\tau,VR}^{\min} \Rightarrow \Phi_{\tau}.$$

The orders of the minimal lifetimes in the  $\check{C}$ - and VR-filtrations are different because cycles with 0 lifetime in the VR-filtration can have a very small positive lifetime in the  $\check{C}$ -filtration. More specifically, for instance in dimension 2, the proof of Theorem 2 reveals that cycles with minimal lifetimes can be taken to be acute triangles that almost have a right angle. Since triangles never have a positive lifetime in the VR-filtration, the persistent cycles with minimal lifetime live substantially longer. As a consequence, we obtain again the scaling of the minimal lifetime.

**Corollary 2 (Scaling of minimal lifetimes – VR-filtration)** *It holds that*

$$(\log n)^{-1} \min_{Z_{i,k,VR} \in W_n} \log L_{i,k,VR} \xrightarrow{P} -1.$$

**Remark 4** The proof of Theorem 1 relies on the topological property that the lifetimes of the persistent  $(d - 1)$ -cycles in the  $\check{C}$ -filtration coincide with the lifetimes of the connected components in the vacant set of a union of balls. In particular, dealing with maximal lifetimes in different dimensions or in the VR-filtration would require a different approach. Moreover, a priori, it is not clear whether typical lifetimes in the VR-filtration should be shorter or longer than those in the  $\check{C}$ -filtration. However, the specific construction in the proof of Theorem 1 suggests it seems plausible that asymptotically the analog of the  $\check{C}$ -filtration threshold  $u_{n,\tau}^{\max}$  in the VR-filtration should be smaller than by a constant factor. Loosely speaking, the characteristic of



persistent  $(d - 1)$ -cycles is a very late death time rather than a very early birth time, and persistent  $(d - 1)$ -cycles are dying later in the  $\check{C}$ - than in the VR-filtration.

### 3 Poisson approximation

The proofs of the Poisson approximation in Theorems 1–3 rely fundamentally on the Chen–Stein method in the form of (Arratia et al. 1990, Theorem 1). To make the presentation self-contained, we sketch now how to tune the general machinery to the setting of extremal life times. The delicate problem-specific conditions are then verified in Sects. 4 and 5.

More precisely, proceeding as in Chenavier (2014), we discretize the space into blocks in order to make the problem amenable to the Chen–Stein framework in the form of (Arratia et al. 1990, Theorem 1). For the convenience of the reader, we outline the most important steps for the maximal life times, noting that the arguments carry over seamlessly to the setting of minimal life times.

The first step is to leverage the percolation conditions described in Sect. 2 in order to introduce a discretization that is coarse enough so that the life times of cavities in non-adjacent blocks are independent. Hence, the size of the discretization is essentially determined by the diameter of the largest cavity that we expect to see in the window  $W_n$ . However, due to the convention that features born first should die last, the actual block size that needs to be considered is a bit larger. Now, we show that vacant or occupied components in continuum percolation whose birth time is bounded away from the coexistence region are at most of poly-logarithmic size with a probability tending to 1. More precisely, recalling that we center cavities at the point which is covered last, we let  $R_n^V(r)$  be the maximal diameter of all bounded connected components of  $V_r(\mathcal{P})$  centered in  $W_n$ . Similarly, we let  $R_n^O(r)$  be the maximal diameter of all bounded connected components of  $O_r(\mathcal{P})$  centered in  $W_n$ . Then, we fix  $\varepsilon_c > 0$  and put

$$E_n^V := \left\{ \sup_{r \notin [r_c^O - \varepsilon_c, r_c^V + \varepsilon_c]} R_n^V(r) \leq (\log n)^2 \right\}, \quad \text{and} \quad E_n^O := \left\{ \sup_{r \notin [r_c^O - \varepsilon_c, r_c^O + \varepsilon_c]} R_n^O(r) \leq (\log n)^2 \right\}.$$

recalling that  $r_c^V, r_c^O$  are the critical radii for vacant and occupied continuum percolation.

We show that the events  $E_n^V$  and  $E_n^O$  occur with a probability tending to 1. As we will explain in the proof of Lemma 1 below, for a fixed sub- or supercritical radius  $r$ , this follows from classical continuum percolation theory. However, since in the definitions of  $E_n^V$  and  $E_n^O$  the radius may vary, we need a small discretization argument. As mentioned in Sect. 2, whereas strong quantitative uniqueness results for occupied phase transition have been established in (Penrose 2003, Sect. 10.6), the analogs in vacant percolation are not readily available in the literature. Hence, for vacant percolation, we do not make any claims in the radius range  $[r_c^O - \varepsilon_c, r_c^V + \varepsilon_c]$ . Note that this issue is only relevant in dimension  $d > 2$ .

**Lemma 1** *It holds that  $\mathbb{P}(E_n^V) \wedge \mathbb{P}(E_n^O) \geq 1 - n^{-4d}$  for all sufficiently large  $n \geq 1$ .*

**Proof** We first explain how to proceed in the vacant setting. We say that an event occurs whp if there exists  $c_0 > 0$  such that its complement occurs with probability at most  $c_0 n^{-5d}$  for large  $n$ . To bound  $R_n^V(r)$ , we distinguish two different regimes for the radii: (1)  $r \geq r_c^V + \varepsilon_c$ , and (2)  $r \leq r_c^O - \varepsilon_c$ .

$r \geq r_c^V + \varepsilon_c$ . All vacant connected components are almost surely bounded, and the size of these components decreases with the radius. Hence, it suffices to show that  $R_n^V(r_c^V + \varepsilon_c) \leq (\log n)^2$  with high probability. We partition  $W_n$  into  $a_n := \lceil 4dn^{1/d} / \varepsilon_c \rceil^d$  sub-boxes  $\mathbf{C}_1, \dots, \mathbf{C}_{a_n}$  with centers  $x_1, \dots, x_{a_n}$  and of side length at most  $\varepsilon_c / (4d)$ . If the vacant set  $V_{r_c^V + \varepsilon_c}(\mathcal{P})$  contains some point of a box  $\mathbf{C}_i$ , then  $\mathbf{C}_i \subseteq V_{r_c^V}(\mathcal{P})$ , where  $r_c^- := r_c^V + \varepsilon_c / 2$ . In particular, since  $\mathcal{P}$  is stationary,

$$\mathbb{P}(R_n^V(r_c^V + \varepsilon_c) \geq (\log n)^2) \leq \sum_{i \leq a_n} \mathbb{P}(R^V(r_-, x_i) \geq (\log n)^2) = a_n \mathbb{P}(R^V(r_-, o) \geq (\log n)^2),$$

where  $R^V(r_-, x_i)$  denotes the diameter of the connected component of  $V_{r_c^-}(\mathcal{P})$  at  $x_i$ . By (Duminil-Copin et al. 2020, Theorem 1.6), there exists  $c > 0$  such that  $\mathbb{P}(R^V(r_-, o) \geq (\log n)^2) \leq \exp(-c(\log n)^2) = n^{-c \log n}$  for all sufficiently large  $n$ . In particular, the event  $\{R_n^V(r_c^V + \varepsilon_c) \leq (\log n)^2\}$  occurs whp.

$r \leq r_c^O - \varepsilon_c$ . All occupied components are bounded and their size decreases with decreasing radius. Now, every bounded vacant component is enclosed by an occupied component. Although this occupied component need not intersect  $W_n$ , its diameter is at least as large as the enclosed vacant component. In particular, if the occupied component contains a point  $x \in \mathbb{R}^d$  and does not intersect  $W_{3n}$ , then it encompasses  $W_n$  so that its diameter is at least  $|x|/4$ . We extend the partition of  $W_{3n}$  from the previous paragraph into a partition  $\cup_{i \geq 1} \mathbf{C}_i = \mathbb{R}^d$  of the entire space. Hence, if  $R_n^V(r) \geq (\log n)^2$  for some  $r \leq r_c^O - \varepsilon_c$ , then there exists a cube  $\mathbf{C}_i$  such that  $\mathbf{C}_i \subseteq O_{r_+}(\mathcal{P})$ , where  $r_+ := r_c^O - \varepsilon_c / 2$ . In particular, again by stationarity, for sufficiently large  $n$ ,

$$\begin{aligned} \mathbb{P}\left(\sup_{r \leq r_c^O - \varepsilon_c} R_n^V(r) \geq (\log n)^2\right) &\leq \sum_{i: \mathbf{C}_i \subseteq W_{3n}} \mathbb{P}(R^O(r_+, x_i) \geq (\log n)^2) + \sum_{i: \mathbf{C}_i \not\subseteq W_{3n}} \mathbb{P}(R^O(r_+, x_i) \geq |x_i|/4) \\ &\leq a_n \mathbb{P}(R^O(r_+, o) \geq (\log n)^2) + \sum_{k \geq n^{1/(2d)}} \#\{\mathbf{C}_i : \lfloor |x_i|/4 \rfloor = k\} \mathbb{P}(R^O(r_+, o) \geq k), \end{aligned}$$

where  $R^O(r_+, x_i)$  denotes the diameter of the connected component of  $O_{r_+}(\mathcal{P})$  at  $x_i$ . We now conclude as in the previous case, by applying (Duminil-Copin et al. 2020, Theorem 1.4) instead of (Duminil-Copin et al. 2020, Theorem 1.6). This proves that the event  $\{\sup_{r \leq r_c^O - \varepsilon_c} R_n^V(r) \leq (\log n)^2\}$  occurs whp.

Next, we deal with the occupied setting. Again, we distinguish different radii regimes, namely: (1)  $r \leq r_c^O - \varepsilon_c$ , (2)  $r \geq r_c^V + \varepsilon_c$ , and (3)  $r_c^O + \varepsilon_c \leq r \leq r_c^V + \varepsilon_c$ . Since the arguments in the first two cases are almost identical to the vacant setting, we only give details on case (3).

$r_c^O + \varepsilon_c \leq r \leq r_c^V + \varepsilon_c$ . Here, we discretize the possible radii that can occur. More precisely, we subdivide the interval  $[r_c^O + \varepsilon_c, r_c^V + \varepsilon_c]$  into  $K_n = O(n^{8d})$  parts  $\{r_i\}_{i \leq K_n}$  such that  $r_{i+1} - r_i \leq n^{-8d}$ . By applying (Penrose 2003, Theorem 10.20), we obtain that

for each fixed  $r_i$ , with probability at least  $1 - \exp(-c(\log n)^2)$ , all bounded connected components at level  $r_i$  centered in  $W_n$ , are of diameter at most  $(\log n)^2$ . The fact that we may choose  $c$  independently of the  $r_i$ 's relies on the assumption that the  $r_i$ 's are bounded away from the critical value  $r_c^O$ . Since we discretize into  $K_n = O(n^{8d})$  different radii, we conclude that whp all bounded connected components at the levels  $r_i, i \leq K_n$  are of diameter at most  $(\log n)^2$ . Hence, it suffices to show that whp for any  $r \in [r_c^O + \varepsilon_c, r_c^V - \varepsilon_c]$  the components at level  $r$  correspond to the components at level  $r_i$  for some  $i \leq K_n$ . In other words, we claim that whp, for each  $i \leq K_n$  there exists at most one pair  $\{X_j, X_k\} \subseteq W_n$  with  $|X_j - X_k|/2 \in (r_i, r_{i+1})$ .

To that end, we fix  $i \leq K_n$ , and set

$$F := \left\{ \{|X_j - X_k|, |X_{j'} - X_{k'}|\} \subseteq (2r_i, 2r_{i+1}) : \text{for some pw. distinct } X_j, X_k, X_{j'}, X_{k'} \in W_n \right\}.$$

Then, by the Slivnyak-Mecke formula (Last and Penrose 2016, Theorem 4.4),

$$\begin{aligned} \mathbb{P}(F) &\leq \mathbb{E} \left[ \sum_{\substack{X_j, X_k, X_{j'}, X_{k'} \in W_n \\ \text{pw. distinct}}} \mathbb{1} \left\{ \{|X_j - X_k|, |X_{j'} - X_{k'}|\} \subseteq (2r_i, 2r_{i+1}) \right\} \right] \\ &= \int_{W_n^4} \mathbb{1} \left\{ \{|x_1 - x_3|, |x_2 - x_4|\} \subseteq (2r_i, 2r_{i+1}) \right\} dx_1 dx_2 dx_3 dx_4 \\ &\leq \int_{W_n^2} |B_{2r_{i+1}}(x_1) \setminus B_{2r_i}(x_1)| |B_{2r_{i+1}}(x_2) \setminus B_{2r_i}(x_2)| dx_1 dx_2 \\ &\leq 2^{2d} n^2 \kappa_d^2 (r_{i+1}^d - r_i^d)^2, \end{aligned}$$

where the final expression is in  $O(n^{2-16d})$ . A similar calculation bounds the probability of the event

$$F' := \left\{ \{|X_j - X_k|, |X_j - X_{k'}|\} \subseteq (2r_i, 2r_{i+1}) : \text{for some pw. distinct } X_j, X_k, X_{k'} \in W_n \right\}.$$

Hence, by the union bound, only with probability at most  $O(n^{2-8d})$ , there exists some  $i \leq K_n$  for which there is more than one pair  $\{X_j, X_k\} \subseteq W_n$  with  $|X_j - X_k|/2 \in (r_i, r_{i+1})$ .

Notice that on the event  $E_n^V$ , for each cavity with  $Z_i \in W_n$ , we have  $L_i < (\log n)^2$ . Now, we partition the sampling window  $W_n$  into  $\mathcal{N}_n \in O(n/(\log n)^{2d})$  blocks  $\{C_j\}_{j \leq \mathcal{N}_n}$  of side length

$$s_n := 2(\log n)^2.$$

To prove that the point process of exceedances converges to a Poisson point process, we have to estimate the number of exceedances. To do it, we first give some notation. More precisely, writing  $S(x, r)$  for the diameter of the connected component of the vacant set  $V_r(\mathcal{P})$  containing  $x \in \mathbb{R}^d$ , we set

$$S_i := \sup_{r \notin [r_c^O - \varepsilon_c, r_c^V + \varepsilon_c] : S(Z_i, r) < \infty} S(Z_i, r)$$

as the diameter of the largest bounded cavity containing  $Z_i$ .

For each  $j \leq \mathcal{N}_n$  and  $B \subseteq W_1$ , we write

$$N_{j,n}(B) := \#\{Z_i \in \mathbf{C}_j \cap (n^{1/d}B) : L_i > u_{n,\tau}^{\max}, S_i \leq (\log n)^2\}$$

for the number of cavities in  $\mathbf{C}_j \cap (n^{1/d}B)$  with diameter-bound at most  $(\log n)^2$ . We denote by

$$N'_n(B) := \#\{j \leq \mathcal{N}_n : N_{j,n}(B) \geq 1\}$$

the number of blocks intersected by  $n^{1/d}B$  for which there exists at least one exceedance with diameter-bound lower than  $(\log n)^2$ . Setting  $s_n^+ := (3s_n)^d$ , the main difficulty is to exclude multiple exceedances in small boxes of size  $s_n^+$ . In what follows, we denote by  $W_{s_n^+}$  a cube with volume  $s_n^+$  and by  $N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max})$  the number of exceedances in  $W_{s_n^+}$ , i.e.

$$N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max}) := \#\left\{Z_i \in W_{s_n^+} : L_i > u_{n,\tau}^{\max}\right\}. \tag{4}$$

**Proposition 1 (No multiple exceedances)** *Let  $\tau > 0$ .*

- (i) *There exists  $\alpha > 0$  such that  $\mathbb{P}(N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max}) \geq 2) \in O(n^{-1-\alpha})$ .*
- (ii) *The same type of property holds for the minimal life time in the VR-filtration with  $1 \leq k \leq d - 1$ , and for the minimal lifetime in the  $\check{C}$ -filtration with  $k = 1$ .*

Proposition 1 is shown in Sect. 5. We now prove the Poisson approximation in Theorem 1.

**Proof (Theorem 1, Poisson approximation)** Let  $\tau > 0$  and  $B \subseteq W_1$  Borel. First, note that

$$\mathbb{E}[\Phi_{n,\tau}^{\max}(B)] = |B| \mathbb{E}[\#\{Z_i \in W_n : L_i > u_{n,\tau}^{\max}\}] = \tau |B|.$$

Hence, by Kallenberg’s theorem (see e.g. Theorem A.1, p.309 in Leadbetter et al. (1983)), it suffices to prove that

$$\mathbb{P}(\Phi_{n,\tau}^{\max}(B) = 0) \xrightarrow[n \rightarrow \infty]{} e^{-\tau|B|}.$$

To achieve this goal, we first note that

$$\begin{aligned} \mathbb{P}(\Phi_{n,\tau}^{\max}(B) \neq N'_n(B)) &\leq \mathbb{P}((E_n^V)^c) + \sum_{j \leq \mathcal{N}_n} \mathbb{P}(\#\{Z_i \in \mathbf{C}_j : L_i > u_{n,\tau}^{\max}\} \geq 2) \\ &\leq \mathbb{P}((E_n^V)^c) + \mathcal{N}_n \mathbb{P}(N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max}) \geq 2), \end{aligned}$$

where the last step uses that  $\mathcal{P}$  is stationary. Therefore, thanks to Lemma 1 and Proposition 1, we have  $\mathbb{P}(\Phi_{n,\tau}^{\max}(B) \neq N'_n(B)) \xrightarrow[n \rightarrow \infty]{} 0$ . Thus, we have to prove that

$\mathbb{P}(N'_n(B) = 0)$  converges to  $e^{-\tau|B|}$ . To do it, it suffices show that  $\mathbb{E}[N'_n(B)] \xrightarrow{n \rightarrow \infty} \tau|B|$  and that  $d_{TV}(N'_n(B), F_n) \xrightarrow{n \rightarrow \infty} 0$ , where  $F_n$  is a Poisson random variable with parameter  $\mathbb{E}[N'_n(B)]$  and where  $d_{TV}(Y, Z)$  denotes the total variation between two random variables  $Y, Z$  with integer values, i.e.,

$$d_{TV}(Y, Z) = 2 \sup_{A \subseteq \mathbb{N}} |\mathbb{P}(Y \in A) - \mathbb{P}(Z \in A)|.$$

First, we deal with  $\mathbb{E}[N'_n(B)]$  and note that  $\tau|B| - \mathbb{E}[N'_n(B)] = \mathbb{E}[\Phi_{n,\tau}^{\max}(B) - N'_n(B)]$ . Now, we use the decomposition

$$\begin{aligned} \Phi_{n,\tau}^{\max}(B) - N'_n(B) &= \#\{i : Z_i \in n^{1/d}B, L_i > u_{n,\tau}^{\max}, S_i \leq (\log n)^2\} - \#\{j \leq \mathcal{N}_n : N_{j,n}(B) \geq 1\} \\ &\quad + \#\{i : Z_i \in n^{1/d}B, L_i > u_{n,\tau}^{\max}, S_i > (\log n)^2\}. \end{aligned} \tag{5}$$

Discretizing over all blocks  $C_j$ , the first quantity on the right-hand side is

$$\sum_{j: N_{j,n}(B) \geq 1} \left( \#\{i : Z_i \in C_j \cap n^{1/d}B, L_i > u_{n,\tau}^{\max}, S_i \leq (\log n)^2\} - 1 \right) = \sum_{j: N_{j,n}(B) \geq 2} (N_{j,n}(B) - 1),$$

because the term inside the sum vanishes for  $N_{j,n}(B) = 1$ . Moreover, it follows from the Hölder’s inequality, that for each  $p, q$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ ,

$$\begin{aligned} \mathbb{E} \left[ \sum_{j: N_{j,n}(B) \geq 2} (N_{j,n}(B) - 1) \right] &\leq \mathcal{N}_n \sup_{j \geq 1} \mathbb{E}[N_{j,n}(B)^p]^{1/p} \mathbb{P}(N_{j,n}(B) \geq 2)^{1/q} \\ &\leq \mathcal{N}_n \sup_{j \geq 1} \mathbb{E}[N_{j,n}(B)^p]^{1/p} \mathbb{P}(N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max}) \geq 2)^{1/q}. \end{aligned}$$

Now, covering the cube  $C_j$  by blocks  $Q_1, \dots, Q_{k_n}$  of side length 1, with  $k_n \in O(s_n^d)$ , we have

$$\mathbb{E}[N_{j,n}(B)^p] \leq k_n^p \sum_{l \leq k_n} \mathbb{E} \left[ \#\{i : Z_i \in Q_l, L_i > u_{n,\tau}^{\max}\}^p \right] = k_n^{p+1} \mathbb{E} \left[ N_{1,\check{C}}^{\max}(u_{n,\tau}^{\max})^p \right].$$

In Sect. 4 below we show that  $u_{n,\tau}^{\max} \geq 2d$  for  $n$  sufficiently large so that there is at most one hole with lifetime exceeding  $u_{n,\tau}^{\max}$  in a box of side length 1. Thus,  $N_{1,\check{C}}^{\max}(u_{n,\tau}^{\max}) \in \{0, 1\}$  and according to (1), we have  $\mathbb{E} \left[ N_{1,\check{C}}^{\max}(u_{n,\tau}^{\max})^p \right] = \mathbb{E} \left[ N_{1,\check{C}}^{\max}(u_{n,\tau}^{\max}) \right] = \frac{\tau}{n}$ . Therefore, for some constant  $c$ ,

$$\mathbb{E} \left[ \sum_{j: N_{j,n}(B) \geq 2} (N_{j,n}(B) - 1) \right] \leq c \mathcal{N}_n k_n^{1+1/p} n^{-1/p} \mathbb{P}(N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max}) \geq 2)^{1/q} \in O(s_n^{d/p} n^{-a/q}),$$

where the last line comes from Proposition 1. This shows that the first expectation in the right-hand side of Eq. (5) converges to 0 as  $n$  goes to infinity. Applying again the Hölder’s inequality, it follows from Lemma 1 that the second expectation of the right-hand side of Eq. (5) also converges to 0. This proves that  $\mathbb{E}[N'_n(B)] \xrightarrow{n \rightarrow \infty} \tau|B|$ .

It remains to prove that  $d_{TV}(N'_n(B), F_n)$  converges to 0. The main idea to do it is to apply the Poisson approximation result (Arratia et al. 1990, Theorem 1),

which is based on the Chen-Stein method. To this end, consider a finite or countable collection  $(Y_\alpha)_{\alpha \in I}$  of  $\{0, 1\}$ -valued random variables. Let  $p_\alpha = \mathbb{P}(Y_\alpha = 1) > 0$ ,  $p_{\alpha\beta} = \mathbb{P}(Y_\alpha = 1, Y_\beta = 1)$ . Moreover, suppose that for each  $\alpha \in I$ , there is a set  $V_\alpha \subseteq I$  that contains  $\alpha$ . Finally, put

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in V_\alpha} p_\alpha p_\beta, \quad b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in V_\alpha} p_{\alpha\beta}, \quad b_3 = \sum_{\alpha \in I} \mathbb{E} \left[ \left| \mathbb{E}[Y_\alpha - p_\alpha | \sigma(Y_\beta : \beta \notin V_\alpha)] \right| \right]. \tag{6}$$

**Theorem 4** (Arratia, Goldstein, Gordon) *Let  $W = \sum_{\alpha \in I} Y_\alpha$ , and assume  $\lambda := \mathbb{E}(W) \in (0, \infty)$ . Then*

$$d_{TV}(W, \text{Po}(\lambda)) \leq 2(b_1 + b_2 + b_3).$$

We apply below Theorem 4 to the collection of random variables  $(\mathbb{1}\{N_{j,n}(B) \geq 1\})_{j \leq \mathcal{N}_n}$ . To do it, for each  $j \leq \mathcal{N}_n$ , we let

$$V_j := \{j\} \cup \{j' \leq \mathcal{N}_n : C_{j'} \text{ is adjacent to } C_j\}.$$

Since  $N_{j,n}(B)$  is equal to the number of cavities in  $C_j \cap (n^{1/d}B)$  with lifetime larger than  $u_{n,\tau}^{\max}$  and with diameter lower than  $(\log n)^2$  and since the side length of each block is  $(\log n)^2$ , the random variables  $N_{j,n}(B)$  and  $N_{j',n}(B)$  are independent provided that  $j' \notin V_j$ . In particular, the term  $b_3$  appearing in Eq. (6) vanishes. Now, to bound  $b_1$ , we note that

$$\mathbb{P}(N_{j,n}(B) \geq 1) \leq \mathbb{E}[\#\{Z_i \in C_j : L_i > u_{n,\tau}^{\max}\}] = \frac{s_n^d \tau}{n},$$

for each  $j \leq \mathcal{N}_n$ . Thus, since  $\#V_j = 3^d$  and  $\mathcal{N}_n s_n^d = n$ ,

$$b_1 = \sum_{j \leq \mathcal{N}_n} \sum_{j' \in V_j} \mathbb{P}(N_{j,n}(B) \geq 1) \mathbb{P}(N_{j',n}(B) \geq 1) \leq \mathcal{N}_n 3^d \left(\frac{s_n^d \tau}{n}\right)^2 \in O(s_n^d n^{-1}).$$

Finally, to bound  $b_2$ , we note that, for each  $j \in \mathcal{N}_n$  and  $j' \in V_j \setminus \{j\}$ ,

$$\mathbb{P}(N_{j,n}(B) \geq 1, N_{j',n}(B) \geq 1) \leq \mathbb{P}(N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max}) \geq 2)$$

since the side length of  $W_{s_n^+}$  equals  $3s_n$ . Hence, by Proposition 1,

$$b_2 = \sum_{\substack{j \leq \mathcal{N}_n \\ j' \in V_j \setminus \{j\}}} \mathbb{P}(N_{j,n}(B) \geq 1, N_{j',n}(B) \geq 1) \leq \mathcal{N}_n (3^d - 1) \mathbb{P}(N_{s_n^+, \check{C}}^{\max}(u_{n,\tau}^{\max}) \geq 2) \in O(s_n^{-d} n^{-\alpha}).$$

Thus, Theorem 4 gives that  $d_{TV}(N'_n(B), F_n) \in O(s_n^{-d} n^{-\alpha})$ . This concludes the proof of Theorem 1.

In a similar way, we obtain the Poisson approximation asserted in Theorems 2 and 3, relying on a correspondingly defined  $E_n^{\text{O}}$  instead of  $E_n^{\text{V}}$ .

**Proof** (Corollary 1) Let  $\varepsilon > 0$ . Then,  $\max_{Z_i \in W_n} L_i \geq (1 + \varepsilon)\gamma_n$  implies that  $\Phi_{n,\tau}^{\max} \neq \emptyset$  a.s. provided that  $n$  is sufficiently large so that  $u_{n,\tau}^{\max} \leq (1 + \varepsilon)\gamma_n$ . In particular, for any  $\tau > 0$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\max_{Z_i \in W_n} L_i \geq (1 + \varepsilon)\gamma_n) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(\Phi_{n,\tau}^{\max} \neq \emptyset) = \mathbb{P}(\Phi_\tau \neq \emptyset) = 1 - \exp(-\tau),$$

and the right-hand side tends to 0 as  $\tau \rightarrow 0$ . Similarly,  $\max_{Z_i \in W_n} L_i \leq (1 - \varepsilon)\gamma_n$  implies that  $\Phi_{n,\tau}^{\max} = \emptyset$  a. s. provided that  $n$  is sufficiently large so that  $u_{n,\tau}^{\max} \geq (1 - \varepsilon)\gamma_n$ . In particular, for any  $\tau > 0$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\max_{Z_i \in W_n} L_i \leq (1 - \varepsilon)\gamma_n) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(\Phi_{n,\tau}^{\max} = \emptyset) = \mathbb{P}(\Phi_\tau = \emptyset) = \exp(-\tau)$$

and the right-hand side tends to 0 as  $\tau \rightarrow \infty$ .

### 4 Scalings of $u_{n,\tau}^{\max}$ , $u_{n,k,\tau,\check{C}}^{\min}$ and $u_{n,k,\tau,VR}^{\min}$

In the present section, we derive the scalings of  $u_{n,\tau}^{\max}$ ,  $u_{n,k,\tau,VR}^{\min}$  and  $u_{n,k,\tau,\check{C}}^{\min}$  asserted in Theorems 1–3.

To begin with, we deal with  $u_{n,\tau}^{\max}$ . As observed in (Bobrowski et al. 2017, Remark 3.2), the upper bound follows from the results in Kahle (2011): according to the scaling of the coverage radius in Hall (1985), above  $\gamma_n$  the union of balls covers  $W_n$ . Hence, there cannot be a cavity centered at  $W_n$  and with lifetime exceeding the coverage radius. The lower bound resorts to a construction inspired by [Lemma 5.1] Bobrowski et al. (2017).

*Scaling of  $u_{n,\tau}^{\max}$  (proof of Theorem 1, scaling).*

We recall from Eq. (4) that for each  $\ell > 0$ , we write  $N_{n,\check{C}}^{\max}(\ell) := \#\{Z_i \in W_n : L_i > \ell\}$  for the number of cavities in  $W_n$  with lifetime at least  $\ell$ .

$u_{n,\tau}^{\max} / \gamma_n \leq 1 + \varepsilon$ . Let  $\varepsilon > 0$ . Writing  $\ell_n^{\max,+} := (1 + \varepsilon)\gamma_n$ , we have to prove that

$$\lim_{n \rightarrow \infty} \mathbb{E} [N_{n,\check{C}}^{\max}(\ell_n^{\max,+})] = 0.$$

By stationarity, it is sufficient to show that  $\mathbb{E} [N_{1,\check{C}}^{\max}(\ell_n^{\max,+})] \in o(n^{-1})$ . In order to bound the expected number of cavities with lifetime at least  $\ell_n^{\max,+}$  that are centered in  $W_1$ , let  $R := \min\{|x| : x \in \mathcal{P}\}$  be the distance of the closest point of  $\mathcal{P}$  to the origin. We claim that the death time of any cavity centered in  $W_1$  is at most  $R + \sqrt{d}$ . Indeed, let  $Z_i \in W_1$  be the center of a cavity. Since the diameter of the cube  $W_1$  equals  $\sqrt{d}$ , there exists  $x \in \mathcal{P}$  such that  $|x - Z_i| \leq R + \sqrt{d}$ . In other words, the cavity dies at the latest at time  $R + \sqrt{d}$ . Next, there can be at most one cavity with lifetime exceeding  $\ell_n^{\max,+}$  and centered in  $W_1$ . Indeed, assume that  $Z_{i_1}, Z_{i_2} \in W_1$  were the centers of two such cavities, and assume that  $Z_{i_2}$  is born after  $Z_{i_1}$ . Then,  $Z_{i_2}$  is born after time  $(R - \sqrt{d})_+$ . Indeed, at time  $(R - \sqrt{d})_+$  the entire cube  $W_1$  is still uncovered so that  $Z_{i_2}$  belongs still to the cavity defined by  $Z_{i_1}$ . Combining this with the previous observation on the death time leads to the contradiction that such a cavity would

have lifetime at most  $2\sqrt{d}$ . Hence, invoking the void probability of the Poisson process  $\mathcal{P}$ ,

$$\mathbb{E}\left[N_{1,\check{C}}^{\max}(\ell_n^{\max,+})\right] \leq \mathbb{P}(\mathcal{P} \cap B_{\ell_n^{\max,+}} - \sqrt{d}(o) = \emptyset) = \exp(-\kappa_d(\ell_n^{\max,+} - \sqrt{d})^d),$$

and the right-hand side is at most  $\exp(-(1 + \varepsilon)^d \log(n)(1 - \sqrt{d}/\ell_n^{\max,+})^d) \in o(n^{-1})$ .  $u_{n,\tau}^{\max}/\gamma_n \geq 1 - \varepsilon$ . With  $\ell_n^{\max,-} := (1 - \varepsilon)\gamma_n$ , we claim that

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[N_{n,\check{C}}^{\max}(\ell_n^{\max,-})\right] = \infty.$$

As indicated at the beginning of this section, we modify a construction of [Lemma 5.1] Bobrowski et al. (2017) by covering a discretized annulus of radius  $\ell_n^{\max,-}$  by Poisson points, see Fig. 2. To that end, let  $W(v) := v + W_1$  be the unit block centered at a lattice point  $v \in \mathbb{Z}^d$ . Now, let

$$E_b := \{\mathcal{P} \cap W(v) \neq \emptyset \text{ for all } v \in \mathbb{Z}^d \text{ with } |v|_\infty = 2\lceil \gamma_n \rceil\}$$

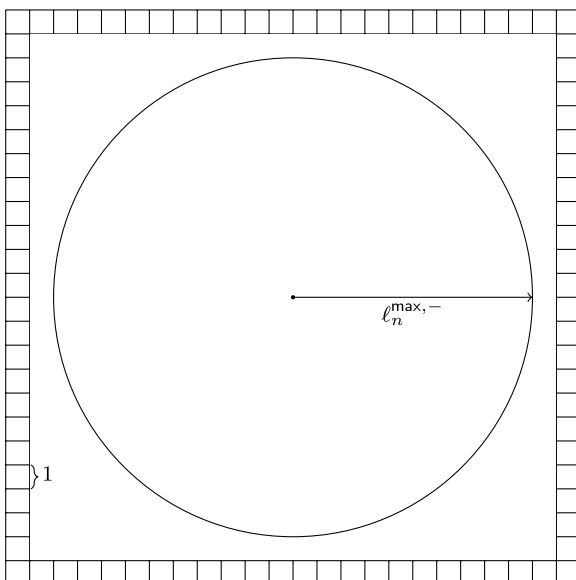
denote the event that each unit block in a  $d_\infty$ -annulus of radius  $2\lceil \gamma_n \rceil$  contains at least one Poisson point. We also write  $m_n = m_{n,\varepsilon}$  for the number of these blocks and note that  $m_n \in O((\log n)^{(d-1)/d})$ . Moreover, we let

$$E_d := \{\mathcal{P} \cap B_{\ell_n^{\max,-}}(o) = \emptyset\}$$

denote the event that there are no Poisson points in the ball with radius  $\ell_n^{\max,-}$ . Then, under the event  $E_b \cap E_d$  we find a cavity with lifetime at least  $\ell_n^{\max,-}$ .

Setting  $\varepsilon' := 1 - (1 - \varepsilon)^d$  the independence property of the Poisson process gives that

**Fig. 2** Template for cavity with maximal lifetime





$$\mathbb{P}(E_b \cap E_d) = \mathbb{P}(E_b)\mathbb{P}(E_d) \geq e^{-cm_n n^{-1+\epsilon'}}$$

for some  $c > 0$ . Finally, we can take the template described by  $E_b \cap E_d$  and shift it to a different location in the window  $W_n$ . Since those configurations are of logarithmic extent, we can arrange at least  $a_n := n^{1-\epsilon'/2}$  of them disjointly in  $W_n$ . In particular, since  $m_n \in O((\log n)^{(d-1)/d})$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E} [N_{n,\check{C}}^{\max}(\ell_n^{\max,-})] \geq \lim_{n \rightarrow \infty} a_n e^{-cm_n n^{-1+\epsilon'}} = \infty.$$

Next, we analyze the scaling of  $u_{n,k,\tau,\check{C}}^{\min}$ . Since we now deal with short rather than long lifetimes, the proof structure is converse to that for  $u_{n,\tau}^{\max}$ . That is, in the lower bound we analyze configurations leading to small lifetimes, whereas the upper bound relies on a specific construction. We also write

$$N_{n,k,\check{C}}^{\min}(\ell) := \#\{Z_{i,k,\check{C}} \in W_n : L_{i,k,\check{C}} < \ell\}$$

for the number of persistent  $k$ -cycles in  $W_n$  with lifetime at most  $\ell$ . Henceforth, let

$$d_{0\dots k} := d(x_0, \dots, x_k) := \inf\{r > 0 : \bigcap_{x \in \{x_0, \dots, x_k\}} B_r(x) \neq \emptyset\} \tag{7}$$

denote the *filtration time* of the simplex  $\{x_0, \dots, x_k\}$ , i.e., the first time it appears in the  $\check{C}$ -filtration. In particular,  $d_{ij} = |x_i - x_j|/2$ .

Moreover, if  $x_0, \dots, x_{k-1} \in \mathbb{R}^d$  are affinely independent and  $r > d(x_0, \dots, x_{k-1})$ , then in every  $k$ -dimensional hyperplane containing  $x_0, \dots, x_{k-1}$ , there are precisely two  $k$ -dimensional balls with radius  $r$  containing  $x_0, \dots, x_{k-1}$  on their boundary, and we let  $D_r(x_0, \dots, x_{k-1})$  denote the union of all such balls. Then, for  $r, \ell > 0$ , we introduce the crescent

$$D_{r,\ell}(x_0, \dots, x_{k-1}) := D_{r+\ell}(x_0, \dots, x_{k-1}) \setminus D_r(x_0, \dots, x_{k-1}),$$

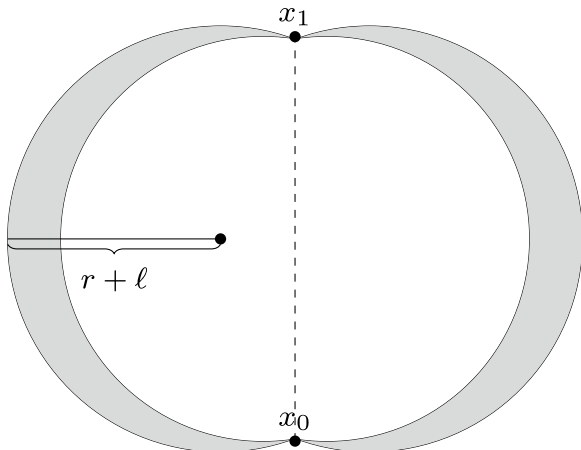
illustrated in Fig. 3 for  $k = 2$ . Having a  $k$ -simplex  $x_0, \dots, x_k$  with filtration time  $s$  means that there exist  $i_0, \dots, i_m \leq k$  and an  $m$ -dimensional ball  $B_s(P)$  centered at some  $P \in \mathbb{R}^d$  such that  $x_0, \dots, x_k \subseteq B_s(P)$  and  $x_{i_0}, \dots, x_{i_m} \in \partial B_s$ . In particular,  $s \in [r, r + \ell]$  means that  $x_{i_m} \in D_{r,\ell}(x_{i_0}, \dots, x_{i_{m-1}})$ .

Scaling of  $u_{n,k,\tau,\check{C}}^{\min}$  (proof of Theorem 2, scaling).  $u_{n,k,\tau,\check{C}}^{\min} \geq n^{-2-\epsilon}$  **and**  $\mathbb{P}(\min_{Z_{i,k,\check{C}} \in W_n} L_{i,k,\check{C}} \leq n^{-2-\epsilon}) \rightarrow 0$ . It suffices to prove the first assertion. We write  $\ell_n^{\min,-} := n^{-2-\epsilon}$  and claim that

$$\lim_{n \rightarrow \infty} \mathbb{E} [N_{n,k,\check{C}}^{\min}(\ell_n^{\min,-})] = 0.$$

For the proof, it will be highly convenient to review the notion of critical simplices discussed in Bobrowski (2019). More precisely, when building the Čech complex by adding simplices according to their filtration time, only a restricted class of simplices can lead to changes the homology. These *critical  $k$ -simplices*  $\sigma$  are determined by the property that the center of the  $k$ -dimensional ball containing the vertices of  $\sigma$  on its boundary is contained in  $\sigma$ , and that the interior of this

**Fig. 3** The crescent set  $D_{r,\ell}(x_0, x_1)$



ball does not contain any other points from  $\mathcal{P}$  [Lemma 2.4] Bobrowski (2019). When a persistent  $k$ -cycle with lifetime  $\ell > 0$  is born at time  $r = d_{0\dots k}$ , then there exist Poisson points  $X_0, \dots, X_k$  such that  $\cup_{i \leq k} B_r(X_i)$  covers the simplex spanned by these points. When a persistent  $k$ -cycle dies at time  $t_d = d_{0\dots k} + \ell$ , then there are Poisson points  $X'_0, \dots, X'_{k+1}$  such that the associated simplex is covered by time  $d_{0\dots k} + \ell$ . We stress that every persistent  $k$ -cycle only dies when a new critical  $(k + 1)$ -simplex joins the complex. After reordering there exists  $-1 \leq j \leq k$  such that  $X'_i = X_i$  for  $i \leq j$  and  $X'_i \neq X_i$  for  $i \geq j + 1$ . Then, the previous observations can be succinctly summarized as

$$X'_{k+1} \in D_{d_{0\dots k}, \ell}(X_0, \dots, X_j, X'_{j+1}, \dots, X'_k).$$

Since we work with persistent  $k$ -cycles in bounded connected components, even in the case  $j = -1$  all of the considered Poisson points are at a distance at most  $s_n$  from  $X_0$ .

Now, we invoke the Slivnyak-Mecke formula from [Theorem 4.4] Last and Penrose (2016). It shows that the expected number of Poisson points  $X_0, \dots, X_k, X'_{j+1}, \dots, X'_{k+1}$  satisfying

$$X'_{k+1} \in D_{d_{0\dots k}, \ell}(X_0, \dots, X_j, X'_{j+1}, \dots, X'_k)$$

is at most

$$\begin{aligned} & \int_{W_{2n} \times B_{s_n}(x_0)^k} \int_{B_{s_n}(x_0)^{k-j}} \int_{D_{d_{0\dots k}, \ell}(x_0, \dots, x_j, x'_{j+1}, \dots, x'_k)} 1 dx'_{k+1} d(x'_{j+1}, \dots, x'_k) d(x_0, \dots, x_k) \\ &= \int_{W_{2n} \times B_{s_n}(x_0)^k} \int_{B_{s_n}(x_0)^{k-j}} |D_{d_{0\dots k}, \ell}(x_0, \dots, x_j, x'_{j+1}, \dots, x'_k)| d(x'_{j+1}, \dots, x'_k) d(x_0, \dots, x_k). \end{aligned}$$

Now, by Lemma 3 in the appendix, the volume of the integrand is bounded above by  $c s_n^d \sqrt{\ell}$  so that

$$\int_{W_{2n} \times B_{s_n}(x_0)^k} \int_{B_{s_n}(x_0)^{k-j}} |D_{d_0 \dots k, \ell}(x_0, \dots, x_j, x'_{j+1}, \dots, x'_k)| d(x'_{j+1}, \dots, x'_k) d(x_0, \dots, x_k) \leq c \kappa_d^{2k-j} s_n^{(2k-j+1)d} \int_{W_{2n}} \sqrt{\ell} dx_0.$$

Recalling that  $s_n = 2(\log n)^2$ , for  $\ell \leq \ell_n^{\min, -}$  the latter expression tends to 0 as  $n \rightarrow \infty$ .  $u_{n,k,\tau,\check{C}}^{\min} \leq n^{-2+\varepsilon}$  and  $\mathbb{P}(\min_{Z_{i,k,\check{C}} \in W_n} L_{i,k,\check{C}} \geq n^{-2+R}) \rightarrow 0$ .

The proof idea is to construct persistent  $k$ -cycles with a short lifetime by relying on perturbations from a deterministic template. More precisely, we take the structure of a cycle whose death time equals the birthtime, and then search for slight perturbations that will lead to very short lifetime. These cycles are such that the (critical)  $k$ -face generating the cycle is on the boundary of the  $(k + 1)$ -face killing it. We first define suitable templates of constant size, which in the final argument will then be copy-pasted throughout the entire window  $W_n$ .

To fix a distinguished template, we consider the unit vectors  $P_i := e_{i+1} \in \mathbb{R}^{k+1} \times \{0\}^{d-k-1}$ ,  $0 \leq i \leq k$ , forming a regular  $k$ -simplex of side length  $\sqrt{2}$  whose circumball is centered at  $M^* := (k + 1)^{-1}(P_0 + \dots + P_k)$  and has circumradius  $\rho_k = \sqrt{k/(k + 1)}$ . Furthermore, fix  $Q^* := M^* + \rho_k M^* / |M^*|$ . Figure 4 illustrates this configuration for persistent 1-cycles. We now claim that  $\max_{i \leq k} |P_i - Q^*| \leq \sqrt{2}$ , which will imply that the faces of the simplex  $\{P_0, \dots, P_k, Q^*\}$  are all covered at time  $\rho_k$ , since the latter is the coverage radius of a regular simplex with side length  $\sqrt{2}$ . Now, recalling that  $P_i$  lies on the ball of radius  $\rho_k$  centered at  $M^*$  and that  $P_i - M^*$  is perpendicular to  $M^* - Q^*$  yields

$$|P_i - Q^*|^2 = |P_i - M^*|^2 + |Q^* - M^*|^2 + 2\langle P_i - M^*, M^* - Q^* \rangle = 2\rho_k^2 = 2 - 2/(k + 1) < 2, \tag{8}$$

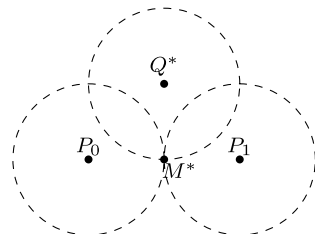
as claimed. Moreover, since the distance between  $P_i$  and  $Q^*$  is strictly smaller than  $\sqrt{2}$ , we deduce that even after perturbing the points of the simplex  $\{P_0, \dots, P_k, Q^*\}$  by a small amount, the  $k$ -face formed by  $P_0, \dots, P_k$  is added after any of the  $k$ -faces involving  $Q^*$ .

Next, we consider random perturbations of this template of a small magnitude  $\delta < 1/2$ . More precisely, we define the event

$$E := \{ \mathcal{P}(B_\delta(P_i)) = 1 \text{ for all } i \leq k, \mathcal{P}(B_{4\delta}(o)) = k + 2 \}$$

that  $\mathcal{P}$  has precisely one point in the  $\delta$ -neighborhood of each of the points in the template in dimension  $k$  and contains precisely one further point inside the ball  $B_{4\delta}(o)$ ,

Fig. 4 Template for persistent 1-cycle of minimal lifetime in dimension



where  $o$  stands for the origin, which we denote by  $Q'$ . Conditioned on  $E$ , we write  $P'_i$  for the Poisson point contained in  $B_\delta(P_i)$ , which is then uniformly distributed at random in  $B_\delta(P_i)$ . We write  $M(P'_0, \dots, P'_k)$  for the center of the  $k$ -dimensional circumball of  $P'_0, \dots, P'_k$ , and  $d' := \inf\{r > 0 : B_r(P'_0) \cap \dots \cap B_r(P'_k) \neq \emptyset\}$  for the circumradius. Then, proceeding similarly as for the definition of  $Q^*$ , we define the set

$$A_\delta(P'_0, \dots, P'_k) := \{B_\delta(M(P'_0, \dots, P'_k) + d'u) : |u| = 1 \text{ and } u \perp \text{span}(P'_1 - P'_0, \dots, P'_k - P'_0)\}. \tag{9}$$

As noted in the paragraph following Eq. (8), if  $Q' \in D_{d', \ell}(P'_0, \dots, P'_k) \cap A_\delta(P'_0, \dots, P'_k)$ , then at the time  $d_{0', \dots, k'}$ , all the  $k$ -faces of the  $(k + 1)$ -simplex  $\{P'_0, \dots, P'_k, Q'\}$  are present. Hence, conditioned on  $E$ , the probability that the  $k$ -cycle formed by the faces of  $\{P'_0, \dots, P'_k, Q'\}$  has lifetime at most  $\ell$  is at least

$$\mathbb{P}(Q' \in D_{d', \ell}(P'_0, \dots, P'_k) \cap A_\delta(P'_0, \dots, P'_k)) = \frac{1}{\kappa_d((4d)^d - d\delta^d)} \mathbb{E}\left[|D_{d', \ell}(P'_0, \dots, P'_k) \cap A_\delta(P'_0, \dots, P'_k)|\right].$$

Then, Lemma 3 in the appendix shows that the volume on the right-hand side is at least  $c\sqrt{\ell}$  for a suitable  $c > 0$ .

Now, for a suitable  $c' > 0$ , we can fit at least  $c'n$  such templates into the window  $W_n$ . Thus, to prove that  $\mathbb{P}(\min_{Z_{i,k,\check{C}} \in W_n} L_{i,k,\check{C}} \geq n^{-2+\epsilon}) \rightarrow 0$ , we take  $\ell = \ell_n^{\min,+} := n^{-2+\epsilon}$  and note that by the independence of the Poisson point process, the probability that none of the shifted templates yields a  $k$ -cycle with lifetime at most  $\ell_n^{\min,+}$  is at most  $(1 - c\sqrt{\ell_n^{\min,+}})^{c'n}$ , which tends to 0, since  $n\sqrt{\ell_n^{\min,+}} \rightarrow \infty$ .

Finally, we move to the scaling of  $u_{n,k,\tau,\text{VR}}^{\min}$ . Henceforth,

$$B_{r,\ell}(x) := B_{r+\ell}(x) \setminus B_r(x)$$

denotes the annulus with center  $x \in \mathbb{R}^d$ , inner radius  $r$  and thickness  $\ell$ .

Scaling of  $u_{n,k,\tau,\text{VR}}^{\min}$  (proof of Theorem 3, scaling).

$u_{n,k,\tau,\text{VR}}^{\min} \geq n^{-1-\epsilon}$ . Writing  $\ell_n^{\min,-} := n^{-1-\epsilon}$ , we assert that

$$\lim_{n \rightarrow \infty} \mathbb{E}[N_{n,k,\text{VR}}^{\min}(\ell_n^{\min,-})] = 0.$$

First, we may assume that  $E_n^O$  occurs. Indeed, by Cauchy-Schwarz,

$$\mathbb{E}[N_{n,k,\text{VR}}^{\min}(\tau_n^{\min,-}) \mathbb{1}\{(E_n^O)^c\}] \leq \mathbb{R}[\mathcal{P}(W_{2n})^d \mathbb{1}\{(E_n^O)^c\}] \leq (\mathbb{E}[\mathcal{P}(W_{2n})^{2d}])^{1/2} \mathbb{P}((E_n^O)^c)^{1/2}.$$

Since  $\mathcal{P}$  is a Poisson point process, the expectation on the right-hand side belongs to  $O(n^{2d})$  so that it remains to invoke Lemma 1.

For the remaining argument, recall that a simplex belongs to the VR-filtration at time  $t$  if the distance between any pair of its vertices is at most  $2t$ . In particular, when a  $k$ -cycle with lifetime  $\ell$  is born at time  $t_b$ , there exist  $X_1, X_2 \in \mathcal{P}$  with  $d_{12} = t_b$ , and when it dies at time  $t_d = t_b + \ell$ , there are Poisson points  $X_3, X_4$  with  $d_{34} = t_b + \ell$ . In other words,  $X_4 \in B_{2d_{12}, 2\ell}(X_3)$ . Moreover, since the event  $E_n^O$  occurs whp, we may assume that  $X_1 \in W_{2n}$  and that  $\max_{i \leq 4} d_{i1} \leq s_n$ .

Here,  $\{X_1, \dots, X_4\}$  consists of at least three different points. First, if they are all pairwise distinct, then the Slivnyak-Mecke formula bounds the expected number of undershoots as

$$\begin{aligned} \int_{W_{2n}} \int_{B_{s_n}(x_1)^2} |B_{2d_{12}, 2\ell_n^{\min,-}}(x_3)| d(x_2, x_3) dx_1 &= \kappa_d^2 s_n^d \int_{W_{2n}} \int_{B_{s_n}(x_1)} ((2d_{12} + 2\ell_n^{\min,-})^d - (2d_{12})^d) dx_2 dx_1 \\ &\leq 4^d d \kappa_d^3 s_n^{3d} \int_{W_{2n}} \ell_n^{\min,-} dx_1 \\ &= 8^d d \kappa_d^3 s_n^{3d} n^{-\varepsilon}, \end{aligned}$$

which tends to 0 as  $n \rightarrow \infty$ .

Second, if say  $X_2 = X_3$ , then we proceed similarly to obtain the bound

$$\begin{aligned} \int_{W_{2n}} \int_{B_{s_n}(x_1)} |B_{2d_{12}, 2\ell_n^{\min,-}}(x_2)| dx_2 dx_1 &= \kappa_d \int_{W_{2n}} \int_{B_{s_n}(x_1)} ((2d_{12} + 2\ell_n^{\min,-})^d - (2d_{12})^d) dx_2 dx_1 \\ &\leq 4^d d \kappa_d^2 s_n^{2d} \int_{W_{2n}} \ell_n^{\min,-} dx_1 \\ &= 8^d d \kappa_d^2 s_n^{2d} n^{-\varepsilon}, \end{aligned}$$

which again tends to 0 as  $n \rightarrow \infty$ .

$u_{n,k,\tau,VR}^{\min} \leq n^{-1+\varepsilon}$ . First, we construct a  $k$ -cycle with a short lifetime for each  $1 \leq k \leq d - 1$ . The  $k$ -cycles depend on a deterministic template illustrated in Fig. 5.

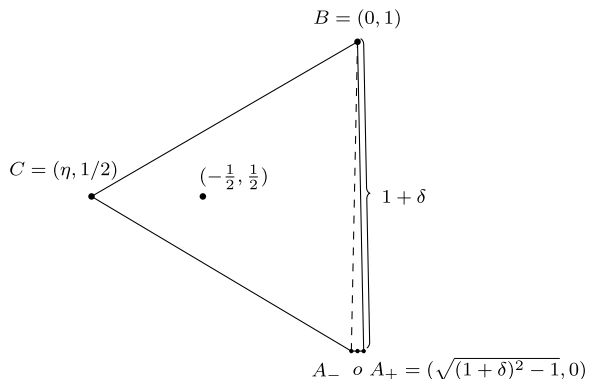
We fix a small value  $\delta = \delta(k, d) > 0$ . Then, we set

$$A_{\pm} := \pm \sqrt{(1 + \delta)^2 - 1} e_1, \quad B := e_2,$$

so that  $|A_- - B| = |A_+ - B| = 1 + \delta$ . We also fix a point  $C := (\eta, 1/2)$ , where  $\eta < 0$  is chosen so that  $|C - A_+| > 1 + \delta > |C - B|$ .

To convey the idea, we sketch how this template gives rise to the desired  $k$ -cycle for  $k = 1$  before moving to higher  $k$ . If we consider the VR-complex on  $\{A_-, A_+, B, C\}$  at level  $1 + \delta$ , then  $A_- A_+ B C$  is a loop. Moreover, removing the edge  $A_- B$  and the attached higher simplices, the complex does not contain any triangles as by construction  $|C - A_+| > 1 + \delta$ . In particular, the loop has positive

**Fig. 5** Template for VR-feature of minimal lifetime



lifetime. However, after adding the edge  $A_-B$ , the loop becomes the boundary of the triangles  $A_-A_+B$  and  $BCA_-$ .

To generalize the construction to higher  $k$ , we introduce additional points  $\{P_{1,\pm}, \dots, P_{k-1,\pm}\}$  to the complex as follows. First, set

$$P_{i,\pm} := \frac{1}{2}(e_2 - e_1) \pm (0.5 + 4\delta)e_{i+2}.$$

In particular, the  $P_{i,\pm}$  are at distance at most  $1 + \delta$  from  $A_-, A_+, B$  and  $C$ . Now, we generalize the above consideration for the VR-complex at level  $|A_+ - B|$  but with the edge  $\{A_-, B\}$  removed. In particular, the complex contains all  $k$ -simplices of the form

$$\{P_{1,\varepsilon_1}, P_{2,\varepsilon_2}, \dots, P_{k-1,\varepsilon_{k-1}}, P, P'\},$$

where  $\varepsilon_i \in \{-, +\}$  and  $\{P, P'\}$  is one of  $\{A_-, A_+\}$ ,  $\{A_+, B\}$ ,  $\{B, C\}$  or  $\{C, A_-\}$ . Then, these simplices form a cycle. Indeed, when removing the vertex  $P_{i,\varepsilon_i}$ , then the resulting face is also contained in the  $k$ -simplex with  $P_{i,\varepsilon_i}$  replaced by  $P_{i,-\varepsilon_i}$ . On the other hand, if for instance  $\{P, P'\} = \{A_-, A_+\}$  and if we remove the point  $A_-$ , then we find the corresponding face also in the simplex with  $\{P, P'\} = \{A_+, B\}$ . However, this complex does not contain any  $(k + 1)$ -simplices. Indeed,  $|P_{i,-} - P_{i,+}| > 1 + \delta$  are at distance larger than  $1 + \delta$ , and for any triple from  $\{A_-, A_+, B, C\}$  at least one edge is also not in the complex. This situation changes drastically if  $A_-B$  belongs to the complex. Then, the boundaries of the  $(k + 1)$ -simplices

$$\{P_{1,\varepsilon_1}, P_{2,\varepsilon_2}, \dots, P_{k-1,\varepsilon_{k-1}}, A_-, A_+, B\} \quad \text{and} \quad \{P_{1,\varepsilon_1}, P_{2,\varepsilon_2}, \dots, P_{k-1,\varepsilon_{k-1}}, B, C, A_-\}$$

yield the cycle constructed before.

It remains to connect this template to random  $k$ -cycles induced by  $\mathcal{P}$ . Similarly as in the  $\check{C}$ -case, set  $\delta_1 = \delta/8$  and consider the event

$$E' := \{\mathcal{P}(B_{\delta_1}(P)) = 1 \text{ for all } P \in \{A_-, A_+, B, C\} \cup \{P_{i,\pm}\}_{i \leq k-1}\} \cap \{\mathcal{P}(B_3(o)) = 2k + 2\}$$

that  $\mathcal{P}$  has precisely one point in the  $\delta_1$ -neighborhood of each of the template points and no other points in the ball  $B_3(o)$ . Conditioned on  $E'$ , we write  $A'_-, A'_+, B', C'$  and  $(P_{i,\pm})'$  for the Poisson points lying in the corresponding neighborhoods of the template points. In particular, under  $E'$ , the primed points are distributed uniformly at random in the corresponding  $\delta_1$ -neighborhoods. Moreover, the construction of the template is sufficiently robust with respect to  $\delta_1$ -perturbations to show that the adjacency findings for the VR-complex continue to hold upon replacing the template points by the perturbed points.

Conditioned on  $E'$ , the probability that this feature has a short lifetime of at most  $\ell_n^{\min,+} := n^{-1+\varepsilon}$  is at least

$$\begin{aligned} & \mathbb{P}(0 < |A'_- - B'| - |A'_+ - B'| < \ell_n^{\min,+} \mid E') \\ &= (\kappa_d \delta_1^d)^{-3} \left| \left\{ (A'_\pm, B') : \max_{P \in \{A_\pm, B\}} |P' - P| \leq \delta_1 \text{ and } 0 < |A'_- - B'| - |A'_+ - B'| < \ell_n^{\min,+} \right\} \right|. \end{aligned}$$

Now, we set  $u := A_+ - B'$  and  $v := A'_- - B'$ , and for  $\alpha \in [0, 1]$  and  $\eta \in \mathbb{R}^d$

$$A_{+, \alpha, \eta} := B' + (|v| - \alpha \ell_n^{\min,+})(u + \eta)/|u + \eta|,$$

so that  $|A_{+, \alpha, \eta} - B'| = |v| - \alpha \ell_n^{\min,+}$ . In particular,  $|A'_- - B'| - |A_{+, \alpha, \eta} - B'| = \alpha \ell_n^{\min,+} \leq \ell_n^{\min,+}$ .

Hence, writing  $\delta_2 := \delta_1/8$ , it suffices to show that

$$|A_{+, \alpha, \eta} - A_+| \leq \delta_1$$

for every  $A'_- \in B_{\delta_2}(A_-)$ ,  $B' \in B_{\delta_2}(o)$ ,  $\alpha \in [0, 1]$  and  $\eta \in B_{\delta_2}(o)$ , because then every such point  $A_{+, \alpha, \eta}$  is an admissible choice for  $A'_+$ . To that end, we leverage the bound

$$\begin{aligned} |A_{+, \alpha, \eta} - A_+| &\leq \left| -u + |v| \frac{u}{|u|} + |v| \left( \frac{u + \eta}{|u + \eta|} - \frac{u}{|u|} \right) \right| + \ell_n^{\min,+} \\ &\leq |u| - |v| + |v| \left| \frac{u + \eta}{|u + \eta|} - \frac{u}{|u|} \right| + \ell_n^{\min,+}. \end{aligned}$$

The first expression is at most  $3\delta_2$  and the second one at most  $4\delta_2$  so that taken together, the right-hand side is indeed at most  $\delta_1$ .

Now, note that

$$|\{A_{+, \alpha, \eta} : \alpha \in [0, 1], \eta \in B_{\delta_2}(o)\}| \geq \ell_n^{\min,+} \kappa_{d-1} \delta_2^{d-1},$$

so that the probability is much larger than  $n^{-1}$ , and consequently the expectation tends to infinity.

### 5 Proof of Proposition 1

In this section, we prove that there are no multiple exceedances in  $W_{s_n^+}$  asymptotically with probability at least  $1 - n^{-\alpha}$  for some  $\alpha > 0$ .

**Proof** (*Proof of Proposition 1 (i)*) By the scaling derived in Sect. 4, it suffices to prove that

$$\mathbb{P}(N_{s_n^+, \check{c}}^{\max}(\ell_n^{\max,-}) \geq 2) \in O(n^{-5/4}).$$

By the scaling in Sect. 4, both cavities  $J_i$  and  $J_j$  die at time at most  $\ell_n^{\max,+}$  whp, so that their birth times are at most  $2\epsilon\gamma_n$ . In particular, for the cavities  $J_i$  and  $J_j$  to be contained in different connected components at the birth time of the later cavity, say  $J_j$ , we have  $d(Z_i, \partial J_i) \geq (1 - \epsilon)\gamma_n$  and  $d(Z_j, \partial J_j) \geq (1 - 3\epsilon)\gamma_n$ . Since  $d(Z_i, Z_j) \geq d(Z_i, \partial J_i) + d(Z_j, \partial J_j)$  at the birth time of  $J_j$ , this implies that  $|v - v'| \geq (2 - 5\epsilon)\gamma_n \geq \gamma_n$ , for  $\epsilon$  small enough, where  $W(v) := v + W_1$ ,  $W(v') := v' + W_1$  are the lattice boxes containing  $Z_i$  and  $Z_j$ , respectively. Therefore,

$$|B_{\ell_n^{\max,-}}(v) \cup B_{\ell_n^{\max,-}}(v')| \geq \kappa_d^{-1} \log(n) |B_{1-\epsilon}(o) \cup B_{1-\epsilon}(e_1)|.$$

Then,

$$\begin{aligned} \mathbb{P}(N_{s_n^+, \check{C}}^{\max}(\ell_n^{\max, -}) \geq 2) &\leq \sum_{\substack{v, v' \in W_{s_n^+} \cap \mathbb{Z}^d \\ |v-v'| \geq \gamma_n}} \mathbb{P}(L_i \wedge L_j \geq \ell_n^{\max, -} \text{ for some } Z_i \in W(v), Z_j \in W(v')) \\ &\leq \sum_{\substack{v, v' \in W_{s_n^+} \cap \mathbb{Z}^d \\ |v-v'| \geq \gamma_n}} \mathbb{P}\left( (B_{\ell_n^{\max, -}}(v) \cup B_{\ell_n^{\max, -}}(v')) \cap \mathcal{P} = \emptyset \right) \\ &\leq |W_{s_n^+}^2| \exp\left(-\kappa_d^{-1} \log(n) |B_{1-\varepsilon}(o) \cup B_{1-\varepsilon}(e_1)|\right). \end{aligned}$$

The last term is in  $O(n^{-5/4})$  provided that  $\varepsilon > 0$  is sufficiently small.

For the remainder of this section, we fix  $\varepsilon := 1/64$ . Next, we move to the minimum in the  $\check{C}$ -filtration.

**Proof** (*Proof of Proposition 1 (ii);  $\check{C}$* ) Again, it suffices to show that

$$\mathbb{P}(N_{s_n^+, \check{C}}^{\min}(\ell_n^{\min, +}) \geq 2) \in O(n^{-9/8}), \tag{10}$$

where we now set  $\ell_n^{\min, +} := n^{-2+\varepsilon}$ .

We need to understand the geometric implications of finding 1-cycles  $Z_i, Z_j \in W_{s_n^+}$  with life times shorter than  $\ell_n^{\min, +}$ . We recall from Sect. 4 that one such undershoot gives Poisson points  $\{X_0, \dots, X_4\}$  such that the triangle  $X_2X_3X_4$  is covered by the union of balls centered at  $X_i, 2 \leq i \leq 4$ . First, we observe that if at least one of  $X_0, X_1$  is not contained in  $\{X_2, X_3, X_4\}$ , then Eq. (10) holds even without taking the second 1-cycle  $Z_j$  into account. Indeed, if for instance  $X_1 = X_2$ , but  $X_0 \notin \{X_2, X_3, X_4\}$ , then the Slivnyak-Mecke formula yields the bound

$$\int_{W_{s_n^+}^3} |B_{2d_{234}-2\ell_n^{\min, +}, 2\ell_n^{\min, +}}(x_2)| d(x_2, x_3, x_4) \in O(s_n^{4d} \ell_n^{\min, +}).$$

We may hence assume to have points  $\{X_0, X_1, X_2\}$  with  $d_{01} < d_{012} \leq d_{01} + \ell_n^{\min, +}$  for the first 1-cycle and similarly points  $\{X'_0, X'_1, X'_2\} \neq \{X_0, X_1, X_2\}$  with  $d_{0'1'} < d_{0'1'2'} \leq d_{0'1'} + \ell_n^{\min, +}$ . Again, several configurations are possible, where the most challenging ones correspond to the cases where  $\{X_0, X_1, X_2\}$  and  $\{X'_0, X'_1, X'_2\}$  differ in one variable.

First, assume that  $X'_0 = X_0, X'_1 = X_1$ , but  $X'_2 \neq X_2$ . Then, invoking Lemma 3 in Appendix 2 gives the volume bound

$$\begin{aligned} \left| \{(x_0, x_1, x_2, x'_2) \in W_{s_n^+}^4 : d_{012}, d_{012'} \in (d_{01}, d_{01} + \ell_n^{\min, +})\} \right| &\leq \int_{W_{s_n^+}^2} |D_{d_{01}, \ell_n^{\min, +}}(x_0, x_1)|^2 d(x_0, x_1) \\ &\leq c_{\text{cresc}}^2 s_n^{4d} \ell_n^{\min, +}. \end{aligned}$$



Since  $\varepsilon = 1/64$ , the last line belongs to  $O(n^{-9/8})$ . This concludes the proof of the first case. Second, assuming that  $X'_1 = X_1, X'_2 = X_2$ , but  $X'_0 \neq X_0$ , Lemmas 3 and 5 in Appendix 2 give that

$$\begin{aligned} & \left| \{(x_0, x_1, x_2, x'_0) \in W_{s_n^+}^4 : d_{01} < d_{012} \leq d_{01} + \ell_n^{\min,+} \text{ and } d_{0'1} < d_{0'12} \leq d_{0'1} + \ell_n^{\min,+}\} \right| \\ & \leq \int_{W_{s_n^+}^2} \int_{D_{d_{01}, \ell_n^{\min,+}}(x_0, x_1)} \left| \{x'_0 \in W_{s_n^+} : d_{0'1} < d_{0'12} \leq d_{0'1} + \ell_n^{\min,+}\} \right| dx_2 d(x_0, x_1) \\ & \leq cn^{-1/4} \int_{W_{s_n^+}^2} |D_{d_{01}, \ell_n^{\min,+}}(x_0, x_1)| d(x_0, x_1) \\ & \leq cc' s_{n,+}^3 n^{-1/4} \sqrt{\ell_n^{\min,+}}, \end{aligned}$$

which is again in  $O(n^{-9/8})$ .

Finally consider the case where  $X'_1 = X_2, X'_2 = X_1$ , but  $X'_0 \neq X_0$ . Then, similarly, Lemmas 3 and 5 in Appendix 2 give that

$$\begin{aligned} & \left| \{(x_0, x_1, x_2, x'_0) \in W_{s_n^+}^4 : d_{01} < d_{012} \leq d_{01} + \ell_n^{\min,+} \text{ and } d_{0'2} < d_{0'21} \leq d_{0'2} + \ell_n^{\min,+}\} \right| \\ & \leq \int_{W_{s_n^+}^2} \int_{D_{d_{01}, \ell_n^{\min,+}}(x_0, x_1)} \left| \{x'_0 \in W_{s_n^+} : d_{0'2} < d_{0'21} \leq d_{0'2} + \ell_n^{\min,+}\} \right| dx_2 d(x_0, x_1), \end{aligned}$$

so that we now conclude as in the previous case.

Finally, we deal with the minimum in the VR-filtration.

**Proof** (*Proof of Proposition 1 (ii); VR*) Again, it suffices to prove that

$$\mathbb{P}(N_{s_n^+, k, VR}^{\min}(\ell_n^{\min,+}) \geq 2) \in O(n^{-33/32}), \tag{11}$$

where we now set  $\ell_n^{\min,+} = n^{-1+\varepsilon}$ . Now, suppose two  $k$ -cycles centered at  $Z_i, Z_j \in W_{s_n^+}$  live shorter than  $\ell_n^{\min,+}$ . By the VR-filtration, this means that for the  $k$ -cycle centered in  $Z_i$ , there exist Poisson points  $X_0, \dots, X_3 \in W_{s_n^+}$  such that

$$d_{01} < d_{23} < d_{01} + \ell_n^{\min,-}.$$

Similarly also the  $k$ -cycle centered in  $Z_j$  gives rise to Poisson points  $X'_0, \dots, X'_3 \in W_{s_n^+}$  such that

$$d_{0'1'} < d_{2'3'} < d_{0'1'} + \ell_n^{\min,-}.$$

Again, not all points need to be distinct, but both  $\{X_0, \dots, X_3\}$  and  $\{X'_0, \dots, X'_3\}$  consist of at least 3 elements. Moreover,  $\{X_0, X_1\} \neq \{X'_0, X'_1\}$  and  $\{X_2, X_3\} \neq \{X'_2, X'_3\}$  since the  $k$ -cycles are distinct.

We now distinguish two cases. The first one being that  $\{X_0, \dots, X_3\} \neq \{X'_0, \dots, X'_3\}$ . Say, for instance  $X'_3 \notin \{X_0, \dots, X_3\}$ . Then, we may apply the Slivnyak-Mecke formula to see that the expected number of configurations is at most

$$\begin{aligned} & \int_{W_{s_n^+}} \mathbb{E} \left[ \#\{X_0, \dots, X_3, X'_0, \dots, X'_2 \in \mathcal{P} \cap W_{s_n^+} : d_{01} < d_{23} < d_{01} + \ell_n^{\min,-} \text{ and } x_3 \in B_{2d_{01}, 2\ell_n^{\min,-}}(X_2)\} \right] dx'_3 \\ & \leq 2^{d+1} ds_{n,+}^{4d} \ell_n^{\min,-} \mathbb{E} \left[ \#\{X_0, \dots, X_3 \in \mathcal{P} \cap W_{s_n^+} : d_{01} < d_{23} < d_{01} + \ell_n^{\min,-}\} \right] \\ & \leq 2c(2^d ds_{n,+}^{4d})^2 (\ell_n^{\min,-})^2 \in O(n^{-33/32}), \end{aligned}$$

for a suitable  $c > 0$ .

It remains to deal with the case  $\{X'_0, \dots, X'_3\} = \{X_0, \dots, X_3\}$ , which necessarily means that they consist of 4 elements. For instance, it may occur that  $X'_0 = X_0$ ,  $X'_1 = X_3$ ,  $X'_2 = X_2$  and  $X'_3 = X_1$ . Then,  $d_{01} < d_{23} < d_{01} + \ell_n^{\min,+}$  and  $d_{03} < d_{12} < d_{03} + \ell_n^{\min,+}$ . Applying the Slivnyak-Mecke formula, the expected number of such configurations in  $W_{s_n^+}$  is at most

$$\left| \{(x_0, \dots, x_3) \in W_{s_n^+}^4 : |d_{23} - d_{01}| \vee |d_{03} - d_{12}| \leq \ell_n^{\min,+}\} \right|.$$

Lemma 4 in Appendix 2 shows that the latter expression is in  $O(n^{-33/32})$ , as asserted.

### Well-definedness of $u_{n,\tau}^{\max}$ , $u_{n,k,\tau,\check{c}}^{\min}$ and $u_{n,k,\tau,\text{VR}}^{\min}$

In this section, we show that the thresholds defined in Eqs. (1), (2) and (3) are well-defined for every  $\tau > 0$  provided that  $n$  is sufficiently large. To that end, introduce

$$F_{k,\check{c}}(\ell) := \mathbb{E}[\#\{Z_{i,k,\check{c}} \in W_1 : L_{i,k,\check{c}} \leq \ell\}], \tag{12}$$

and similarly  $F_{k,\text{VR}}$ .

**Lemma 2 (Well-definedness of  $u_{n,\tau}^{\max}$ ,  $u_{n,k,\tau,\check{c}}^{\min}$  and  $u_{n,k,\tau,\text{VR}}^{\min}$ )** *Let  $1 \leq k \leq d - 1$ . Then,  $F_{k,\check{c}}, F_{k,\text{VR}} : (0, \infty) \rightarrow (0, \infty)$  are continuous and strictly increasing with  $\lim_{\ell \rightarrow 0} F_{k,\check{c}}(\ell) = \lim_{\ell \rightarrow 0} F_{k,\text{VR}}(\ell) = 0$ .*

**Proof**  $\lim_{\ell \rightarrow 0} F_{k,\check{c}/\text{VR}}(\ell) = 0$ . First, by right-continuity  $\lim_{\ell \rightarrow 0} F_{k,\check{c}/\text{VR}}(\ell) = F_{k,\check{c}/\text{VR}}(0)$ , where  $F_{k,\check{c}/\text{VR}}(0)$  describes the expected number of persistent  $k$ -cycles with lifetime 0 centered in  $W_1$ . However, by definition, the lifetime of a persistent  $k$ -cycle is always strictly positive, so that  $F_{k,\check{c}/\text{VR}}(0) = 0$ .

**Continuity.** Let  $\ell > 0$  be arbitrary. To show continuity, we establish that there are no persistent  $k$ -cycles with lifetime exactly  $\ell$ . We start by proving the claim for VR-filtration. In that case, writing  $d_{ij} = |X_i - X_j|/2$ , there would exist points  $X_0, X_1, X_2, X_3 \in \mathcal{P}$  such that  $d_{23} = d_{01} + \ell$  with  $X_3 \notin \{X_0, X_1, X_2\}$ . Then,

$$\begin{aligned} & \mathbb{E} \left[ \#\{X_3 : X_3 \in \partial B_{2d_{01}+2\ell}(X_2) \text{ for some } X_0, X_1, X_2 \in \mathcal{P}\} \right] \\ &= \int_{\mathbb{R}^d} \mathbb{P} \left( x \in \partial B_{2d_{01}+2\ell}(X_2) \text{ for some } X_0, X_1, X_2 \in \mathcal{P} \right) dx \\ &= \mathbb{E} \left[ \left| \bigcup_{X_0, X_1, X_2 \in \mathcal{P}} \partial B_{2d_{01}+2\ell}(X_2) \right| \right] \\ &= 0. \end{aligned}$$

For the  $\check{C}$ -filtration, the argumentation is similar. Indeed, we recall from Sect. 4 that a persistent  $k$ -cycle gives Poisson points  $X_0, \dots, X_k \in \mathcal{P}$  and  $X'_0, \dots, X'_{k+1} \in \mathcal{P}$  with  $X'_{k+1} \notin \{X'_0, \dots, X'_k\}$  such that relying on the filtration times from Eq. (7), we have  $d_{0' \dots (k+1)'} = d_{0 \dots k} + \ell$ . Hence, similarly as in the VR-filtration,

$$\begin{aligned} & \mathbb{E} \left[ \#\{X'_{k+1} : d_{0' \dots (k+1)'} = d_{0 \dots k} + \ell \text{ for some } X_0, \dots, X_k, X'_0, \dots, X'_k \in \mathcal{P}\} \right] \\ &= \mathbb{E} \left[ \left| \bigcup_{\substack{X_0, \dots, X_k \in \mathcal{P} \\ X'_0, \dots, X'_k \in \mathcal{P}}} \partial D_{d_{0 \dots k}}(X'_0, \dots, X'_k) \right| \right] \\ &= 0. \end{aligned}$$

**Strict monotonicity.** Strict monotonicity means that for  $\ell < \ell'$  there is a positive probability to observe a persistent  $k$ -cycle with lifetime in  $(\ell, \ell')$ . Note that scaling all points by a factor also scales the lifetime by that factor. Hence, it suffices to show that for some fixed  $b$  and every  $\varepsilon > 0$ , there is a positive probability to have a persistent  $k$ -cycle with lifetime in  $(b - \varepsilon, b + \varepsilon)$ .

We start with the VR-filtration. Consider a persistent  $k$ -cycle described by a cross-polytope  $\{\pm e_i\}_{1 \leq i \leq k+1}$ . Then, this persistent  $k$ -cycle has birth time  $1/\sqrt{2}$  and death time 1. Therefore its lifetime is  $b := 1 - 1/\sqrt{2}$ . If we allow the vertices of the persistent  $k$ -cycle to be perturbed by at most  $\varepsilon/2$ , the lifetime is in the interval  $(b - \varepsilon, b + \varepsilon)$ .

Finally, we deal with the  $\check{C}$ -filtration. Consider a persistent  $k$ -cycle described by a regular simplex  $\{e_i\}_{1 \leq i \leq k+1}$ . Then, this persistent  $k$ -cycle has birth time  $\rho := \sqrt{(k-1)/k}$  and death time  $\rho' := \sqrt{k/(k+1)}$ . Therefore, its lifetime is  $b := \rho' - \rho$ , and we conclude as in the VR-filtration.

### Volume computations

In this section, we compute volume bounds for the specific configurations occurring in the proofs in Sects. 4 and 5. All computations rely only on findings from elementary geometry, but are still bit tedious when written out in detail.

First, we bound the volumes of crescents. For any affinely independent points  $x_0, \dots, x_k \in \mathbb{R}^d$ , recall the definition of the set  $A_\delta(x_0, \dots, x_k)$  introduced in Eq.

(9). As in Sects. 4 and 5, we write  $d_{i_0 \dots i_\ell}$  for the filtration time when the simplex  $\{x_{i_0}, \dots, x_{i_\ell}\}$  appears in the  $\check{C}$ -filtration.

**Lemma 3 (Volume of crescents)** *There exists  $c_{\text{cresc}} = c_{\text{cresc}}(d) > 0$  with the following properties. Let  $r > 0$ ,  $k \leq d - 1$ ,  $\ell < 1$ ,  $x_0, \dots, x_k \in \mathbb{R}^d$  be affinely independent. Then,*

$$|D_{r,\ell}(x_0, \dots, x_k)| \leq c_{\text{cresc}}(r + \ell)^k \sqrt{\ell}.$$

Moreover, if  $d_{0 \dots k} \geq 1/2$  and  $\delta \in (\ell, 1)$ , then

$$|D_{d_{0 \dots k}, \ell}(x_0, \dots, x_k) \cap A_\delta(x_0, \dots, x_k)| \geq c_{\text{cresc}}^{-1} \delta^k \sqrt{\ell}.$$

**Proof** To ease notation, set  $a := d_{0 \dots k}$ . Since both  $D_{d_{0 \dots k}, \ell}(x_0, \dots, x_k)$  and  $A_\delta(x_0, \dots, x_k)$  are defined through their sections with  $(k + 1)$ -dimensional planes containing  $x_0, \dots, x_k$ , we may assume that  $k = d - 1$ . Furthermore, by rotating and shifting we may assume that  $x_0, \dots, x_{d-1}$  are contained in  $\{0\} \times \mathbb{R}^{d-1}$  and that their circumcenter is the origin. Finally, the set  $D_{r,\ell}(x_0, \dots, x_{d-1})$  is rotationally symmetric around the axis  $\mathbb{R}e_1$ .

**Upper bound.** Defining  $r_+(b), r_-(b) > 0$  for any  $b \in \mathbb{R}$  through

$$D_{r,\ell}(-ae_2, ae_2) \cap (\{b\} \times \mathbb{R}) = \{b\} \times ([r_-(b), r_+(b)] \cup [-r_+(b), -r_-(b)]),$$

we obtain by Fubini that

$$|D_{r,\ell}(x_0, \dots, x_{d-1})| = \int_0^{2r+2\ell} |B_{r_+(b)}(o) \setminus B_{r_-(b)}(o)| db = \kappa_{d-1} \int_0^{2r+2\ell} (r_+(b)^{d-1} - r_-(b)^{d-1}) db.$$

Now,  $r_+(b)^{d-1} - r_-(b)^{d-1}$  is at most  $d(r + \ell)^{d-2}(r_+(b) - r_-(b))$ , so that

$$\begin{aligned} |D_{r,\ell}(x_0, \dots, x_{d-1})| &\leq d\kappa_{d-1}(r + \ell)^{d-2} \int_0^{2r+2\ell} (r_+(b) - r_-(b)) db \\ &= \frac{d}{2} \kappa_{d-1}(r + \ell)^{d-2} |D_{r,\ell}(-ae_2, ae_2)|. \end{aligned}$$

Hence, we have now reduced the upper bound to the special setting where  $d = 2$ . Here, an elementary geometric argument that is elucidated in (Biscio et al. 2020, [Lemma 9.8]) concludes the proof.

**Lower bound.** For a point  $x \in \mathbb{R}^{d-1}$  with  $|x| \leq \delta$ , we let

$$I(x) := \{b \geq 0 : (b, x) \in D_{a,\ell}(x_0, \dots, x_{d-1})\}$$

denote the interval consisting of all points inside  $D_{a,\ell}$  projecting onto  $x$ . Then, it suffices to show that there exists a constant  $c > 0$  such that for each such  $x$  we have  $|I(x)| \geq c\sqrt{\ell}$ . Again, after rotation, we may reduce to the two-dimensional setting, i.e., assume that  $x = (0, x, 0, \dots, 0)$  for some  $x \leq \delta$ .

To derive the lower bound on  $|I(x)|$ , note that the first coordinate of one of the midpoint of the circle through  $\pm ae_2$  with radius  $a + \ell$  equals

$$b_0 := \sqrt{(a + \ell)^2 - a^2} = \sqrt{\ell} \sqrt{2a + \ell}.$$

Thus, as illustrated in Fig. 6,

$$|I(x)| = \sqrt{(a + \ell)^2 - x^2} + b_0 - \sqrt{a^2 - x^2} = b_0 + \ell \frac{2a + \ell}{\sqrt{(a + \ell)^2 - x^2} + \sqrt{a^2 - x^2}},$$

which is bounded below by a scalar multiple of  $\sqrt{\ell}$  since  $b_0 \geq \sqrt{2a\ell} \geq \sqrt{\ell}$ . □

Next, we bound the volume of almost parallelograms, recalling that  $s_n^+ = 3^d s_n^d$ .

**Lemma 4 (Almost parallelogram)** *It holds that*

$$\left| \{ (x_0, \dots, x_3) \in W_{s_n^+}^4 : |d_{23} - d_{01}| \vee |d_{03} - d_{12}| \leq n^{-63/64} \} \right| \in O(n^{-33/32}).$$

**Proof** First, by rotating  $x_3$  into the plane spanned by  $x_0, x_1$  and  $x_2$ , we may reduce to the two-dimensional setting. Second, we may assume that  $\min_{i \neq j} d_{ij} \geq n^{-1/16}$ . Indeed, for instance if  $i = 0$  and  $j = 1$ , then

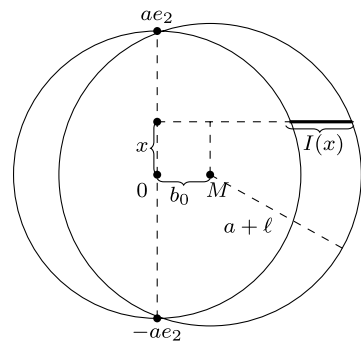
$$\left| \{ (x_0, \dots, x_3) \in W_{s_n^+}^4 : d_{01} \leq n^{-1/16} \text{ and } |d_{23} - d_{01}| \leq \ell_n \} \right| \in O(n^{-33/32}),$$

where we set  $\ell_n := n^{-63/64}$ . Similarly, we may also assume that the angles  $\angle x_i x_j x_k$  are at least  $n^{-1/16}$  for all pairwise distinct  $i, j, k$ . Note that we did not attempt to optimize the exponent  $-1/16$ .

After these simplifications, it remains to bound the annuli-intersection area  $|B_{d_{12}, \ell_n}(x_0) \cap B_{d_{01}, \ell_n}(x_2)|$ , as illustrated in Fig. 7. To bound this quantity, we rely on the co-area formula from (Morgan 2016, Chapter 3). More precisely, we have

$$|B_{d_{12}, \ell_n}(x_0) \cap B_{d_{01}, \ell_n}(x_2)| = \int_{d_{01}}^{d_{01} + \ell_n} |\partial B_r(x_0) \cap B_{d_{01}, \ell_n}(x_2)| dr.$$

**Fig. 6** Lower bound on crescent volume



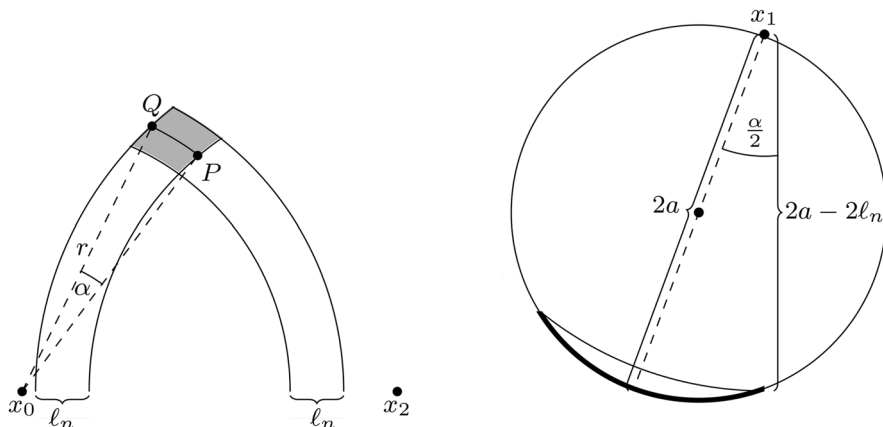


Fig. 7 Intersection area of annuli (left); Arc describing those  $x_0$  with  $d_{01} \geq d_{012} - \ell_n$  (right)

Hence, if we write  $P, Q$  for the endpoints of one of the two arcs  $\partial B_r(x_0) \cap B_{d_{01}, \ell_n}(x_2)$  and  $\alpha$  for the enclosed angle, it suffices to show that  $\alpha$  is of order  $O(n^{-1/2})$ .

To that end, note that

$$\cos(\angle x_2 x_0 P) = \frac{d_{02}^2 + r^2 - d_{01}^2}{2d_{02}r} \quad \text{and} \quad \cos(\angle x_2 x_0 Q) = \frac{d_{02}^2 + r^2 - (d_{01} + \ell_n)^2}{2d_{02}r}.$$

Since  $\alpha = \angle x_2 x_0 Q - \angle x_2 x_0 P$ , the mean-value theorem yields some  $\alpha' \in [\angle x_2 x_0 P, \angle x_2 x_0 Q]$  with

$$\alpha = \frac{\cos(\angle x_2 x_0 Q) - \cos(\angle x_2 x_0 P)}{\sin(\alpha')} = \ell_n \frac{2d_{01} + \ell_n}{2d_{02}r \sin(\alpha')}.$$

To finish the proof, note that the denominator is of order at least  $n^{-1/4}$  by the assumptions at the beginning of proof. □

We conclude the appendix with a final elementary geometric volume bound.

**Lemma 5** *It holds that*

$$|\{(x_0, x_1, x_2) \in W_{s_n^3} : d_{01} < d_{012} \leq d_{01} + n^{-127/64}\}| \in O(n^{-1/4}).$$

**Proof** To ease notation, we put  $\ell_n := n^{-127/64}$ . Arguing similarly as in the proof of Lemma 4, we may leverage rotational symmetry around the axis formed by  $x_1 x_2$  to reduce the proof to the setting  $d = 2$ . Similarly, we may assume that  $d_{12} \geq 2\ell_n$ .

Then, for fixed  $a > d_{12}$ , as observed in the proof of Lemma 3, the location of all  $x_0$  with  $d_{012} = a$  is the union of the two circles with radius  $a$  passing through  $x_1, x_2$ . In particular, the location of  $x_0$  with  $d_{01} \geq a - \ell_n$  is given by an arc in each of these circles, see Fig. 7 (right). Hence, applying the co-area formula (Morgan 2016, Chapter 3) to the level sets of the function  $u(x_0) := d_{012}$ , we obtain that

$$\begin{aligned}
 & \left| \{(x_0, x_1, x_2) \in W_{s_n^+}^3 : d_{01} \geq d_{012} - \ell_n\} \right| \\
 &= \int_{\ell_n}^{\infty} \left| \{(x_0, x_1, x_2) \in W_{s_n^+}^3 : d_{01} \geq a - \ell_n\} \cap \{d_{012} \leq a\} \right| |\nabla u(x_0)|^{-1} da.
 \end{aligned}$$

Since the gradient of  $u$  is bounded away from 0, it suffices to show that the length of these arcs is of order  $O(\sqrt{a\ell_n})$ . By construction, the angle  $\alpha$  associated with one of these arcs satisfies  $\cos(\alpha/2) = (2a - 2\ell_n)/2a$ , so that

$$\sin^2(\alpha/2) = 1 - \cos^2(\alpha/2) \leq \frac{\ell_n}{a},$$

Since  $\ell_n/a \leq 1/2$ , we deduce the asserted  $a\alpha \leq a\sqrt{\frac{8\ell_n}{a}} = \sqrt{8a\ell_n}$ . □

**Acknowledgements** We thank two anonymous referees for reading through the entire manuscript thoroughly and for providing us with high-quality and constructive feedback. In particular, one of the referees suggested the use of critical simplices. Their suggestions improved the presentation in many places.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Arratia, R., Goldstein, L., Gordon, L.: Poisson approximation and the Chen-Stein method. *Statist. Sci.* **5**(4), 403–434 (1990)

Biscio, C.A.N., Chenavier, N., Hirsch, C., Svane, A.M.: Testing goodness of fit for point processes via topological data analysis. *Electron. J. Stat.* **14**(1), 1024–1074 (2020)

Björner, A.: Topological methods. In: *Handbook of Combinatorics*, pp. 1819–1872. Elsevier, Amsterdam (1995)

Bobrowski, O.: Homological connectivity in random Čech complexes. arXiv preprint [arXiv:1906.04861](https://arxiv.org/abs/1906.04861) (2019)

Bobrowski, O., Kahle, M., Skraba, P.: Maximally persistent cycles in random geometric complexes. *Ann. Appl. Probab.* **27**(4), 2032–2060 (2017)

Bobrowski, O., Skraba, P.: Homological Percolation: The Formation of Giant k-Cycles. *Int. Math. Res. Not* (2020)

Boissonnat, J.D., Chazal, F., Yvinec, M.: *Geometric and Topological Inference*. Cambridge University Press, Cambridge (2018)

Calka, P., Chenavier, N.: Extreme values for characteristic radii of a Poisson-Voronoi tessellation. *Extremes* **17**(3), 359–385 (2014)

Chenavier, N.: A general study of extremes of stationary tessellations with examples. *Stochastic Process. Appl.* **124**(9), 2917–2953 (2014)

- Duminil-Copin, H., Raoufi, A., Tassion, V.: Subcritical phase of  $d$ -dimensional Poisson-Boolean percolation and its vacant set. *Annales Henri Lebesgue* **3**, 677–700 (2020)
- Ghrist, R.: Barcodes: the persistent topology of data. *Bull. Am. Math. Soc.* **45**(1), 61–75 (2008)
- Gidea, M., Katz, Y.A.: Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications* **491**, 820–834 (2018)
- Hall, P.: On the coverage of  $k$ -dimensional space by  $k$ -dimensional spheres. *Ann. Probab.* **13**(3), 991–1002 (1985)
- Kahle, M.: Random geometric complexes. *Discrete Comput. Geom.* **45**(3), 553–573 (2011)
- Last, G., Penrose, M.D.: *Lectures on the Poisson Process*. Cambridge University Press, Cambridge (2016)
- Leadbetter, M.R., Lindgren, G., Rootzén, H.: *Extremes and related properties of random sequences and processes*. Springer-Verlag, New York (1983)
- Meester, R., Roy, R.: *Continuum Percolation*. Cambridge University Press, Cambridge (1996)
- Morgan, F.: *Geometric Measure Theory: A Beginner's Guide*. Elsevier, Amsterdam (2016)
- Owada, T., Adler, R.J.: Limit theorems for point processes under geometric constraints (and topological crackle). *Ann. Probab.* **45**(3), 2004–2055 (2017)
- Penrose, M.D.: The longest edge of the random minimal spanning tree. *Ann. Appl. Probab.* **7**(2), 340–361 (1997)
- Penrose, M.D.: *Random Geometric Graphs*. Oxford University Press, Oxford (2003)
- Pranav, P., Edelsbrunner, H., van de Weygaert, R., Vegter, G., Kerber, M., Jones, B.J.T., Wintraecken, M.: The topology of the cosmic web in terms of persistent Betti numbers. *Monthly Notices of the Royal Astronomical Society* **465**(4), 4281–4310 (2016)
- Rocks, J.W., Liu, A.J., Katifori, E.: Revealing structure-function relationships in functional flow networks via persistent homology. *Phys. Rev. Research* **2**, 033234 (2020)
- Saadatfar, M., Takeuchi, H., Robins, V., Francois, N., Hiraoka, Y.: Pore configuration landscape of granular crystallization. *Nature Communications* **8**(1), 1–11 (2017)
- Schulte, M., Thäle, C.: The scaling limit of Poisson-driven order statistics with applications in geometric probability. *Stochastic Process. Appl.* **122**(12), 4096–4120 (2012)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.