Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 3, 1869–1897
https://doi.org/10.1214/19-AIHP1020

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Thresholds for vanishing of 'Isolated' faces in random Čech and Vietoris–Rips complexes

Srikanth K. Iyer[a,1] and D. Yogeshwaran[b,2]

[a]*Department of Mathematics, Indian Institute of Science, Bangalore, India. E-mail: skiyer@iisc.ac.in*
[b]*Theoretical Statistics and Mathematics unit, Indian Statistical Institute, Bangalore, India. E-mail: d.yogesh@isibang.ac.in*

Received 5 March 2018; revised 23 August 2019; accepted 25 August 2019

Abstract. We study combinatorial connectivity for two models of random geometric complexes. These two models – Čech and Vietoris–Rips complexes – are built on a homogeneous Poisson point process of intensity n on a d-dimensional torus, $d > 1$, using balls of radius r_n. In the former, the k-simplices/faces are formed by subsets of $(k+1)$ Poisson points such that the balls of radius r_n centred at these points have a mutual interesection and in the latter, we require only a pairwise intersection of the balls. Given a (simplicial) complex (i.e., a collection of k-simplices for all $k \geq 1$), we can connect k-simplices via $(k+1)$-simplices ('up-connectivity') or via $(k-1)$-simplices ('down-connectivity). Our interest is to understand these two combinatorial notions of connectivity for the random Čech and Vietoris–Rips complexes asymptotically as $n \to \infty$. In particular, we analyse in detail the threshold radius for vanishing of isolated k-faces for up and down connectivity of both types of random geometric complexes. Though it is expected that the threshold radius $r_n = \Theta((\frac{\log n}{n})^{1/d})$ in coarse scale, our results give tighter bounds on the constants in the logarithmic scale as well as shed light on the possible second-order correction factors. Further, they also reveal interesting differences between the phase transition in the Čech and Vietoris–Rips cases. The analysis is interesting due to non-monotonicity of the number of isolated k-faces (as a function of the radius) and leads one to consider 'monotonic' vanishing of isolated k-faces. The latter coincides with the vanishing threshold mentioned above at a coarse scale (i.e., $\log n$ scale) but differs in the $\log\log n$ scale for the Čech complex with $k=1$ in the up-connected case. For the case of up-connectivity in the Vietoris–Rips complex and for r_n in the critical window, we also show a Poisson convergence for the number of isolated k-faces when $k \leq d$.

Résumé. Nous étudions la connectivité combinatoire pour deux modèles de complexes géométriques aléatoires. Ces deux modèles – les complexes de Čech et de Vietoris–Rips – sont construits sur la base d'un processus de Poisson homogène d'intensité n sur un tore de dimension d, $d > 1$, en utilisant des boules de rayon r_n. Dans le premier, les k-simplexes/faces sont formés par les sous-ensembles de $k+1$ points du processus de Poisson tels que l'intersection des boules de rayon r_n centrées en ces points est non vide, et dans le second, nous demandons seulement que les intersections deux-à-deux des boules soient non vides. Étant donné un complexe simplicial (c'est-à-dire une collection de k-simplexes pour tous $k \geq 1$), nous pouvons connecter les k-simplexes via les $(k+1)$-simplexes (connectivité par le haut) ou via les $(k-1)$-simplexes (connectivité par le bas).

Notre objectif est de comprendre ces deux notions combinatoires de connectivité pour les complexes de Čech et Vietoris–Rips asymptotiquement lorsque $n \to \infty$.

En particulier, nous analysons en détail le rayon critique pour la disparition des k-faces isolées pour la connectivité par le haut et par le bas dans les deux types de complexes géométriques aléatoires. Bien qu'il soit attendu que le rayon critique soit $r_n = \Theta((\frac{\log n}{n})^{1/d})$ dans une échelle grossière, nos résultats donnent des bornes plus fines sur les constantes dans l'échelle logarithmique et suggère les possibles facteurs correctifs de second ordre. De plus, ils révèlent aussi des différences intéressantes entre les transitions de phase entre les cas de Čech et Vietoris–Rips.

L'analyse est intéressante du fait de la non monotonie du nombre de k-faces isolées (comme fonction du rayon) ce qui conduit à considérer une version monotone de la disparition des k-faces. Cette dernière coïncide avec le seuil de disparition mentionné précédemment à une échelle grossière (c'est-à-dire à une échelle $\log n$) mais diffère à l'échelle $\log\log n$ pour le complexe de Čech avec $k=1$ pour la connectivité par le haut.

Dans le cas de la connectivité par le haut dans le cas du complexe de Vietoris–Rips et pour r_n dans la fenêtre critique, nous montrons aussi une convergence vers un processus de Poisson pour le nombre de k-faces isolées quand $k \leq d$.

[1]SKI's research was supported in part by SERB Matrix grant and UGC-CAS.
[2]DY's research was supported in part by DST-INSPIRE Faculty fellowship and CPDA from the Indian Statistical Institute.

MSC: Primary 60D05; 05E45; secondary 60B99; 05C80

Keywords: Random geometric complexes; Random hypergraphs; Connectivity; Maximal faces; Phase transition; Poisson convergence

1. Introduction

Let $\mathcal{P}_n = \{X_1, \dots, X_{N_n}\}$ be a collection of points in the d-dimensional torus U with $\{X_i\}_{i\geq 1}$ being a sequence of i.i.d. uniform random variables in U and N_n, an independent Poisson random variable with mean n. In other words, $\mathcal{P}_n$ is a homogeneous Poisson point process with intensity n on U. A classical model of random graph $G(\mathcal{P}_n, r)$ (for $r \in (0, \infty)$) introduced by Gilbert in 1961 [22] called the random geometric graph or Gilbert graph is as follows: The vertex set is $\mathcal{P}_n$ and X_i, X_j share an edge if $|X_i - X_j| \leq 2r$. Though Gilbert introduced it on the plane, we shall study it on the torus U to avoid boundary effects. This is a common simplification especially when studying sharp thresholds for connectivity properties. A seminal result in the subject was the determining of exact connectivity threshold [1,47,48]. The precise statement of the sharp phase transition result [48, Theorem 13.10] is that for any sequence $w(n) \to \infty$, the following holds:

$$\mathbb{P}\big\{G(\mathcal{P}_n, r_n) \text{ is connected}\big\} \to \begin{cases} 0 & \text{if } n\theta_d 2^d r_n^d = \log n - w(n) \\ 1 & \text{if } n\theta_d 2^d r_n^d = \log n + w(n), \end{cases} \tag{1.1}$$

where θ_d denotes the volume of the unit ball in $\mathbb{R}^d$. An important step towards the proof of the above result was a similar phase transition result for vanishing of $J_{n,0}$, the number of isolated nodes in $G(\mathcal{P}_n, r_n)$ i.e., thresholds for $\mathbb{P}\{J_{n,0} \geq 1\}$. That the threshold for vanishing of isolated nodes and threshold for connectivity coincide for a random geometric graph was inspired by a similar phenomenon observed in the case of the Erdös–Rényi random graphs [20] though the proof in the former case is a lot more involved.

$J_{n,0}$ is nothing but the number of isolated 1-cliques and it is natural to wonder if there is a sharp phase transition for vanishing of higher-order cliques and if so, is it related to any higher-dimensional topological phase transitions in random geometric graphs? We denote by $J_{n,k}$[3] the number of 'isolated' $(k+1)$-cliques in $G(\mathcal{P}_n, r)$, i.e., the number of $(k+1)$-cliques that do not belong to a $(k+2)$-clique. In other words, $J_{n,k}$ is the number of maximal cliques of order $(k+1)$. Due to non-monotonicity of $J_{n,k}$ in r, a threshold need not even exist.

The question of weak/sharp thresholds for higher-order connectivity entails two steps – (1) Determining the weak/sharp threshold for vanishing of 'isolated' clique counts and (2) Show that this coincides with the weak/sharp threshold of the corresponding notion of 'connectivity'. In this article, we shall focus on the first step. One of our results will give thresholds (i.e., r_n) for vanishing of $J_{n,k}$ on $G(\mathcal{P}_n, r_n)$ and for the case of $1 \leq k \leq d$, we shall show that our thresholds are sharp by showing that $J_{n,k}$ converges to a suitable Poisson random variable. Of course, such thresholds will have implications for the second step too i.e., 'connectivity' thresholds. In the next subsection (Section 1.1), we shall discuss in detail about the background literature on these combinatorial notions of connectivity and then preview our results in Section 1.2 as well as explaining further connections to existing results. The combinatorial topology notions that we shall need for these two subsections as well as the article are defined rigorously in Section 2.1. Our main results are stated in Sections 2.2–2.4 and we end with proofs in Section 3. We have included the Appendix describing some basic elements of simplicial homology and discrete Morse theory for ready reference random topology results in the paper.

1.1. *Up, down connectivity and related literature*

A natural higher-dimensional generalization of graphs are simplicial complexes, which can be concisely defined as hypergraphs closed under the operation of taking subsets of hyper-edges. We shall give precise definitions in Section 2.1. The analogous notion of clique counts in simplicial complexes is 'face counts'. We provide weak thresholds for vanishing of 'isolated' face counts in two models of random geometric complexes – Vietoris–Rips and Čech complexes. In each of these models, we shall consider two notions of connectivity – up and down – and hence two notions of 'isolation'. We shall define them shortly.

First, we would like to mention that our question or answer is not without precedence. Specifically, it was shown by Kahle in [31,33] that the threshold for vanishing of higher Betti numbers (a notion of higher-order connectivity) was linked to the threshold for vanishing of 'isolated' clique counts of *Erdös–Rényi flag/clique complexes*. A brief but formal definition of Betti numbers as well as some simplicial homology results are described in the Appendix. The

[3] Once we introduce different types of connectivity and complexes, we shall qualify this notation further.

0th Betti number is nothing but the number of connected components and hence this sharp phase transition result is a generalization of the result for the connectivity threshold for Erdös–Rényi random graphs. An earlier generalization of Erdös–Rényi result for a different model of random complexes called the *random d-complex* was shown by Linial and Meshulam and later by Meshulam and Wallach [37,38]. Both of these are models that generalize Erdös–Rényi graphs. We shall not discuss these models of random complexes any further apart from referring the reader to [34] for more details. The search for a geometric counterpart to the above results is still on despite a significant recent contribution by Bobrowski and Weinberger [11] which we shall discuss later. This is the broader aim towards which we take a step in this article. Betti numbers represent an algebraic notion of higher-dimensional connectivity and there are other more combinatorial notions of connectivity as we have indicated above and shall discuss now.

Let $\mathcal{K}$ be a finite simplicial complex (to be abbreviated as complex in future) which is a collection of "faces". Each face is assigned a non-negative integer-valued dimension and is composed of lower dimensional faces. Suppose $S_k(\mathcal{K})$ denotes the set of k-faces in $\mathcal{K}$, and $\sigma \in S_k(\mathcal{K})$ then we must have $\sigma' \in \mathcal{K}$ for every $\sigma' \subset \sigma$. In other words, if a face is in the complex then all lower dimensional faces contained in it are also part of the complex. The simple notion of connectivity in the graph case generalises to multiple notions of connectivity on complexes. We shall examine two such notions on two random geometric complexes. Given a complex $\mathcal{K}$, define the graph of 'up-connectivity', $G_k^{\mathcal{K},\mathcal{U}}$ as follows: The vertex set is $S_k(\mathcal{K})$ and $\sigma, \tau \in S_k(\mathcal{K})$ have an edge if $\sigma \cup \tau \in S_{k+1}(\mathcal{K})$. On $\mathcal{K}$, one can also define the graph of 'down-connectivity', $G_k^{\mathcal{K},\mathcal{D}}$ as follows: The vertex set is $S_k(\mathcal{K})$ and $\sigma, \tau \in S_k(\mathcal{K})$ have an edge if $\sigma \cap \tau \in S_{k-1}(\mathcal{K})$. In each of the next four paragraphs, we shall explain four different and unrelated contexts in which 'up-connectivity' and 'down-connectivity' have been considered. Thus, we hope to convince the reader that these notions of connectivity are worthy of further research not only for their intrinsic challenge and interest but also for their applications.

Having defined a graph, one can naturally consider connected components of the graph, random walk on the graph and the corresponding Laplacian. We shall denote the number of connected components of $G_k^{\mathcal{K},\mathcal{U}}$ and $G_k^{\mathcal{K},\mathcal{D}}$ as P_k (P-vector) and Q_k (Q-vector) respectively, $k = 0, 1, \dots$. Since $S_{-1}(\mathcal{K}) = \emptyset$ by convention, trivially $Q_0 = 0$. The Q-vectors are one of the central invariants in Q-analysis pioneered by R. Atkins in [2,3] to understand combinatorial connectivity of complexes. It is not possible to describe Q-analysis in detail and so we briefly mention that Q-analysis entails understanding 'combinatorial holes' in a complex via Q-vectors and related quantities. This has been later developed into a general theory of combinatorial connectivity of complexes known as combinatorial homotopy or A-homotopy theory. For more on this combinatorial homotopy theory and its applications, please refer to [4,5,36]. This is the first context in which the notion of 'down-connectivity' is relevant.

Now to the second context. One procedure to construct a complex from a graph G is to define $S_k(\mathcal{K})$ to be the set of all $(k+1)$-cliques in G. Such a complex is called clique complex of the graph G and we denote it by $\mathcal{K}(G)$. Viewing cliques as communities and to investigate overlapping of communities, Derenyi et al. in [18] studied percolation on the graph $G_k^{\mathcal{K}(G),\mathcal{D}}$ and termed it as clique percolation (though not using the terminology of simplicial complexes). This and further variants of clique percolation on Erdös–Rényi graphs was studied by [14] and the corresponding question for percolation in the up-connectivity graph of random geometric complexes was addressed in [7]. For a survey of this direction of research, see [45].

Even though the notion of down-connectivity has implicitly been used in Q-analysis, combinatorial homotopy and clique percolation without stating them explicitly, we shall now reference literature where these terms have appeared explicitly. These are the very recent studies of Laplacians on simplicial complexes [24,29,40,46], which is the third context where both 'up' and 'down' connectivity have been studied. Here again, there are two notions of Laplacians – up and down – and as expected they are related to up and down connectivities respectively. Irreducibility of the two Laplacians are related to the connectivity of $G_k^{\mathcal{K},\mathcal{U}}$ and $G_k^{\mathcal{K},\mathcal{D}}$ respectively. As we can observe that there are varied contexts in which the notions of up and down connectivity crop up but barring these few papers on Laplacians of random complexes and face percolation on random complexes, this is very much a fertile terrain. Our results give a lower bound on the thresholds for irreducibility of the two Laplacians and triviality of the Q-vector.

Now the fourth context, which we have touched upon earlier, is the study of Betti numbers of random complexes. This is also the direction of more extensive research on random complexes. Betti numbers are another way of quantifying connectivity of complexes. In fact, Betti numbers has been the main focus of most studies on random complexes. We point the reader to the surveys [10,34] for details on this growing area lying in the intersection of probability and topology. Motivated by applications to topological data analysis [15,16], random geometric complexes were introduced in [32] and among other things, upper bounds for thresholds on vanishing of Betti numbers for the Vietoris–Rips and Čech complexes on Poisson point processes were given. Similar thresholds were later proven for more general stationary point processes in [52] and for Poisson point processes on compact, closed manifolds without a boundary in [11]. To briefly allude to applications of such results in topological data analysis, we mention that a very weak threshold was used in the pioneering work of [42] to find homology of submanifolds from random samples. Since Betti numbers are algebraic quantities and the notions of up-connectivity and down-connectivity are combinatorial, apriori it not obvious why the two need to be related. However, for the random d-complex, it was shown that the threshold for up-connectivity was same as the threshold for vanishing of Betti numbers [35, Theorem 1.8]. It is worth repeating that this work is a step towards the geometric counterpart of such a result [4]

1.2. *Preview*: *Coarse-scale asymptotics*

Now, we shall survey some results of relevance to us before stating a few of our results. We refer the reader to the Appendix for a formal definition of Betti numbers and other simplicial homology results referred to below. Our results stated in this section shall be at the coarse-scale as the finer-scale results involve considerably more notation and are postponed to Section 2. Now onwards, let $\mathcal{P}_n$ be the homogeneous Poisson point process with intensity n on the unit d-dimensional torus for some $d \geq 2$. Also, when we refer to random Vietoris–Rips ($\mathcal{R}(\mathcal{P}_n, r_n)$) and Čech complex ($\mathcal{C}(\mathcal{P}_n, r_n)$), we refer to these complexes constructed on $\mathcal{P}_n$ using balls of radius r_n (precisely defined in Definitions 2.1 and 2.2). In short, the k-simplices/faces of the Čech complex (resp. Vietoris–Rips complex) are formed by subsets of $(k+1)$ distinct points of $\mathcal{P}_n$ such that the mutual interesection (resp. all pairwise intersections) of balls of radius r_n centred at these points is non-empty. For the random Čech complex, a significant contribution refining the afore-mentioned vanishing thresholds appeared recently in [13]. In particular, it was shown that (see [13, Theorem 5.4]) for a sequence $w(n) \to \infty$, the following holds: For $1 \leq k \leq d$,

$$\mathbb{E}\big[\beta_k\big(\mathcal{C}(\mathcal{P}_n, r_n)\big)\big] \to \begin{cases} \infty & \text{if } n\theta_d r_n^d = \log n + (k-2)\log\log n - w(n) \\ \beta_k(U) & \text{if } n\theta_d r_n^d = \log n + k\log\log n + w(n), \end{cases} \tag{1.2}$$

where $\beta_k(\cdot)$ denotes the kth Betti number. Apart from the gap between the upper and lower thresholds, this is an extension of (1.1) to higher dimensions in the algebraic sense. The above threshold result was extended to a phase transition result for the event $\{\beta_k(\mathcal{C}(\mathcal{P}_n, r_n)) \neq \beta_k(U)\}$ at a coarser scale [13, Corollary 5.5]: We have for $1 \leq k \leq d$ and any $\epsilon \in (0,1)$

$$\mathbb{P}\big\{\beta_k\big(\mathcal{C}(\mathcal{P}_n, r_n)\big) = \beta_k(U)\big\} \to \begin{cases} 0 & \text{if } n\theta_d r_n^d = (1-\epsilon)\log n \\ 1 & \text{if } n\theta_d r_n^d = (1+\epsilon)\log n. \end{cases} \tag{1.3}$$

The above threshold also corresponds with that of thresholds for complete coverage [21,27] and surprisingly reveals that β_0 (number of connected components) equals one at a much lower threshold than the other Betti numbers which all vanish "nearly" together and correspond to the threshold for complete coverage i.e., the event $\{U \subset \bigcup_{x \in \mathcal{P}_n} B_x(r_n)\}$.

As should be obvious by now, the question of connectivity in higher dimensions can be posed in at least three ways (up, down and Betti numbers) and for at least two different models of geometric complexes (Vietoris–Rips and Čech). Though thresholds for Betti numbers of random Čech complexes have been partially addressed in [13,32] but thresholds for other notions of connectivity remain still open. While Betti numbers of Čech complexes are non-trivial only for $k \leq d$ but the question of up and down connectivity are relevant for any k.

The specific geometry of random Čech complexes enables one to study them via coverage processes as well as Morse theory (see the Appendix). Indeed, the key tool in [13] is investigation of the critical points of the distance function $\rho_n : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto \min_{1 \leq i \leq N_n}\{|x - X_i|\}$. Since $\rho_n^{-1}[0, r] = \bigcup_{X \in \mathcal{P}_n} B_X(r)$, critical points of ρ_n are related to the Betti numbers of $\bigcup_{X \in \mathcal{P}_n} B_X(r)$ via Morse theory and the later are identical to those of random Čech complexes due to the nerve theorem (see the Appendix, [6, Theorem 10.7]). However, both these tools are either unavailable or insufficient for study of other notions of connectivity in the two models of geometric complexes. A possibly more universal approximation for study of connectivity thresholds in random complexes are 'isolated' face counts and this is the main reason why we focus on these objects in this article.

Let $G_k^{p,q}(\mathcal{P}_n, r_n)$, $k \geq 1$ denote the up and down connected graphs for $q \in \mathcal{I}_2 := \{\mathcal{U}, \mathcal{D}\}$ and the Čech and Vietoris–Rips complexes for $p \in \mathcal{I}_1 := \{\mathcal{C}, \mathcal{R}\}$. Recall that our Čech and Vietoris–Rips complexes are formed on $\mathcal{P}_n$ using balls of radius r_n. In the above notation, $\mathcal{C}$, $\mathcal{R}$ refer to the Čech and Vietoris–Rips cases respectively and $\mathcal{U}$, $\mathcal{D}$ refer to the up and down connectivity respectively. As a trailer for our results, we state the following coarse scale phase transition result for isolated k-faces in the random Vietoris–Rips and Čech complexes. The constants $m_k^{p,q}$ that appear in the Theorem are defined in Section 2.1. In what follows, we abbreviate $r_{n,k}^{p,q}$ by r_n.

[4]During the final revision of our draft, a couple of recent works [8,17] have made some progress and we shall briefly remark on these later.

Theorem 1.1. *Let $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$ and $k \geq 1$. Let $J_{n,k}^{p,q}(r)$ denote the number of isolated nodes in $G_k^{p,q}(\mathcal{P}_n, r)$. Then the following holds for any $\epsilon \in (0,1)$.*

$$\mathbb{P}\big\{J_{n,k}^{p,q}(r_n) = 0\big\} \to \begin{cases} 0 & \text{if } nm_k^{p,q} r_n^d = (1-\epsilon)\log n \\ 1 & \text{if } nm_k^{p,q} r_n^d = (1+\epsilon)\log n. \end{cases} \tag{1.4}$$

Indeed, for $nm_k^{p,q} r_n^d = (1-\epsilon)\log n$, we have that $\mathbb{E}[J_{n,k}^{p,q}(r_n)] \to \infty$ and $\frac{\log J_{n,k}^{p,q}(r_n)}{\log \mathbb{E}[J_{n,k}^{p,q}(r_n)]} \xrightarrow{P} 1$ as $n \to \infty$.

Since $J_{n,k}^{p,q}(r_n)$ is a non-monotonic functional in r for $k \geq 1$, even presence of a phase transition is not obvious. However, the above theorem shows a phase-transition for the existence of isolated nodes at a fixed radius r_n in the up/down-connected graphs. To gain a better understanding of the phase-transition, we consider the existence of isolated nodes for some radii $s \geq r_n$. This is a monotonic event and at the coarse scale has the same threshold for vanishing as existence of isolated nodes.

Theorem 1.2. *Let $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$ and $k \geq 1$. Let $J_{n,k}^{p,q}(r)$ denote the number of isolated nodes in $G_k^{p,q}(\mathcal{P}_n, r)$. Then the following holds for any $\epsilon \in (0,1)$*

$$\mathbb{P}\left\{\bigcap_{r \geq r_n} \big\{J_{n,k}^{p,q}(r) = 0\big\}\right\} \to \begin{cases} 0 & \text{if } nm_k^{p,q} r_n^d = (1-\epsilon)\log n \\ 1 & \text{if } nm_k^{p,q} r_n^d = (1+\epsilon)\log n. \end{cases} \tag{1.5}$$

We shall shortly see evidence that at a finer scale the thresholds for the events in Theorems 1.1 and 1.2 need not coincide at least for the Čech complex. Considering Theorem 1.1 as the first step towards determining thresholds for up/down-connectivity in both the random geometric complexes, here is the second step. The following theorem shows that whenever 'isolated' nodes vanish in the up/down-connected graphs, components of finite but fixed order also vanish.

Theorem 1.3. *Let $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$, $k \geq 1$ and $L \geq 1$. Let $J_{n,k}^{p,q}(r, L)$ denote the number of components in $G_k^{p,q}(\mathcal{P}_n, r)$ with exactly L vertices. Then for any $\epsilon > 0$ and $nm_k^{p,q} r_n^d = (1+\epsilon)\log n$, we have that*

$$\mathbb{E}\big[J_{n,k}^{p,q}(r_n, L)\big] \to 0.$$

Vanishing of isolated nodes in $G_k^{p,q}(\mathcal{P}_n, r_n)$ is a necessary condition for connectivity of $G_k^{p,q}(\mathcal{P}_n, r_n)$ but as mentioned before, this is also a sufficient condition in many random graph models. It is not completely obvious at this point if this is true even in the models we have considered above. We conjecture it to be true at least in the coarse scale with Theorem 1.3 as a partial eveidence.

Finer phase transition results for expectations and a distributional result inside the critical window for $J_{n,k}^{\mathcal{R},\mathcal{U}}$ with $k \leq d$ are stated in Section 2. An analogous distributional result is currently unavailable for other cases or similar statistics such as Morse critical points in random geometric complexes. An important tool in obtaining the distributional result for the Vietoris–Rips complex is the purely deterministic geometric Lemma 3.2 and such a result is not available for the Čech complex or Morse critical points. Stating these finer results shall involve considerably more notation and are hence postponed to Section 2. However, in the special case of $k = 1$, we can state these results with no additional notation as we shall do so now. Note that $J_{n,1}^{\mathcal{R},\mathcal{D}} = J_{n,1}^{\mathcal{C},\mathcal{D}}$.

Proposition 1.4. *Let $J_{n,1}^{p,q}(r_n)$ denote the number of isolated nodes in $G_1^{p,q}(\mathcal{P}_n, r_n)$. Let w_n be any real sequence converging to ∞ and $nr_n^d \to \infty$ as $n \to \infty$. For $(p,q) \in \mathcal{I}_1 \times \mathcal{I}_2 \setminus \{(\mathcal{C}, \mathcal{U})\}$ we have*

$$\mathbb{E}\big[J_{n,1}^{p,q}(r_n)\big] \to \begin{cases} \infty & \text{if } nm_1^{p,q} r_n^d = \log n - w_n \\ 0 & \text{if } nm_1^{p,q} r_n^d = \log n + w_n, \end{cases} \tag{1.6}$$

whereas

$$\mathbb{E}\big[J_{n,1}^{\mathcal{C},\mathcal{U}}(r_n)\big] \to \begin{cases} \infty & \text{if } nm_1^{\mathcal{C},\mathcal{U}} r_n^d = \log n - \log\log n - w_n \\ 0 & \text{if } nm_1^{\mathcal{C},\mathcal{U}} r_n^d = \log n - \log\log n + w_n. \end{cases} \tag{1.7}$$

Notice the difference in the thresholds at the level of expectation between Čech and Vietoris–Rips complexes. We shall also see that there is a difference between the two complexes when we look at a finer phase transition result corresponding to the event in Theorem 1.2. To be more precise, consider the event $\bigcup_{s\geq r}\{J^{p,q}_{n,1}(s)\geq 1\}$. We denote by $J^{p,q,*}_{n,k}(r)$ the number of k-faces (i.e., $k+1$ points) that contribute to the event or in other words, the number of k-faces that are 'isolated' for some $s\geq r$. This is defined more precisely in (2.12). Clearly, all 'isolated' faces at radius r are included i.e., $J^{p,q}_{n,k}(r)\leq J^{p,q,*}_{n,k}(r)$ and the latter is non-increasing in r. Thus, we have that

$$\{J^{p,q,*}_{n,k}(r)=0\}=\bigcap_{s\geq r}\{J^{p,q}_{n,k}(s)=0\}$$

and we have shown in Theorems 1.1 and 1.2 that the thresholds for vanishing of $J^{p,q}_{n,k}(\cdot)$ and $J^{p,q,*}_{n,k}(\cdot)$ coincide at the coarse scale. However, we shall see now that at a finer scale whether they coincide or not depends on the geometry of the complex.

Proposition 1.5. *Let $p\in\mathcal{I}_1$, $q\in\mathcal{I}_2$ and $k=1$. Let $J^{p,q,*}_{n,1}(r)$ denote the number of 1-faces that are isolated in $G^{p,q}_1(\mathcal{P}_n,s)$ for some $s\geq r$. Let w_n be any real sequence converging to ∞ and $nr_n^d\to\infty$ as $n\to\infty$. Then*

$$\mathbb{E}\big[J^{p,q,*}_{n,1}(r_n)\big]\to\begin{cases}\infty & \text{if } nm_1^{p,q}r_n^d=\log n-w_n\\ 0 & \text{if } nm_1^{p,q}r_n^d=\log n+w_n.\end{cases}\tag{1.8}$$

Note the missing $\log\log n$ factor for the Čech case i.e., as claimed earlier the thresholds for vanishing of $J^{\mathcal{C},\mathcal{U}}_{n,1}(\cdot)$ and $J^{\mathcal{C},\mathcal{U},*}_{n,1}(\cdot)$ do not coincide at least at the level of expectations. More importantly, this means that if we choose $w_n\to\infty$ such that $w_n=o(\log\log n)$ and $nm_1^{\mathcal{C},\mathcal{U}}r_n^d=\log n-\log\log n+w_n$, then

$$\mathbb{E}\big[J^{\mathcal{C},\mathcal{U}}_{n,1}(r_n)\big]\to 0,\quad \text{but } \mathbb{E}\big[J^{\mathcal{C},\mathcal{U},*}_{n,1}(r_n)\big]\to\infty.$$

In other words, for any fixed "$r\in(\log n-\log\log n,\log n)$" and for n large, we are unlikely to observe an 'isolated' face at r but we expect to see large number of them in the interval.

We shall now explain the consequences of our results for topological phase transitions as well as contrast them with the related results of [13]. While our results are on thresholds for vanishing of isolated nodes but as mentioned before, our tendentious view is that these are indicative of similar results for the corresponding connectivity thresholds.

Remark 1.6.

- Since $m_k^{\mathcal{C},\mathcal{U}}=\theta_d$ for all $k\geq 1$, we note that at the coarse scale the threshold for vanishing of isolated nodes in $G_k^{\mathcal{C},\mathcal{U}}(\mathcal{P}_n,r_n)$ matches with that of $\beta_k(\mathcal{P}_n,r_n)$. And like we pointed out for Betti numbers, $J^{\mathcal{C},\mathcal{U}}_{n,0}$ vanishes much earlier compared to $J^{\mathcal{C},\mathcal{U}}_{n,k}$ for $k\geq 1$ which all vanish "nearly together".
- At a finer scale (see Proposition 2.4), the threshold for vanishing of isolated nodes in $G_k^{\mathcal{C},\mathcal{U}}(\mathcal{P}_n,r_n)$ is analogous to that of vanishing of homology groups in (1.2) and in the case of $k=1$ matches exactly with the lower threshold in (1.2) (see Proposition 2.6). In [33,37,38], it was shown that thresholds for vanishing of isolated nodes in $G_k^{p,\mathcal{U}}(\mathcal{P}_n,r_n)$, $p\in\mathcal{I}_1$, was same as that of vanishing of homology groups in Erdös-Rènyi-like random complexes. Our results together with (1.2) offer evidence of such a phenomenon holding true even for random geometric Čech complexes under up-connectivity.
- Perhaps, a little ambitiously one can conjecture that the threshold for vanishing of isolated nodes in $G_k^{\mathcal{C},\mathcal{U}}(\mathcal{P}_n,r_n)$ should be $n\theta_d r_n^d=\log n+(k-2)\log\log n$, which corresponds to the lower threshold in (1.2). Recently in [8, Theorem 3.1], it is shown that thresholds for monotonic vanishing of Betti numbers (defined similar to $J^{p,q,*}_{n,k}$) is $n\theta_d r_n^d=\log n+(k-1)\log\log n$ and conjectured that the thresholds for vanishing (not monotonic) of Betti numbers will be at $n\theta_d r_n^d=\log n+(k-2)\log\log n$ (see remarks following [8, Corollary 7.14]). In the light of these recent results, our results in Propositions 1.4 and 1.5 offer evidence of similar behaviour for up-connectivity in Čech complexes and actually a proof in the case of $k=1$. Further, it indicates that the Vietoris–Rips complexes could exhibit a different behaviour. The difference between Čech and Vietoris–Rips is not only in that the thresholds are different and the latter thresholds depend on k but that there could be a difference between vanishing of Betti numbers and monotonic vanishing of Betti numbers.
- The four thresholds corresponding to vanishing of isolated nodes in $G_k^{p,q}(\mathcal{P}_n,r_n)$, $p\in\mathcal{I}_1$, $q\in\mathcal{I}_2$ are all different even in the coarser scale, since the corresponding constants $m_k^{p,q}$'s are different.

- For $p \in \mathcal{I}_1$ since $m_k^{p,\mathcal{U}} \le m_k^{p,\mathcal{U}}$, it follows that vanishing of isolated nodes in $G_k^{p,\mathcal{U}}(\mathcal{P}_n, r_n)$ occur for a radius that is at least as large as that required in $G_k^{p,\mathcal{D}}$ even in the coarse scale. A similar reasoning yields that even in the coarse scale, the threshold for vanishing of isolated nodes in $G_k^{\mathcal{C},q}(\mathcal{P}_n, r_n)$ is larger than in $G_k^{\mathcal{R},q}(\mathcal{P}_n, r_n)$ for any $q \in \mathcal{I}_2$.
- Since $m_k^{\mathcal{R},q}$ is strictly decreasing in k for $k \le d$, vanishing of isolated nodes of $G_k^{\mathcal{R},q}(\mathcal{P}_n, r_n)$ happen later than $G_j^{\mathcal{R},q}(\mathcal{P}_n, r_n)$ for $j < k \le d$ even at the coarse scale for any $q \in \mathcal{I}_2$. This is in contrast to the scenario for $G_k^{\mathcal{C},\mathcal{U}}(\mathcal{P}_n, r_n)$ for which $m_k^{\mathcal{C},\mathcal{U}} = \theta_d$ for all $k \ge 1$.

We shall now say a few words on our proof methods. While the main tools for obtaining asymptotics for 'isolated' faces are the classical Palm theory and Campbell–Mecke formula, the specific geometric analysis pertaining to 'isolated faces' differ from that of Morse critical points. Apart from this, we shall see that the non-monotonicity of 'isolated' nodes complicates the analysis. In such high-density regimes, the asymptotics are usually determined by the 'minimal configuration' contributing to the functional. However, it is very important to understand the behaviour in the neighbourhood of the 'minimal configuration'. This we shall see is far from clear for 'isolated' faces in contrast to Morse critical points. In special cases such as $k = 1$, we are able to understand the neighbourhood of the 'minimal configuration' to be able to give more detailed results though the techniques vary considerably from case to case. As is evident in our results, such problems are also a matter of scale. Often in coarse scale, we are able to overcome these issues with respect to 'the minimal configuration' more easily.

The connection between Morse critical points and Betti numbers is a deterministic fact whereas the connection between 'isolated' faces and Betti numbers arises mainly in random contexts. This is one reason why translating our results to thresholds for Betti numbers is incomplete at the moment. However, for the two discrete models of random simplicial complexes – Erdös–Rényi clique complexes [33] and random d-complexes [37,38] – using different methods it has been shown that the threshold for vanishing of 'isolated' faces in the up-connected graph corresponds to the threshold for vanishing of homology.

In the graph case, closely related to connectivity threshold is the largest edge-weight on a minimal spanning tree. It was shown in [47] that asymptotically the longest edge on a minimal spanning tree on the complete graph on $\mathcal{P}_n$ with weights as the Euclidean distance is same as the largest nearest neighbour distance i.e., the smallest radius at which all the vertices are non-isolated in the random geometric graph on $\mathcal{P}_n$. Simplicial analogues of spanning tree are called as *spanning acycles* and their behaviour on randomly weighted complexes have been investigated in [28,51]. In particular, see [51, Section 3] for relations between spanning acycles, Betti numbers and 'isolated' faces. We see our work as another step towards establishing such relations for random geometric complexes.

To end the introduction, we shall point a few more directions in which our work could be extended apart from the natural program of computing exact thresholds and investigating the critical window. One important problem would be to investigate higher-dimensional or distributional analogues of Propositions 1.4 and 1.5. One could also consider geometric complexes on compact Riemannian manifolds and Poisson point processes with non-uniform densities. For ideas on the former extension, see [8,12,13,17] and for the latter see [25,30,43,44,48].

2. Main results

2.1. *Some combinatorial topology notions*

A subset $\mathcal{K} \subset 2^{\mathcal{X}}$ for a finite point-set $\mathcal{X}$ is said to be an *abstract simplicial complex* (abbreviated as complex in future) if $A \in \mathcal{K}$ and $B \subset A$ implies that $B \in \mathcal{K}$. The elements of $\mathcal{K}$ are called faces or simplices and the dimension $\dim \sigma$ of a face σ is $|\sigma| - 1$. We shall denote a k-face by $[v_1, \dots, v_{k+1}]$. The maximal faces (faces that are not included in any other faces) are called *facets*. By convention, $\emptyset \in \mathcal{K}$ and $\dim(\emptyset) = -1$. The collection of k-faces is denoted by $S_k(\mathcal{K})$ and the k-skeleton of $\mathcal{K}$ is the complex $\mathcal{K}^k := \bigcup_{i=-1}^{k} S_i(K)$. A complex is said to be *pure* if all maximal faces have the same dimension. Note that $\mathcal{K}^1$ is nothing but a graph.

Denote by $B_x(r)$ a closed ball of radius r centered at x. $|\cdot|$ will denote the cardinality of a finite set as well as the Lebesgue measure and $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$. There are two types of complexes defined on point processes. These complexes are the Vietoris–Rips complex and the Čech complex which we now define. Let $U = [0, 1]^d$ be equipped with the *toroidal metric* i.e.,

$$d(x, y) = \inf\{\|x - y + z\| : z \in \mathbb{Z}^d\}, \quad x, y \in U.$$

For $\mathbf{x} = (x_1, \dots, x_{k+1}) \in \mathbb{R}^{d(k+1)}$, let $B_{\mathbf{x}}(r) = \bigcup_{i=1}^{k+1} B_{x_i}(r)$, $h(\mathbf{x}) = h(x_1, \dots, x_{k+1})$ for $h : \mathbb{R}^{d(k+1)} \to \mathbb{R}$ and $d\mathbf{x} = dx_1 \dots dx_{k+1}$. Let $\mathbf{1} = (1, \dots, 1)$. Let $\mathcal{X}$ be a finite set in U and we shall use $\mathcal{X}^{(k)}$ to denote the set of k-tuples of distinct points in $\mathcal{X}$.

Definition 2.1 (Vietoris–Rips complex). The abstract simplicial complex $\mathcal{R}(\mathcal{X}, r)$ constructed as below is called the Vietoris–Rips complex associated to $\mathcal{X}$ and r.

(1) The 0-simplices of $\mathcal{R}(\mathcal{X}, r)$ are the points in $\mathcal{X}$.
(2) For $k \geq 1$, a k-simplex, or k-dimensional 'face', $\sigma = [x_{i_1}, \dots, x_{i_{k+1}}]$ is in $\mathcal{R}(\mathcal{X}, r)$ if $B_{x_{i_j}}(r) \cap B_{x_{i_m}}(r) \neq \varnothing$ for every $1 \leq j < m \leq k+1$ and where $(x_{i_1}, \dots, x_{i_{k+1}}) \in \mathcal{X}^{(k+1)}$.

Definition 2.2 (Čech complex). The abstract simplicial complex $\mathcal{C}(\mathcal{X}, r)$ constructed as below is called the Čech complex associated to $\mathcal{X}$ and r.

(1) The 0-simplices of $\mathcal{C}(\mathcal{X}, r)$ are the points in $\mathcal{X}$,
(2) For $k \geq 1$, a k-simplex, or k-dimensional 'face', $\sigma = [x_{i_1}, \dots, x_{i_{k+1}}]$ is in $\mathcal{C}(\mathcal{X}, r)$ if $\bigcap_{j=1}^{k+1} B_{x_{i_j}}(r) \neq \varnothing$ and where $(x_{i_1}, \dots, x_{i_{k+1}}) \in \mathcal{X}^{(k+1)}$.

Observe that the faces of a Vietoris–Rips complex are nothing but cliques of a random geometric graph and the 1-skeletons of both the Vietoris–Rips and Čech complexes coincide with the random geometric graph.

The functionals that we study in this paper are the isolated simplex counts in the Čech and the Vietoris–Rips complexes. In addition there are two notions of connectivity, up and down which in turn determines what constitutes an isolated simplex. For any $k \geq 0$, let $S_k(\mathcal{X}, r)$ be the collection of all k-simplices of the Čech complex on $\mathcal{X}$. Consider the graph $G_k^{\mathcal{C},\mathcal{U}}(\mathcal{X}, r)$ with vertex set $S_k(\mathcal{X}, r)$ and with an edge between any two elements $\sigma_1, \sigma_2 \in S_k(\mathcal{X}, r)$ provided they are up-connected, that is, $\sigma_1 \cup \sigma_2 \in S_{k+1}(\mathcal{X}, r)$. Similarly for any $k \geq 1$, we can define the graphs $G_k^{\mathcal{C},\mathcal{D}}(\mathcal{X}, r)$ with edges between elements $\sigma_1, \sigma_2 \in S_k(\mathcal{X}, r)$ that are down connected, that is, $\sigma_1 \cap \sigma_2 \in S_{k-1}(\mathcal{X}, r)$. The graphs $G_k^{\mathcal{R},\mathcal{U}}(\mathcal{X}, r)$ and $G_k^{\mathcal{R},\mathcal{D}}(\mathcal{X}, r)$ are defined similarly by taking $S_k(\mathcal{X}, r)$ to be the collection of all k-simplices in the Vietoris–Rips complex. See Figure 1 for an illustration of the two complexes and their maximal faces. Fix $k \geq 2$. For $\mathbf{x} = (x_1, \dots, x_{k+1}) \in \mathbb{R}^{d(k+1)}$ and $r > 0$ define the functions

$$h_k^{\mathcal{C}}(\mathbf{x}, r) = \mathbf{1}\left[\bigcap_{j=1}^{k+1} B_{x_j}(r) \neq \varnothing\right],$$

$$h_k^{\mathcal{R}}(\mathbf{x}, r) = \prod_{1 \leq j < m \leq k+1} \mathbf{1}\left[B_{x_j}(r) \cap B_{x_m}(r) \neq \varnothing\right]. \tag{2.1}$$

Recall that $\mathcal{I}_1 = \{\mathcal{C}, \mathcal{R}\}$ and $\mathcal{I}_2 = \{\mathcal{U}, \mathcal{D}\}$. We will denote by $h_k^p(\mathbf{x})$ the function $h_k^p(\mathbf{x}, 1)$ for $p \in \mathcal{I}_1$, $k \geq 1$. For $A \subset U$ and $s > 0$ let

$$A^{(s)} := \{x \in U : \text{there exists } y \in A \text{ such that } d(x, y) \leq s\} = \{x + y : x \in A, y \in B_O(s)\}$$

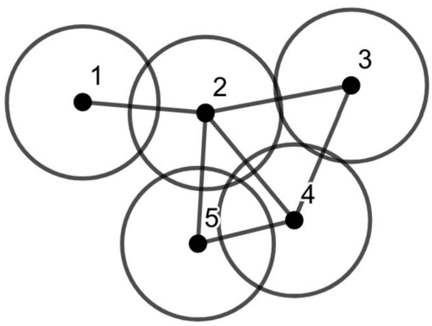

Fig. 1. $\{[1, 2], [2, 3], [3, 4], [2, 4, 5]\}$ are the maximal faces of the Čech complex on the point set $\mathcal{X} = \{1, \dots, 5\}$ whereas $\{[1, 2], [2, 3, 4], [2, 4, 5]\}$ are the maximal faces of the Vietoris–Rips complex on $\mathcal{X}$.

be the closed s-neighbourhood of the subset A and O denotes the origin. For $x \in \mathbb{R}^{d(k+1)}$ and $r, s > 0$ define the set-valued functions

$$\begin{aligned} Q_k^{\mathcal{C},\mathcal{U}}(\mathbf{x}, r, s) &= \left(\bigcap_{j=1}^{k+1} B_{x_j}(r)\right)^{(s)} \\ Q_k^{\mathcal{C},\mathcal{D}}(\mathbf{x}, r, s) &= \bigcup_{i=1}^{k+1}\left(\bigcap_{j=1, j\neq i}^{k+1} B_{x_j}(r)\right)^{(s)} \\ Q_k^{\mathcal{R},\mathcal{U}}(\mathbf{x}, r, s) &= \bigcap_{j=1}^{k+1}\big(B_{x_j}(r)\big)^{(s)} = \bigcap_{j=1}^{k+1} B_{x_j}(r+s) \\ Q_k^{\mathcal{R},\mathcal{D}}(\mathbf{x}, r, s) &= \bigcup_{i=1}^{k+1}\left(\bigcap_{j=1, j\neq i}^{k+1} B_{x_j}(r+s)\right). \end{aligned} \tag{2.2}$$

We will also use the abbreviated forms

$$Q_k^{p,q}(\mathbf{x}, r) = Q_k^{p,q}(\mathbf{x}, r, r) \quad \text{and} \quad Q_k^{p,q}(\mathbf{x}) = Q_k^{p,q}(\mathbf{x}, 1), \quad p \in \mathcal{I}_1, q \in \mathcal{I}_2. \tag{2.3}$$

Let

$$A_k^p := \big\{\mathbf{y} \in \big(\mathbb{R}^d\big)^k : h_k^p(O, \mathbf{y}) = 1\big\}, \tag{2.4}$$

be the set of configurations that form a k-simplex with the origin. Clearly $A_k^p \subset B_O(2)^k$ and since our complexes are defined using closed balls, A_k^p is compact. For any $\mathbf{y} \in A_k^p$ and finite point set $\mathcal{X} \subset \mathbb{R}^d$, the k-simplex comprising the points $\{O, \mathbf{y}\} \in (\mathbb{R}^d)^{k+1}$ is isolated in $G_k^{p,q}(\mathcal{X} \cup \{O, \mathbf{y}\}, 1)$ provided in $Q_k^{p,q}(O, \mathbf{y}) \cap \mathcal{X} = \varnothing$. Let

$$m_k^{p,q} = \inf\big\{\big|Q_k^{p,q}(O, \mathbf{y})\big| : \mathbf{y} \in A_k^p\big\}, \quad \text{and} \quad M_k^{p,q} = \sup\big\{\big|Q_k^{p,q}(O, \mathbf{y})\big| : \mathbf{y} \in A_k^p\big\}. \tag{2.5}$$

Often, the choice of k is clear from the context and hence we shall suppress it. It is easy to see that $m_k^{\mathcal{C},\mathcal{U}} = \theta_d$ and $m_k^{\mathcal{C},\mathcal{D}} = 2^d\theta_d$ for all $k \geq 1$. The former occurs for any configuration of points that are as far apart as possible such that the common intersection of balls is a single point and the latter happens when all the points coincide. Note that $m_k^{\mathcal{C},q}$ does not depend on k. For $k \leq d$, $m_k^{\mathcal{R},\mathcal{U}}$ is the volume of the lens of intersection of $(k+1)$ balls of radius 2 that can be placed in $\mathbb{R}^d$ so that their centers are exactly at distance 2 from each other. The region $Q^{p,q}(\mathbf{x})$ for a *minimal* configuration is illustrated in Figures 2–4 for the case $d = 2$ and $k = 1$. The volume of the shaded region in the figures is $m^{p,q}$. The balls in Figure 2 are of unit radii and are of radii 2 in Figures 3, 4.

2.2. *Expectation asymptotics for isolated face counts*

In this section we present a radius regime under which the expected isolated complex count stabilizes in the limit. This regime involves a parameter sequence whose asymptotic behavior is described. Later part of the section contains a result on the rate of convergence of this parameter sequence which has some interesting implications that will be discussed.

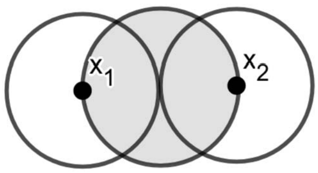

Fig. 2. $p = \mathcal{C}, q = \mathcal{U}$.

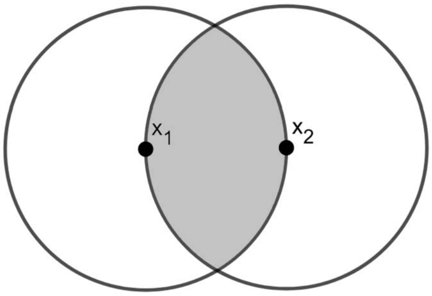

Fig. 3. $p = \mathcal{R}, q = \mathcal{U}$.

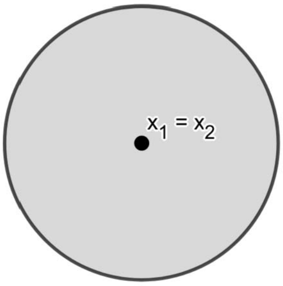

Fig. 4. $p \in \mathcal{I}_1, q = \mathcal{D}$.

Definition 2.3 (Isolated face counts). Let $k \geq 1$. For $p \in \mathcal{I}_1$ and $q \in \mathcal{I}_2$ the number of isolated simplices in the graph $G_k^{p,q}(\mathcal{X}, r)$ is defined as

$$J_k^{p,q}(\mathcal{X}, r) := \frac{1}{(k+1)!} \sum_{\mathbf{x} \in \mathcal{X}^{(k+1)}} h_k^p(\mathbf{x}, r) \mathbf{1}\big[\mathcal{X} \cap Q_k^{p,q}(\mathbf{x}, r) = \mathbf{x}\big]. \tag{2.6}$$

For example, $J_k^{\mathcal{R},\mathcal{U}}(\mathcal{X}, r)$ counts the number of maximal $(k+1)$-cliques in the random geometric graph. As a more concrete example, in Figure 1, we have that $J_1^{\mathcal{R},\mathcal{D}}(\mathcal{X}, r) = J_1^{\mathcal{C},\mathcal{D}}(\mathcal{X}, r) = 0$, $J_1^{\mathcal{R},\mathcal{U}}(\mathcal{X}, r) = 1$, $J_1^{\mathcal{C},\mathcal{U}}(\mathcal{X}, r) = 3$.

Let $\mathcal{P}_n$ be a Poisson point process with intensity $n\mathbf{1}_U(\cdot)$ where $U = [0, 1]^d$ is the unit cube with the toroidal metric. Our first result is on the radius regime r_n that stabilizes the expected number of isolated simplices in the connectivity regime. For any $k \geq 1$, $\alpha \in \mathbb{R}$, $p \in \mathcal{I}_1$ and $q \in \mathcal{I}_2$, $c > 0$ define the sequence of radial functions $\{r_n^{p,q}(c)\}_{n \geq n_0}$ as

$$r_n(c) = r_{n,k}^{p,q}(c) = \left(\frac{\log n + k \log\log n + \log|A^p| + \alpha - k \log m^{p,q} - \log(k+1)!}{nc} \right)^{\frac{1}{d}}, \tag{2.7}$$

where n_0 is defined so that for all $n \geq n_0$, $r_n > 0$. Note that n_0 does not depend on c.

We will denote $J_k^{p,q}(\mathcal{P}_n, r)$ by $J_{n,k}^{p,q}(r)$. We abbreviate A_k^p, $M_k^{p,q}$, $m_k^{p,q}$, $Q_k^{p,q}(\cdot)$ by A, M, m, Q respectively.

Proposition 2.4. *Let $k \geq 1$, $\alpha \in \mathbb{R}$, $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$ and $r_n(c)$ be as defined in (2.7). Then there exists a sequence $c_n = c_{n,k}^{p,q} \in [m, M]$ such that $c_n \to m$ and*

$$\mathbb{E}\big[J_{n,k}^{p,q}\big(r_n(c_n)\big)\big] \to e^{-\alpha}, \quad \text{as } n \to \infty. \tag{2.8}$$

Though the constant $c_n \to m$, one cannot replace c_n by m in Proposition 2.4 as shown by the following result.

Proposition 2.5. *Let $k \geq 1$, $\alpha \in \mathbb{R}$, $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$ and r_n be such that $nmr_n^d = \log n + k \log\log n + w_n^1$ where w_n^1 is a sequence bounded from below i.e.*, $\liminf_{n\to\infty} w_n^1 > -\infty$. *Then*

$$\mathbb{E}\big[J_{n,k}^{p,q}\big(r_n(m)\big)\big] \to 0, \quad \text{as } n \to \infty. \tag{2.9}$$

It would be desirable to obtain precise estimates on the sequence c_n in Proposition 2.4. We explore this and a few other results for the case $k = 1$ in the next subsection.

2.2.1. *The* 1*-simplex*

For the case $k = 1$, we derive a result on the rate of convergence of the sequence c_n. The maximal face in a 1-complex is an edge. An edge is isolated in the sense of up-connectivity provided it is not part of a triangle i.e., a 2-simplex. By Definition 2.1, three nodes form a triangle in the Vietoris–Rips complex provided the balls centered at these points have non-trivial pairwise intersection. In the Čech complex (see Definition 2.2), the balls centered at these vertices must have a mutual intersection. Down connectivity for both the Rips and the Čech cases are identical. An edge is isolated in the down sense if it does not share a vertex with any other edge.

From (2.7), we have that

$$n\big(r_n(c_n)\big)^d c_n - \log n - \log\log n$$

is a bounded sequence. If we re-write $r_n(c_n)$ in terms of m, then it alters the coefficient of the $\log\log n$ term. This is the content of the first claim in the following proposition. A consequence of this result yields a rate of convergence for the sequence c_n.

Proposition 2.6. *For any $\alpha \in \mathbb{R}$ let $a^{\mathcal{C},\mathcal{U}} = 2$ and $a^{p,q} = 1$ for $(p,q) \in \mathcal{I}_1 \times \mathcal{I}_2 \setminus (\mathcal{C},\mathcal{U})$. Let $r_n(c_n) = r^{p,q}_{n,1}(c^{p,q}_{n,1})$, $p \in \mathcal{I}_1$ and $q \in \mathcal{I}_2$ be the sequence as in Proposition 2.4 for which the expected number of isolated edges, $\mathbb{E}[J^{p,q}_{n,1}(r_n(c_n))] \to e^{-\alpha}$ with $k = 1$. Then*

$$n\big(r_n(c_n)\big)^d m - \log n - \big(1 - a^{p,q}\big)\log\log n \tag{2.10}$$

is a bounded sequence. Further,

$$\lim_{n\to\infty}\left(\frac{c_n}{m} - 1\right)\frac{\log n}{\log\log n} = a^{p,q}. \tag{2.11}$$

Though the statement of the above theorem covers all the cases via a single equation, the estimates in different cases require somewhat different ideas. It is somewhat surprising that $a^{\mathcal{C},\mathcal{U}} = 2$ in the contrast to the other cases and the implication of this for the threshold for vanishing of isolated faces has been mentioned in Remark 1.6.

2.3. *Monotonic vanishing of isolated faces*

Let $p \in \mathcal{I}_1$ and $q \in \mathcal{I}_2$. Given $\mathbf{x} = (x_1, \dots, x_{k+1}) \in \mathbb{R}^{d(k+1)}$, define $R^p_k(\mathbf{x}) := \inf\{r : h^p(\mathbf{x}, r) = 1\}$. For example, in the Vietoris–Rips complex, we have that $2R^{\mathcal{R}}_k(\mathbf{x}) = \max_{i\neq j}|x_i - x_j|$. Define *the number of isolated faces at r and above* as follows:

$$J^{p,q,*}_k(\mathcal{X}, r) := J^{p,q}_k(\mathcal{X}, r) + \sum_{\mathbf{x}\in\mathcal{X}^{(k+1)}} \mathbf{1}\big[R^p_k(\mathbf{x}) > r\big]\mathbf{1}\big[\mathcal{X} \cap Q^{p,q}_k(\mathbf{x}), R^p_k(\mathbf{x})) \equiv x\big], \tag{2.12}$$

where the $Q^{p,q}$'s are defined in (2.3). By definition, it is clear that $J^{p,q,*}_k(\mathcal{X}, r) \geq J^{p,q}_k(\mathcal{X}, r)$. We have already seen coarse-scale thresholds for vanishing of $J^{p,q,*}_{n,k}(r) := J^{p,q,*}_k(\mathcal{P}_n, r)$, $k \geq 1$ in Theorem 1.2 and a finer threshold for vanishing of $J^{p,q,*}_{n,1}(r)$ in Proposition 1.5. We now give a finer upper bound for the threshold for vanishing of $J^{p,q,*}_{n,k}(r)$ for all $k \geq 1$.

Proposition 2.7. *Fix $k \geq 1$, $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$ and let $J^{p,q,*}_{n,k}(r_n) := J^{p,q,*}_{n,k}(\mathcal{P}_n, r_n)$ where $nmr_n^d = \log n + k\log\log n + w^1_n$ for some sequence w^1_n bounded from below i.e., $\liminf_{n\to\infty} w^1_n > -\infty$. Then we have that*

$$\mathbb{E}\big[J^{p,q,*}_{n,k}(r_n)\big] \to 0, \quad \text{as } n \to \infty.$$

2.4. *Poisson convergence for isolated Vietoris–Rips simplices under up-connectivity*

Our next result is a weak convergence result for the number of isolated simplices in the Vietoris–Rips complex.

Theorem 2.8. *Let $\alpha \in \mathbb{R}$, $d \geq 2$, $0 \leq k \leq d$, $r_n(c) = r^{\mathcal{R},\mathcal{U}}_{n,k}(c)$ be as defined in (2.7) and $c_n = c^{\mathcal{R},\mathcal{U}}_{n,k}$ be the sequence as in Proposition 2.4 i.e., $\mathbb{E}[J^{\mathcal{R},\mathcal{U}}_{n,k}(r_n(c_n))] \to e^{-\alpha}$. Then the number of isolated k-simplices $J^{\mathcal{R},\mathcal{U}}_{n,k}(r_n(c_n))$ converges in distribution to a Poisson random variable with mean $e^{-\alpha}$.*

The above distributional result extends to finite connected components in $G^{\mathcal{R},\mathcal{U}}(\mathcal{P}_n, r_n^{\mathcal{R},\mathcal{U}}(c_n))$. Recall that $J_{n,k}^{\mathcal{R},\mathcal{U}}(r,\ell)$ denotes the number of components in $G_k^{\mathcal{R},\mathcal{U}}(\mathcal{P}_n, r)$ with ℓ vertices (see Theorem 1.3).

Theorem 2.9. *Let $\alpha \in \mathbb{R}$, $d \geq 2$ and $0 \leq k \leq d$. Suppose that $c_n = c_{n,k}^{\mathcal{R},\mathcal{U}}$ is the sequence as in Proposition 2.4 i.e., $\mathbb{E}[J_{n,k}^{\mathcal{R},\mathcal{U}}(r_n(c_n))] \to e^{-\alpha}$. Let $L \geq 1$. Then $\sum_{l=1}^{L} J_{n,k}^{\mathcal{R},\mathcal{U}}(r_n(c_n), l)$ converges in distribution to a Poisson random variable with mean $e^{-\alpha}$.*

3. Proofs

Since in our proofs, we shall be working with a fixed k, we will drop the subscript k unless we want to emphasize the particular value of k or if there is any ambiguity. For example, we shall denote $Q_k^{p,q}$ by $Q^{p,q}$ etc. We shall also drop the superscripts p, q and use, for instance, J_n for $J_{n,k}^{p,q}$ unless we are carrying out a case-by-case analysis. In what follows, $C_1, C_2, \dots$ will denote finite constants whose values will change from place to place.

3.1. *Proofs of results in Section 2.2*

Proof of Proposition 2.4. Fix $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$, $k \geq 1$ but as mentioned above we shall drop these subscripts and superscripts for brevity of notation. For $\mathbf{x} = (x_1, \dots, x_{k+1})$, let $h_n(\mathbf{x}) = h(\mathbf{x}, r_n)$ and $Q_n(\mathbf{x}) = Q(\mathbf{x}, r_n)$ be as defined in (2.1) and (2.2). Without loss of generality, we can assume that $n \geq n_0$, where n_0 is as chosen in (2.7).

For any sequence of radial functions r_n, by the Campbell–Mecke formula (see [48, Theorem 1.6]) we have

$$\mathbb{E}\big[J_n(r_n)\big] = \frac{n^{k+1}}{(k+1)!} \int_{U^{k+1}} h_n(\mathbf{x}) e^{-n|Q_n(\mathbf{x})|} \, d\mathbf{x}. \tag{3.1}$$

Setting $x_i = x_1 + r_n y_i$, $i = 2, \dots, k+1$, in (3.1) and using the fact that $|Q_n(\mathbf{x})| = r_n^d |Q(\mathbf{x})|$ we get

$$\mathbb{E}\big[J_n(r_n)\big] = \frac{n(nr_n^d)^k}{(k+1)!} \int_{U \times ((r_n)^{-1}(U - x_1))^k} h(x_1, \mathbf{y}) e^{-nr_n^d |Q(x_1,\mathbf{y})|} \, dx_1 \, d\mathbf{y}, \tag{3.2}$$

where $\mathbf{y} = (y_2, \dots, y_{k+1})$. Let $n_1 \geq n_0$ be such that for all $n \geq n_1$ such that $\bigcup_{x \in U} B_x(1) \subset (r_n)^{-1}(U - x_1))$ for all $x_1 \in U$. Since the metric is toroidal, we obtain for all $n \geq n_1$ that

$$\mathbb{E}\big[J_n(r_n)\big] = \frac{n(nr_n^d)^k}{(k+1)!} \int_A e^{-nr_n^d |Q(O,\mathbf{y})|} \, d\mathbf{y}, \tag{3.3}$$

where $A = A^p$ is as defined in (2.4).

For any $c \in \mathbb{R}_+$ and $\alpha \in \mathbb{R}$, recall that $r_n(c)$ was defined in (2.7) as follows:

$$r_n(c) = \left(\frac{\log n + k \log\log n + \log|A| + \alpha - k \log m - \log(k+1)!}{nc} \right)^{\frac{1}{d}}.$$

The function

$$c \mapsto \int_A e^{-nr_n(c)^d |Q(O,\mathbf{y})|} \, d\mathbf{y}$$

is continuous in c, and for all $n > n_1$, tends to 0 as $c \to 0$ and tends to $|A|$ as $c \to \infty$. Since $e^{-nr_n(c)^d c} \in (0,1)$ does not depend on c, by the intermediate value theorem there exists a sequence c_n such that

$$\int_A e^{-nr_n(c_n)^d |Q(O,\mathbf{y})|} \, d\mathbf{y} = |A| e^{-nr_n(c_n)^d c_n}. \tag{3.4}$$

With the above choice of c_n, we obtain from (3.3)–(3.4) that

$$\begin{aligned}
\mathbb{E}\big[J_n\big(r_n(c_n)\big)\big] &= \frac{n(nr_n(c_n)^d)^k}{(k+1)!}|A|e^{-nr_n(c_n)^d c_n} \\
&= \frac{n(nr_n(c_n)^d)^k}{(k+1)!}|A|e^{-(\log n + k\log\log n + \log|A| + \alpha - k\log m - \log(k+1)!)} \\
&= \left(\frac{nr_n(c_n)^d m}{\log n}\right)^k e^{-\alpha} \\
&= \left(\frac{\log n + k\log\log n + \log|A| + \alpha - k\log m - \log(k+1)!}{\log n}\right)^k \left(\frac{m}{c_n}\right)^k e^{-\alpha}. \qquad (3.5)
\end{aligned}$$

Thus the proof is complete provided we show that $c_n \in (m, M)$ and $c_n \to m$. From the definition of m, M and the fact that the function $Q(O, \mathbf{y})$ achieves these values only on a set of zero measure, we have

$$|A|e^{-nr_n(c)^d M} < \int_A e^{-nr_n(c)^d|Q(O,\mathbf{y})|}\,d\mathbf{y} < |A|e^{-nr_n(c)^d m},$$

which together with (3.4) implies that $c_n \in (m, M)$.

Suppose $\limsup_{n\to\infty} c_n > m$. Then we can choose $m_1 > m$, $\epsilon > 0$ and a subsequence $\{n_j\}_{j\geq 1}$ such that $c_{n_j}(1-\epsilon) > m_1$ for all $j \geq 1$. From (3.3) we derive by calculations similar to the one in (3.5) that

$$\begin{aligned}
\mathbb{E}\big[J_{n_j}\big(r_{n_j}(c_{n_j})\big)\big] &\geq \frac{n_j(n_j r_{n_j}^d)^k}{(k+1)!}\int_{A\cap\{Q(O,\mathbf{y})<m_1\}} e^{-n_j r_{n_j}^d|Q(O,\mathbf{y})|}\,d\mathbf{y} \\
&\geq Cn_j\big(n_j r_{n_j}^d\big)^k \exp\left(-\frac{m_1}{c_{n_j}}\big(\log n_j + k\log\log n_j + \log|A| + \alpha - k\log m - \log(k+1)!\big)\right) \\
&\geq Cn_j\big(n_j r_{n_j}^d\big)^k \exp\big(-(1-\epsilon)\big(\log n_j + k\log\log n_j + \log|A| + \alpha - k\log m - \log(k+1)!\big)\big) \to \infty.
\end{aligned}$$

This contradicts the fact that from (3.5) and $c_n > m$, we must have $\limsup_{n\to\infty}\mathbb{E}[J_n(r_n(c_n))] \leq e^{-\alpha}$. Hence $\limsup_{n\to\infty} c_n = m$ and since $c_n \in (m, M)$, it follows that $\lim_{n\to\infty} c_n = m$. □

Proof of Proposition 2.5. Dropping the superscripts p, q we have from (3.3) that

$$\begin{aligned}
\mathbb{E}\big[J_n\big(r_n(m)\big)\big] &= \frac{n(nr_n^d)^k}{(k+1)!}e^{-nmr_n^d}\int_{A^p} e^{-nr_n^d(|Q(O,\mathbf{y})|-m)}\,d\mathbf{y} \\
&\leq C_1 e^{-w_n^1}\int_{A^p} e^{-nr_n^d(|Q(O,\mathbf{y})|-m)}\,d\mathbf{y} \to 0, \quad n\to\infty. \qquad (3.6)
\end{aligned}$$

The convergence above follows by the bounded convergence theorem because A^p is compact, $|Q(O,\mathbf{y})| \geq m$ on A^p and $e^{-nr_n^d(|Q(O,\mathbf{y})|-m)} \to 0$ almost surely on A^p. □

In the following proof, we shall use the Steiner's formula (see [50, (1.2)]) which states that for any non-empty convex body (compact convex subset) $K \subset \mathbb{R}^d$ and $r > 0$, we have that

$$V_d\big(K \oplus B_O(r)\big) = \sum_{i=0}^d r^{d-i} c_{i,d} V_i(K), \qquad (3.7)$$

where $V_i(K)$ is the *ith intrinsic volume* of K and $c_{i,d} = \theta_{d-i}$, the volume of the $(d-i)$-dimensional unit ball. Indeed, the intrinsic volumes can be also defined via the Steiner's formula once it is shown that $V_d(K \oplus B_O(r))$ is a polynomial in r. However, for our purposes, the following properties of the intrinsic volumes will suffice (see [50, Section 14.2]): (i) monotonicity under set inclusion and additivity for non-empty convex bodies and (ii) homogeneity i.e., $V_j(tK) = t^j V_j(K)$ for all $t > 0$, $0 \leq j \leq d$ and K being a convex body.

Proof of Proposition 2.6. First, we shall prove (2.10) via a case-by-case analysis. Since $k=1$ we need to consider only three cases as isolated simplex counts for 'down-connectivity' in both the Vietoris–Rips and Čech complexes are identical for $k=1$. We shall write $J_n^{\mathcal{R},\mathcal{U}}$ for $J_{n,1}^{\mathcal{R},\mathcal{U}}$ and $r_n(c_n)$ for $r_{n,1}^{\mathcal{R},\mathcal{U}}(c_{n,1}^{\mathcal{R},\mathcal{U}})$ etc.

Case 1. We first consider the case $p=\mathcal{R}$, $q=\mathcal{U}$. Since $k=1$, the function $Q^{\mathcal{R},\mathcal{U}}(O,y)$ in (3.3) equals $|B_O(2)\cap By(2)|$, where $y\in A=B_O(2)$. Substituting in (3.3) and changing to polar coordinates we get

$$\mathbb{E}\big[J_n^{\mathcal{R},\mathcal{U}}(r)\big]=n\big(nr^d\big)\theta_d\int_0^2 s^{d-1}e^{-nr^d|B_O(2)\cap B_{se_1}(2)|}\,ds, \tag{3.8}$$

where $e_1=(1,0,\dots,0)\in\mathbb{R}^d$ is the unit vector along the first coordinate axis. The volume of the lens of intersection $B_O(2)\cap B_{se_1}(2)$ equals $2^d\theta_d\eta(s)$ (see [23, (7.5)] and [39, (6)]) where

$$\eta(s)=1-\frac{\theta_{d-1}}{\theta_d}\int_0^{s/2}\left(1-\frac{t^2}{4}\right)^{\frac{d-1}{2}},\quad 0\le s\le 2. \tag{3.9}$$

Substituting from (3.9) in (3.8) we get

$$\mathbb{E}\big[J_n^{\mathcal{R},\mathcal{U}}(r)\big]=n\big(nr^d\big)\theta_d e^{-nr^d\theta_d 2^d\eta(2)}\int_0^2 s^{d-1}e^{-nr^d\theta_d 2^d(\eta(s)-\eta(2))}\,ds. \tag{3.10}$$

For $0\le t\le 1$ we have $\frac{3}{4}\le(1-\frac{t^2}{4})\le 1$ and hence

$$\frac{\theta_{d-1}}{\theta_d}\left(\frac{3}{4}\right)^{\frac{d-1}{2}}\left(1-\frac{s}{2}\right)\le\eta(s)-\eta(2)\le\frac{\theta_{d-1}}{\theta_d}\left(1-\frac{s}{2}\right). \tag{3.11}$$

Using the lower bound for $(\eta(s)-\eta(2))$ from (3.11) in (3.10) and noting that $m=m^{\mathcal{R},\mathcal{U}}=\theta_d 2^d\eta(2)$, we obtain

$$\mathbb{E}\big[J_n^{\mathcal{R},\mathcal{U}}(r)\big]\le n\big(nr^d\big)\theta_d e^{-nr^d m}\int_0^2 s^{d-1}e^{-nr^d 2^d\theta_{d-1}(\frac{3}{4})^{\frac{d-1}{2}}(1-\frac{s}{2})}\,ds. \tag{3.12}$$

Let $a=2^d(\frac{3}{4})^{\frac{d-1}{2}}\theta_{d-1}$. Making the change of variable $u=nr^d2^d\theta_{d-1}(\frac{3}{4})^{\frac{d-1}{2}}(1-\frac{s}{2})$ and replacing r by r_n we get

$$\mathbb{E}\big[J_n^{\mathcal{R},\mathcal{U}}(r_n)\big]\le C_2 n e^{-nr_n^d m}\int_0^{anr_n^d}\left(1-\frac{u}{anr_n^d}\right)^{d-1}e^{-u}\,du.$$

If $nr_n^d\to\infty$ as $n\to\infty$, then

$$\int_0^{anr_n^d}\left(1-\frac{u}{anr_n^d}\right)^{d-1}e^{-u}\,du\to 1,$$

and hence

$$\mathbb{E}\big[J_n^{\mathcal{R},\mathcal{U}}(r_n)\big]\le e^{-nr_n^d m+\log n+C_3}. \tag{3.13}$$

Since $\mathbb{E}[J_n^{\mathcal{R},\mathcal{U}}(r_n(c_n))]\to e^{-\alpha}\in(0,\infty)$, we must have

$$n\big(r_n(c_n)\big)^d m-\log n\le C_4.$$

Similarly using the upper bound for $(\eta(s)-\eta(2))$ from (3.11) in (3.10) and proceeding as above will yield

$$\mathbb{E}\big[J_n^{\mathcal{R},\mathcal{U}}(r_n)\big]\ge e^{-nr_n^d m+\log n+C_5}, \tag{3.14}$$

and again using the fact that $\mathbb{E}[J_n^{\mathcal{R},\mathcal{U}}(r_n(c_n))]\to e^{-\alpha}$, we obtain

$$n\big(r_n(c_n)\big)^d m-\log n\ge C_6.$$

Case 2. Let $p = \mathcal{C}$, $q = \mathcal{U}$. For $k = 1$, $Q^{\mathcal{C},\mathcal{U}}(O, y) = V_d((B_O(1) \cap B_y(1)) \oplus B_O(1))$ where V_d denotes the volume in $\mathbb{R}^d$. Substituting in (3.3) and changing to polar coordinates we obtain

$$\mathbb{E}\big[J_n^{\mathcal{C},\mathcal{U}}(r)\big] = n\big(nr^d\big)\theta_d \int_0^2 s^{d-1} e^{-nr^d |V_d((B_O(1)\cap B_{se_1}(1))\oplus B_O(1))|}\, ds. \tag{3.15}$$

By the Steiner's formula (3.7), we have that

$$V_d\big(\big(B_O(1) \cap B_{se_1}(1)\big) \oplus B_O(1)\big) = \theta_d + \sum_{j=1}^{d} c_{j,d} V_j\big(B_O(1) \cap B_{se_1}(1)\big). \tag{3.16}$$

For $0 \le s \le 2$, the lens $B_O(1) \cap B_{se_1}(1)$ contains the line segment $\ell(s)$ joining the points $\frac{s}{2}e_1 - \sqrt{1 - (\frac{s}{2})^2}e_2$ and $\frac{s}{2}e_1 + \sqrt{1 - (\frac{s}{2})^2}e_2$. To see this, consider the projection of the balls $B_O(1)$, $B_{se_1}(1)$ on the coordinate plane determined by the first two coordinates. Hence,

$$V_1\big(B_O(1) \cap B_{se_1}(1)\big) \ge V_1\big(\ell(s)\big) = \sqrt{(2-s)(2+s)} \ge \sqrt{2(2-s)}. \tag{3.17}$$

Using (3.16) in (3.15) and the lower bound from (3.17), we obtain

$$\mathbb{E}\big[J_n^{\mathcal{C},\mathcal{U}}(r)\big] \le n\big(nr^d\big)\theta_d 2^{d-1} e^{-nr^d\theta_d} \int_0^2 e^{-nr^d c_{1,d}\sqrt{2(2-s)}}\, ds. \tag{3.18}$$

Making the change of variables $u = nr^d c_{1,d}\sqrt{2(2-s)}$, we obtain

$$\mathbb{E}\big[J_n^{\mathcal{C},\mathcal{U}}(r)\big] \le C_1 \frac{1}{r^d} e^{-nr^d\theta_d}. \tag{3.19}$$

(3.19) along with the fact that $\mathbb{E}[J_n^{\mathcal{C},\mathcal{U}}(r_n(c_n))] \to e^{-\alpha} \in (0, \infty)$ implies that

$$nr_n^d\theta_d + \log r_n^d \le C_2. \tag{3.20}$$

Substituting for r_n from (2.7) with $k = 1$ in the above inequality with $C_\alpha = \log|A| + \alpha - \log m - \log 2$ we obtain

$$nr_n^d\theta_d + \log\left(\frac{\log n + \log\log n + C_\alpha}{nc_n}\right) \le C_2.$$

Adding and subtracting $\log\log n$ in the above expression, we obtain

$$nr_n^d\theta_d - \log n + \log\log n \le C_2 - \log\left(\frac{\log n + \log\log n + C_\alpha}{\log n}\right) + \log c_n \le C_3 < \infty,$$

since c_n is bounded.

To obtain the bound in the other direction, note that the lens $B_O(1) \cap B_{se_1}(1)$, $0 \le s \le 2$, is contained in a ball of radius $\frac{\sqrt{4-s^2}}{2}$ centered at $\frac{s}{2}e_1$. Hence

$$V_j\big(B_O(1) \cap B_{se_1}(1)\big) \le \left(\frac{\sqrt{4-s^2}}{2}\right)^j V_j\big(B_O(1)\big) \le (2-s)^{\frac{j}{2}}\, V_j\big(B_O(1)\big). \tag{3.21}$$

Substituting from (3.16) in (3.15) and then using the upper bound from (3.21) yields

$$\mathbb{E}\big[J_n^{\mathcal{C},\mathcal{U}}(r)\big] \ge n\theta_d\big(nr^d\big)e^{n\theta_d r^d} \int_0^2 s^{d-1} e^{-nr^d \sum_{j=1}^d c_{j,d}(2-s)^{\frac{j}{2}} V_j(B_O(1))}\, ds. \tag{3.22}$$

Changing variables to $u = c_{1,d}(2-s)^{\frac{1}{2}} V_1(B_O(1))nr^d = C_1 nr^d(2-s)^{\frac{1}{2}}$ we obtain

$$\begin{aligned}\mathbb{E}\big[J_n^{\mathcal{C},\mathcal{U}}(r_n)\big] &\ge C_2 \frac{n(nr_n^d)}{(nr_n^d)^2} e^{-n\theta_d r_n^d} \int_0^{\sqrt{2}C_1 nr_n^d} \left(2 - \left(\frac{u}{C_1 nr_n^d}\right)^2\right)^{d-1} u e^{-(u+\sum_{j=2}^d c_{j,d}(\frac{u}{C_1 nr_n^d})^j V_j(B_O(1))nr_n^d)}\, du \\ &\ge C_3 \frac{1}{r_n^d} e^{-n\theta_d r_n^d},\end{aligned} \tag{3.23}$$

where the last inequality holds for all n sufficiently large provided $nr_n^d \to \infty$ and by the dominated convergence theorem. From (3.23) and the fact that $\mathbb{E}[J_n^{\mathcal{C},\mathcal{U}}(r_n(c_n))] \to e^{-\alpha}$ we get

$$nr_n^d\theta_d + \log r_n^d \geq C_4.$$

Comparing with (3.20) we observe that the inequality is reversed and we have a different constant. Thus, the lower bound is obtained by a computation similar to the one following (3.20).

Case 3. Finally we consider the case $p \in \mathcal{I}_1$, $q = \mathcal{D}$. Computations here are similar to those in the case $p = \mathcal{R}$, $q = \mathcal{U}$ and so we will skip some of the details. We start by observing that for $k = 1$, $Q^{p,\mathcal{D}}(O, y) = V_d(B_O(2) \cup B_y(2))$, $y \in B_O(2)$ and $m^{p,\mathcal{D}} = 2^d\theta_d$. For y of the form se_1, $0 \leq s \leq 2$, we have the bounds

$$B_O(1) \cup B_{(2+\frac{s}{2})e_1}\left(\frac{s}{2}\right) \subset B_O(2) \cup B_{se_1}(2) \subset B_{\frac{s}{2}e_1}\left(2 + \frac{s}{2}\right). \tag{3.24}$$

Since $B_O(2) \cap B_{(2+\frac{s}{2})e_1}(\frac{s}{2}) = \varnothing$, the inclusion on the left in (3.24) implies the following inequality.

$$V_d\big(B_O(2) \cup B_{se_1}(2)\big) \geq V_d\left(B_O(2) \cup B_{(2+\frac{s}{2})}\left(\frac{s}{2}\right)\right) = V_d\big(B_O(2)\big) + \left(\frac{s}{2}\right)^d V_d\big(B_O(1)\big). \tag{3.25}$$

Changing to polar coordinates in (3.3) and using (3.25) we get,

$$\mathbb{E}\big[J_n^{p,\mathcal{D}}(r)\big] \leq n\big(nr^d\big)\theta_d e^{-nr^d 2^d\theta_d} \int_0^2 s^{d-1} e^{-n\theta_d r^d (\frac{s}{2})^d}\, ds. \tag{3.26}$$

Making the change of variable $u = n\theta_d r^d(\frac{s}{2})^d$ in (3.26) and simplifying as we did in the first two cases, we get

$$\mathbb{E}\big[J_n^{p,\mathcal{D}}(r)\big] \leq C_1 n e^{-nr^d 2^d\theta_d} \tag{3.27}$$

The rest of the proof is by now a standard computation as in *Case 1* (see computation following (3.13)). For a bound in the other direction we use the right hand inclusion in (3.24) to write

$$\mathbb{E}\big[J_n^{p,\mathcal{D}}(r)\big] \geq n\big(nr^d\big)\theta_d e^{-nr^d 2^d\theta_d} \int_0^2 s^{d-1} e^{-nr^d\theta_d((2+\frac{s}{2})^d - 2^d)}\, ds. \tag{3.28}$$

Now using the binomial expansion for $((2 + \frac{s}{2})^d - 2^d)$ and making the change of variable $u = \frac{nr^d\theta_d s^d}{2^d}$, we obtain

$$\mathbb{E}\big[J_n^{p,\mathcal{D}}(r)\big] \geq C_1 n e^{-nr^d 2^d\theta_d} \int_0^{nr^d\theta_d} e^{-u - \sum_{j=1}^{d-1} c_{j,d}(\frac{u}{nr^d})^{\frac{j}{d}}}\, du.$$

Proceeding as in the proof for the lower bound (3.23), we derive that

$$\mathbb{E}\big[J_n^{p,\mathcal{D}}(r_n)\big] \geq C_2 n e^{-n2^d\theta_d r_n^d}, \tag{3.29}$$

where the last inequality holds for all n sufficiently large provided $nr_n^d \to \infty$. This completes the proof of (2.10).

We now prove (2.11). We have from (2.10) that

$$C_1 \leq n\big(r_n(c_n)\big)^d m - \log n - (1-a)\log\log n \leq C_2, \tag{3.30}$$

for some finite constants C_1, C_2. Substituting from (2.7) in (3.30) with $C_\alpha := \log|A| + \alpha - \log m - \log 2$, we obtain

$$C_1 \leq (\log n + \log\log n + C_\alpha)\frac{m}{c_n} - \log n - (1-a)\log\log n \leq C_2,$$

which simplifies to

$$C_1 \leq \left(\frac{m}{c_n} - 1\right)\log n + \left(\frac{m}{c_n} - (1-a)\right)\log\log n + \frac{m}{c_n}C_\alpha \leq C_2.$$

The result now follows since $\frac{m}{c_n} \to 1$ as $n \to \infty$ by Proposition 2.4. This completes the proof of Proposition 2.6. $\square$

3.2. *Proof of result in Section 2.3*

Proof of Proposition 2.7. Let us fix $k \geq 1$ and $q \in \mathcal{I}_2$. We first consider the case $p = \mathcal{R}$. So, we shall again drop these subscripts and superscripts for rest of the calculation. Further set $\hat{J}_n(r) = J_n^*(r) - J_n(r)$. Given a $r > 0$, using (2.12) and Campbell–Mecke formula, we derive an upperbound for $\mathbb{E}[\hat{J}_n(r)]$.

$$\begin{aligned}
\mathbb{E}\big[\hat{J}_n(r)\big] &= \frac{n^{k+1}}{(k+1)!}\int_{U^{k+1}} \mathbf{1}\big[R(\mathbf{x}) > r\big] e^{-n|Q(\mathbf{x},R(\mathbf{x}))|}\,d\mathbf{x} \\
&= \frac{n^{k+1}}{(k+1)!}\int_{U^{k+1}} \sum_{i,j=1,i\neq j}^{k+1} \mathbf{1}\big[R(\mathbf{x}) > r, 2R(\mathbf{x}) = |x_i - x_j|\big] \times e^{-n|Q(\mathbf{x},|x_i-x_j|/2)|}\,d\mathbf{x} \\
&= \frac{n^{k+1}}{2(k-1)!}\int_{U^{k+1}} \mathbf{1}\big[2R(\mathbf{x}) = |x_1 - x_2|\big]\mathbf{1}\big[|x_1 - x_2| > 2r\big] \times e^{-n|Q(\mathbf{x},|x_1-x_2|/2)|}\,d\mathbf{x} \\
\big(\mathbf{x} \to \mathbf{x} - (x_1, \dots, x_1)\big) &\leq \frac{n^{k+1}}{2(k-1)!}\int_{U^{k}} \mathbf{1}\big[2R(O,\mathbf{x}) = |x_2|\big]\mathbf{1}\big[|x_2| > 2r\big] e^{-n|Q((O,\mathbf{x}),|x_2|/2)|}\,d\mathbf{x},
\end{aligned}$$

where $\mathbf{x} = (x_2, \dots, x_{k+1})$ in the final expression. Changing the variable $\mathbf{x} \to r\mathbf{x}$ yields

$$\mathbb{E}\big[\hat{J}_n(r)\big] \leq \frac{n(nr^d)^k}{2(k-1)!}\int_{(\mathbb{R}^d)^k} \mathbf{1}\big[2R(O,\mathbf{x}) = |x_2|, |x_2| > 2\big] e^{-nr^d|Q((O,\mathbf{x}),|x_2|/2)|}\,d\mathbf{x}.$$

Changing the variable $x_2/2$ to polar co-ordinates and then $\mathbf{x} \to s\mathbf{x}$, we obtain (with $\mathbf{x} = (x_3, \dots, x_{k+1})$)

$$\begin{aligned}
\mathbb{E}\big[\hat{J}_n(r)\big] &\leq \frac{n2^d\theta_d(nr^d)^k}{2(k-1)!}\int_1^\infty s^{d-1}\,ds \int_{B_O(2s)^{k-1}} 1\big[R\big((O, 2se_1, \mathbf{x})\big) = s\big] \\
&\quad \times e^{-nr^d|Q((O,2se_1,\mathbf{x}),s))|}\,d\mathbf{x} \\
&= \frac{n2^d\theta_d(nr^d)^k}{2(k-1)!}\int_1^\infty s^{dk-1}\,ds \int_{B_O(2)^{k-1}} 1\big[R\big((O, 2e_1, \mathbf{x})\big) = 1\big] \\
&\quad \times e^{-nr^d s^d|Q((O,2e_1,\mathbf{x}),1))|}\,d\mathbf{x}.
\end{aligned}$$

Since $Q(\cdot) \geq m$ for $R(\cdot) = 1$, we have

$$\mathbb{E}\big[\hat{J}_n(r)\big] \leq C_1 n\big(n\theta_d 2^d r^d\big)^k \int_1^\infty s^{dk-1} e^{-nr^d s^d m}\,ds.$$

Making the change of variables $t = nr^d s^d m$, we obtain

$$\begin{aligned}
\mathbb{E}\big[\hat{J}_n(r)\big] &\leq C_2 \times n \int_{nr^d m}^\infty t^{k-1} e^{-t}\,dt \\
&\leq C_3 \times n e^{-nr^d m} \sum_{j=0}^{k-1} \frac{(nr^d m)^j}{j!},
\end{aligned}$$

where in the last inequality we have used integral formulas for the upper gamma function. A simple substitution now yields that if $nmr_n^d = \log n + k\log\log n + w_n^1$ for some sequence w_n^1 bounded from below, we have that

$$\mathbb{E}\big[\hat{J}_n(r_n)\big] \leq C_4 e^{-w_n^1 - \log\log n} \to 0.$$

From Proposition 2.5, we know that if $nmr_n^d = \log n + k\log\log n + w_n^1$ for some sequence w_n^1 bounded from below, then

$$\mathbb{E}\big[J_n(r_n)\big] \to 0.$$

Thus, if $nmr_n^d = \log n + k \log\log n + w_n^1$ for some sequence w_n^1 bounded from below, then

$$\mathbb{E}\big[J_{n,k}^{\mathcal{R},q,*}(r_n)\big] = \mathbb{E}\big[J_n(r_n)\big] + \mathbb{E}\big[\hat{J}_n(r_n)\big] \to 0,$$

for any $q \in \mathcal{I}_2$.

We now consider the Čech case i.e., $p = \mathcal{C}$. In this case, the computation is a little more involved but more along the lines of that for critical points in the proof of [13, Proposition 6.1]. Define $\hat{J}_n^{\mathcal{C}}(r) = J_n^{\mathcal{C},q,*}(r) - J_n^{\mathcal{C},q}(r)$. By definition of the Čech complex, $R^{\mathcal{C}}(x_0,\dots,x_k) = \inf\{r : \cap_{i=0}^k B_r(x_i) \neq \varnothing\}$ and further we have that $\{C(x_0,\dots,x_k)\} = \bigcap_{i=0}^k B_{R^{\mathcal{C}}(x_0,\dots,x_k)}(x_i)$ for some point $C(x_0,\dots,x_k) \in \mathbb{R}^d$. Now using this observation and proceeding as in the Vietoris–Rips complex case using translation and scaling we have that (dropping the superscripts $\mathcal{C}$, q as usual)

$$\mathbb{E}\big[\hat{J}_n(r)\big] \le \frac{n^{k+1}r^{dk}}{(k+1)!}\int_{(\mathbb{R}^d)^k} \mathbf{1}\big[R(O,\mathbf{x}) > 1\big] e^{-nr^d|Q((O,\mathbf{x}),R(O,\mathbf{x}))|}\,d\mathbf{x}$$

The RHS in the above equation is exactly of the form [13, (8.8)] with $h_1(O,\mathbf{x})$ there replaced by $\mathbf{1}[R(O,\mathbf{x}) > 1]$ and $\theta_d R(O,\mathbf{x})^d$ replaced by $|Q((O,\mathbf{x}),R(O,\mathbf{x}))|$ in the exponent. Observe that both $|Q((O,\mathbf{x}),R(O,\mathbf{x}))|$ and $R(O,\mathbf{x})$ are rotation invariant and also $|Q((O,s\mathbf{x}),R(O,s\mathbf{x}))| = s^d|Q((O,\mathbf{x}),R(O,\mathbf{x}))|$ for any $s > 0$. So, we can now follow the derivations in [13, (8.8)–(8.10)] and using the bound that $Q((O,\mathbf{x}),1) \ge m$ derive that

$$\mathbb{E}\big[\hat{J}_n(r)\big] \le C_1 n\big(nr^d\big)^k \int_1^\infty s^{dk-1} e^{-nr^d s^d m}\,ds.$$

The above integral can be simplified and evaluated as in the Vietoris–Rips case above to obtain that

$$\mathbb{E}\big[\hat{J}_n(r)\big] \le C_2 \times n e^{-nr^d m} \sum_{j=0}^{k-1} \frac{(nr^d m)^j}{j!}.$$

Thus again combining with the Proposition 2.5, we have that if $nmr_n^d = \log n + k\log\log n + w_n^1$ for some sequence w_n^1 bounded from below, then $\mathbb{E}[J_{n,k}^{\mathcal{C},q,*}(r_n)] \to 0$ for any $q \in \mathcal{I}_2$. □

3.3. *Proofs of results in Section 1.2*

Proof of Theorem 1.1. Fix $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$ and we shall drop the superscripts p, q in the rest of the proof. Let $nmr_n^d = (1+\epsilon)\log n$. Substituting in (3.3) and observing that $|Q(O,y)| \ge m$ and A is bounded, we obtain

$$\begin{aligned}\mathbb{E}\big[J_n(r_n)\big] &\le C_1 n\big(nr_n^d\big)^k e^{-nr_n^d m}\\ &\le C_2 \frac{(\log n)^k}{n^\epsilon}.\end{aligned} \tag{3.31}$$

By the Markov's inequality and the bound obtained in (3.31), we have

$$\mathbb{P}\big\{J_n(r_n) \ge 1\big\} \le C_2 \frac{(\log n)^k}{n^\epsilon} \to 0,$$

as $n \to \infty$. This proves the second assertion in (1.4).

Let $nmr_n^d = (1-\epsilon)\log n$. To prove the first assertion we use the second moment approach. Since

$$\mathbb{P}\big\{J_n(r_n) \ge 1\big\} \ge \frac{(\mathbb{E}[J_n(r_n)])^2}{\mathbb{E}[J_n(r_n)^2]},$$

to prove the first assertion it suffices to show that

$$\frac{(\mathbb{E}[J_n(r_n)])^2}{\mathbb{E}[J_n(r_n)^2]} \to 1, \quad \text{as } n \to \infty. \tag{3.32}$$

To this end, we evaluate $\mathbb{E}[J_n^2]$. From (2.6) we can write

$$J_n(r_n)^2 = C_1 \sum_{\mathbf{x},\mathbf{y}\in\mathcal{P}_n^{k+1}} h(\mathbf{x}, r_n)h(\mathbf{y}, r_n)\mathbf{1}\big[\mathcal{P}_n\big(Q(\mathbf{x}, r_n) \cup Q(\mathbf{y}, r_n)\big) = 0\big]$$

$$= \sum_{j=0}^{k+1} J_n^{(j)}, \tag{3.33}$$

where

$$J_n^{(j)} = C_1 \sum_{\substack{\mathbf{x},\mathbf{y}\in\mathcal{P}_n^{k+1},\\ |\mathbf{x}\cap\mathbf{y}|=j}} h(\mathbf{x}, r_n)h(\mathbf{y}, r_n)\mathbf{1}\big[\mathcal{P}_n\big(Q(\mathbf{x}, r_n) \cup Q(\mathbf{y}, r_n)\big) = 0\big], \quad j = 0, 1, 2, \dots, (k+1), \tag{3.34}$$

is the contribution to J_n^2 when the two complexes share j vertices. For $j = 1, \dots, (k+1)$, we have by the Campbell–Mecke formula

$$\mathbb{E}\big[J_n^{(j)}\big] \le C_2 n^{2k+2-j} \int_{\mathbf{x}\in U^{(k+1)}} \int_{\mathbf{z}\in U^{(k+1-j)}} h(\mathbf{x}, r_n)h(\mathbf{y}, r_n)e^{-|Q(\mathbf{x},r_n)\cup Q(\mathbf{y},r_n)|}\, d\mathbf{x}\, d\mathbf{z},$$

where $\mathbf{y} = (x_1, \dots, x_j, \mathbf{z})$, $\mathbf{z} = (z_1, \dots, z_{k+1-j})$. In the last inequality, $h(\mathbf{y}, r_n) = 0$ if any of the variables z_i, $i = 1, \dots, (k+1-j)$ lies outside a ball of radius $6kr_n$ from x_1 and hence, we derive that

$$\mathbb{E}\big[J_n^{(j)}\big] \le C_3\big(nr_n^d\big)^{k+1-j} n^{k+1} \int_{\mathbf{x}\in U^{(k+1)}} h(\mathbf{x}, r_n)e^{-|Q(\mathbf{x},r_n)|}\, d\mathbf{x}. \tag{3.35}$$

Comparing the right hand side of (3.35) with (3.1) and using the definition of r_n we obtain

$$\mathbb{E}\big[J_n^{(j)}\big] \le C_4(\log n)^{k+1-j}\mathbb{E}[J_n], \quad j = 1, 2, \dots (k+1). \tag{3.36}$$

Now consider $\mathbb{E}[J_n^{(0)}]$. By the Campbell–Mecke formula we have

$$\mathbb{E}\big[J_n^{(0)}\big] = \big[(k+1)!\big]^{-2} n^{2k+2} \int_{\mathbf{x}\in U^{(k+1)}} \int_{\mathbf{z}\in U^{(k+1)}} h(\mathbf{x}, r_n)h(\mathbf{z}, r_n)e^{-|Q(\mathbf{x},r_n)\cup Q(\mathbf{z},r_n)|}\, d\mathbf{x}\, d\mathbf{z}. \tag{3.37}$$

Divide the inner integral in (3.37) into two parts, one over the region where $\min_{1\le r,s\le(k+1)} |x_r - z_s| \le 6kr_n$ and the second its complement. Over the first region we proceed as in (3.35), (3.36) to obtain the bound

$$n^{2k+2} \int_{\mathbf{x}\in U^{(k+1)}} \int_{\mathbf{z}\in U^{(k+1)}} \mathbf{1}\Big[\min_{1\le r,s\le(k+1)} |x_r - z_s| \le 6kr_n\Big] h(\mathbf{x}, r_n)h(\mathbf{z}, r_n)e^{-|Q(\mathbf{x},r_n)|}\, d\mathbf{x}\, d\mathbf{z}$$

$$\le C_5(\log n)^{k+1}\mathbb{E}[J_n]. \tag{3.38}$$

Over the region where $\min_{1\le r,s\le(k+1)} |x_r - z_s| > 6kr_n$, we have $|Q(\mathbf{x}, r_n) \cup Q(\mathbf{z}, r_n)| = |Q(\mathbf{x}, r_n)| + |Q(\mathbf{z}, r_n)|$ which yields the bound

$$\big[(k+1)!\big]^{-2} n^{2k+2} \int_{\mathbf{x}\in U^{(k+1)}} \int_{\mathbf{z}\in U^{(k+1)}} \mathbf{1}\Big[\min_{1\le r,s\le(k+1)} |x_r - z_s| > 6kr_n\Big] h(\mathbf{x}, r_n)h(\mathbf{z}, r_n)e^{-|Q(\mathbf{x},r_n)|+|Q(\mathbf{z},r_n)|}\, d\mathbf{x}\, d\mathbf{z}$$

$$\le \big(\mathbb{E}[J_n]\big)^2. \tag{3.39}$$

From (3.33), (3.36)–(3.39) we obtain

$$\mathbb{E}\big[J_n^2\big] \le C_7(\log n)^{k+1}\mathbb{E}[J_n] + \big(\mathbb{E}[J_n]\big)^2. \tag{3.40}$$

Choose $\delta > 0$ sufficiently small so that $\frac{(m+\delta)(1-\epsilon)}{m} = 1 - \frac{\epsilon}{2}$. With this choice of δ, substituting for r_n in (3.3) we obtain

$$\mathbb{E}[J_n] \ge \frac{n(nr_n^d)^k}{(k+1)!} \int_A \mathbf{1}\big[\big|Q(O, \mathbf{y})\big| \le m + \delta\big] e^{-nr_n^d|Q(O,\mathbf{y})|}\, d\mathbf{y}$$

$$\ge C_8 n(\log n)^k e^{-\frac{(m+\delta)(1-\epsilon)\log n}{m}}$$

$$= C_8 n^{\frac{\epsilon}{2}} (\log n)^k. \tag{3.41}$$

It now follows from (3.40) and (3.41) that

$$\liminf_{n\to\infty} \frac{(\mathbb{E}[J_n])^2}{\mathbb{E}[J_n^2]} \geq 1.$$

This proves (3.32) and hence the first assertion in (1.4). For any $\delta \in (0,1)$, we have by Chebyshev's inequality

$$P\big(J_n \geq \mathbb{E}[J_n]^{1+\delta}\big) \leq \frac{\mathbb{E}[J_n^2]}{(\mathbb{E}[J_n])^{2(1+\delta)}} \to 0,$$

where the convergence follows from (3.41) and (3.32). This proves

$$P\left(\frac{\log J_n}{\log \mathbb{E}[J_n]} \leq 1+\delta\right) \to 1.$$

Now, we prove the converse bound. From (3.41), (3.32), and the Paley–Zygmund inequality we have for any sequence $\theta_n \in (0,1)$ such that $\theta_n \to 0$ and $\theta_n \mathbb{E}[J_n] \to \infty$,

$$P\big(J_n \geq \theta_n \mathbb{E}[J_n]\big) \geq (1-\theta_n)^2 \frac{(\mathbb{E}[J_n(r_n)])^2}{\mathbb{E}[J_n(r_n)^2]} \to 1, \quad \text{as } n \to \infty.$$

In particular, taking $\theta_n = (\mathbb{E}[J_n])^{-\delta}$ for some $\delta \in (0,1)$ in the above inequality and using (3.41) we obtain

$$P\left(\frac{\log J_n}{\log \mathbb{E}[J_n]} \geq 1-\delta\right) \to 1.$$

Thus, we derive that $\frac{\log J_n}{\log \mathbb{E}[J_n]} \xrightarrow{P} 1$. □

Proof of Theorem 1.2. The first statement in (1.5) follows trivially from the corresponding statement in (1.4) and the second statement now follows from Proposition 2.7 and Markov's inequality. □

Proof of Theorem 1.3. Fix $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$, $k \geq 1$ and $L \geq 1$. We shall now onwards drop superscripts p, q in the rest of the proof except to avoid ambiguity. For $M \geq 1$ let $\Delta(L, M)$ denote the set of feasible (up/down)-connected graphs Γ formed by L k-faces of the Čech or Vietoris–Rips complex such that there are a total of M vertices in the k-faces and each of the M vertices is present in at least one of the k-faces. More precisely, $\Gamma \in \Delta(L, M)$ if there exists $\{x_1, \dots, x_M\} \subset \mathbb{R}^d$ with $S_k(\{x_1, \dots, x_M\}, 1) = L$, $G_k(\{x_1, \dots, x_M\}, 1) \cong \Gamma$ and further each x_i, $1 \leq i \leq M$ belongs to at least one k-face. Note that it is possible that $\Delta(L, M) = \varnothing$ for certain choices of M and L either due to the combinatorics or the geometry. Trivially, $\Delta(L, M) = \varnothing$ for $M > L(k+1)$ and $M \leq k$. Hence setting $\Delta(L) = \bigcup_{M=k+1}^{L(k+1)} \Delta(L, M)$, we see that $\Delta(L)$ is the set of all feasible (up/down)-connected graphs that can be formed on L faces. Note that both $\Delta(L, M)$ and $\Delta(L)$ depend on p and q but we omit the same. Thus, we have that

$$J_n(r, L) = \sum_{M=k+1}^{L(k+1)} \sum_{\Gamma \in \Delta(L,M)} \bar{J}_n(r, \Gamma), \tag{3.42}$$

where $\bar{J}_n(r, \Gamma)$ is the number of induced Γ components in $G_k(\mathcal{P}_n, r)$ formed by M vertices i.e.,

$$\bar{J}_n(r, \Gamma) := \sum_{\{X_1,\dots,X_M\} \subset \mathcal{P}_n} \mathbf{1}\big[G_k\big(\{X_1, \dots, X_M\}, r\big) \cong \Gamma\big] \mathbf{1}\left[\mathcal{P}_n\left(\bigcup_{i=1}^{L} Q\big(\big(\mathbf{X}^i\big), r\big)\right) = \{X_1, \dots, X_M\}\right], \tag{3.43}$$

where $\mathbf{X}^i$ is a $(k+1)$-subset of $\{X_1, \dots, X_M\}$ such that the $\mathbf{X}^i, i = 1, \dots, L$ are the vertices in $G_k(\{X_1, \dots, X_M\}, r)$ i.e., the k-faces in the corresponding geometric complex. Since $\Delta(L)$ is a finite set, it is enough if we show that for all $\Gamma \in \Delta(L, M)$,

$$\mathbb{E}\big[\bar{J}_n(r_n, \Gamma)\big] \to 0$$

for r_n such that $nmr_n^d = (1+\epsilon)\log n$ for any $\epsilon > 0$.

Set $h_{\Gamma,r} := \mathbf{1}[G_k(\{x_1, \dots, x_M\}, r) \cong \Gamma]$ and $h_\Gamma := h_{\Gamma,1}$. Since Γ is connected, we note that there exists a $K > 0$ (possibly depending on M, L) such that $h_\Gamma(O, x_2, \dots, x_M) = 0$ if $\max_{i=2,\dots,M} |x_i| > K$.

Further, whenever $G_k(\{x_1, \dots, x_M\}, r) \cong \Gamma$, we denote the L vertices (i.e., k-faces) by $\mathbf{x}^1, \dots, \mathbf{x}^L$. Let $r > 0$. As usual, we start with the Campbell–Mecke formula and then use translation and scaling relations in the below derivation:

$$\mathbb{E}\big[\bar{J}_n(r, \Gamma)\big] = \frac{n^M}{M!} \int_{U^M} h_{\Gamma,r}(x_1, \dots, x_M) e^{-n|\cup_{i=0}^{L} Q(\mathbf{x}^i, r)|}\, dx_1 \dots dx_M$$

$$(\text{change } x_i \to x_i + r x_1, i \ge 1) \le \frac{n(nr^d)^{M-1}}{M!} \int_{(\mathbb{R}^d)^{M-1}} h_\Gamma(O, x_2, \dots, x_m) e^{-nr^d |\cup_{i=0}^{L} Q(\mathbf{x}^i)|}\, dx_2 \dots dx_M$$

$$\left(\text{by } \left|\bigcup_{i=0}^{L} Q(\mathbf{x}^i)\right| \ge m\right) \le \frac{n(n\theta_d K^d r^d)^{M-1}}{M!} e^{-nr^d m}.$$

Now choosing r_n such that $nmr_n^d = (1+\epsilon)\log n$ for an $\epsilon > 0$, we have using the above bound that

$$\mathbb{E}\big[\bar{J}_n(r_n, \Gamma)\big] \le \left(\frac{\theta_d K^d (1+\epsilon)^d}{m}\right)^{M-1} n^{-\epsilon} (\log n)^{M-1} \to 0.$$

□

Proof of Proposition 1.4. The results are a straightforward consequence of the inequalities (3.13), (3.14), (3.27), (3.29), (3.19) and (3.23) obtained in the proof of Theorem 2.6. □

Proof of Proposition 1.5. Fix a $p \in \mathcal{I}_1$, $q \in \mathcal{I}_2$. Recall that $\hat{J}_{n,1}(r) = J^*_{n,1}(r) - J_{n,1}(r)$. We note that $R(x_1, x_2) = |x_1 - x_2|/2$. Again, we know asymptotics of $\mathbb{E}[J_{n,1}(r)]$ from Proposition 2.6. So, we shall only derive asymptotics for $\mathbb{E}[\hat{J}_{n,1}(r)]$. Again, starting with Campbell–Mecke formula and using translation, change to polar coordinates as in the above calculations

$$\mathbb{E}\big[\hat{J}_{n,1}(r)\big] = \frac{n^2}{2} \int_{U^2} \mathbf{1}\big[R(\mathbf{x}) > r\big] e^{-n|Q(\mathbf{x}, R(\mathbf{x}))|}\, d\mathbf{x}$$

$$= \frac{n^2}{2} \int_U dx_0 \int_{U - x_0} \mathbf{1}\big[|x - x_0| > 2r\big] e^{-n|Q((x_0, x), |x|/2)|}\, dx$$

$$\big((x_0, x) \to (O, x - x_0)\big) = \frac{n^2}{2} \int_U \mathbf{1}\big[|x| > 2r\big] e^{-n|Q((O,x),|x|/2|)}\, dx$$

$$(\text{change } x/2r \text{ to polar co-ordinates}) = n^2 \theta_d 2^{d-1} r^d \int_1^\infty s^{d-1} e^{-nr^d |Q((O, 2se_1), s)|}\, ds$$

$$= n^2 \theta_d 2^{d-1} r^d \int_1^\infty s^{d-1} e^{-nr^d s^d |Q((O, 2e_1), 1)|}\, ds$$

$$(\text{by definition of } m_1) = n^2 \theta_d 2^{d-1} r^d \int_1^\infty s^{d-1} e^{-nr^d s^d m_1}\, ds$$

$$= \frac{2^{d-1}\theta_d}{dm_1} n e^{-nr^d m_1}.$$

Thus, combining with Proposition 1.4, the proof is complete. □

3.4. *Proofs of results in Section 2.4*

Theorem 2.8 is proved using the criterion derived in [49, Theorem 3.1], a simpler version of which is stated below. In order to state this Theorem, we need some notation. Let η be a finite Poisson point process in $\mathbb{R}^d$ with intensity measure μ and $\mathcal{N}$ be the space of all finite subsets of $\mathbb{R}^d$ with the sigma-algebra on $\mathcal{N}$ generated by the functions $\xi \to |\xi \cap B|$ for all bounded Borel sets $B \subset \mathbb{R}^d$. Let $k \in \mathbb{N}$ and let $f : (\mathbb{R}^d)^k \times \mathcal{N} \to \{0, 1\}$ be a measurable function. For any $\xi \in \mathcal{N}$, set

$$F(\xi) := \sum_{\psi \subset \xi : |\psi| = k} f(\psi, \xi \setminus \psi).$$

For $x_1, \dots, x_k \in \mathbb{R}^d$, set $p(x_1, \dots, x_k) = \mathbb{E}[f(\{x_1, \dots, x_k\}, \eta)]$.

Theorem 3.1 ([49, Theorem 3.1]). *Let $W = F(\eta)$ with η and F as defined above. Suppose that $w : (\mathbb{R}^d)^k \to [0,\infty)$ is a measurable function and that for μ^k-almost every $\mathbf{x} = (x_1, \dots, x_k) \in (\mathbb{R}^d)^k$ with $p(\mathbf{x}) > 0$, we can find coupled random variables $U^{\mathbf{x}}$, $V^{\mathbf{x}}$ such that*

- $U^{\mathbf{x}} \overset{d}{=} W$
- $1 + V^{\mathbf{x}} \overset{d}{=} F(\bigcup_{i=1}^k \{x_i\} \cup \eta) | f(\{x_1, \dots, x_k\}, \eta) = 1$
- $\mathbb{E}[|U^{\mathbf{x}} - V^{\mathbf{x}}|] \le w(\mathbf{x})$.

Then the total variation distance between the law of W and a Poisson random variable with mean $\mathbb{E}[W]$ satisfies

$$d_{\mathrm{TV}}\big(W, \mathrm{Poi}\big(\mathbb{E}[W]\big)\big) \le \frac{1 \wedge (\mathbb{E}[W])^{-1}}{k!} \int w(\mathbf{x}) p(\mathbf{x}) \mu^k(d\mathbf{x}).$$

The following geometric lemma is crucial in the proofs of Theorems 2.8 and 2.9. The lemma essentially says that two ‘near-minimal’ configurations of points that form ‘isolated cliques’ are well seperated from each other. The lack of such a geometric lemma hinders extending these results to the Čech complex.

Lemma 3.2. *For $\delta \ge 0$, $j \le k$, define*

$$D_{j,\delta} := \big\{(\mathbf{x}, \mathbf{z}) \in B_O(2)^k \times B_O(6)^{k-j+1} : 2 - \delta \le |x_i|, |x_i - x_j|, |z_i|, |x_i - z_j|, |z_i - z_j|,$$
$$\forall i \ne j, h\big((O, \mathbf{x})\big) = h(\mathbf{y}) = 1\big\},$$

where $\mathbf{x} = (x_2, \dots, x_{k+1}) \in \mathbb{R}^{dk}$, $\mathbf{z} = (z_1, \dots, z_{k-j+1}) \in \mathbb{R}^{d(k-j+1)}$ and $\mathbf{y} = (x_{k-j+2}, \dots, x_{k+1}, z_1, \dots, z_{k-j+1}) \in \mathbb{R}^{d(k+1)}$. Then there exists a $\delta_0 > 0$ such that for any $0 \le \delta < \delta_0$, we can find a $\beta := \beta(\delta) > 0$ for which $|Q(\mathbf{y}) \setminus Q(O, \mathbf{x})| \ge \beta$ on the set $D_{j,\delta}$.

Proof. Since the function $|Q(\mathbf{y}) \setminus Q(O, \mathbf{x})|$ is continuous in $(\mathbf{x}, \mathbf{z})$, it suffices to show the result with $\delta = 0$. Then any $(\mathbf{x}, \mathbf{z}) \in D_{j,0}$ must satisfy the following conditions. Firstly, since $h((O, \mathbf{x})) = h(\mathbf{y}) = 1$, we have that $|x_i|, |x_i - x_\ell|, |z_i - z_\ell| = 2$, $\forall i \ne \ell$ and also $|x_i - z_\ell| = 2, \forall \ell$ and $\forall i \in \{k-j+2, \dots, k+1\}$. Secondly, $|z_\ell|, |x_i - z_\ell| \ge 2, \forall \ell$ and $\forall i \in \{2, \dots, k-j+1\}$. Since $Q(O, \mathbf{x}) \subset B_O(2)$, it suffices to show $|Q(\mathbf{y}) \setminus B_O(2)| \ge \beta$ for some $\beta > 0$.

We now state two geometric claims for generic points in $\mathbb{R}^d$ which will be proven later.

Claim 1: Define $\delta_1 := \min\{\max |u_i - u_\ell| : u_i \in \mathbb{R}^d, 1 \le i, \ell \le d+2, |u_i - u_\ell| \ge 2, \ \forall i \ne \ell\} - 2$. The first claim is that $\delta_1 > 0$.

Claim 2: Let $\delta_1 > 0$ be as in Claim 1. If $u_1, \dots, u_{d+1} \in \mathbb{R}^d$ are such that $|u_i - u_\ell| = 2$, $\forall i \ne \ell$, then there exists an $u' \in \mathbb{R}^d$ such that $|u' - u_1| \ge 2 + \delta_1$ and $|u' - u_i| = 2$ for all $i \in \{2, \dots, d+1\}$.

Claim 1 is fairly intuitive in that it asserts that no collection of $(d+2)$ points in $\mathbb{R}^d$ can have all pairwise distances to be exactly two and by continuity, we have that the maximum of pairwise distances is at least $2 + \delta_1$ for some $\delta_1 > 0$. Suppose we have $(d+1)$ points in $\mathbb{R}^d$ with all pairwise distances being exactly two. Then Claim 2 asserts that there exists a point x with distance at least $2 + \delta_1$ from one point and distance exactly two from all other points. Using the above two claims, we now complete the proof.

First consider the case when $k = d$. Since $\delta = 0$, $x_{k-j+2}, \dots, x_{k+1}, z_1, \dots, z_{k-j+1}$ are at distance exactly two from each other and $|x_i| = 2$ for all $1 \le i \le k+1$. Now applying Claim 1 to $\{O, x_{k-j+2}, \dots, x_{k+1}, z_1, \dots, z_{k-j+1}\}$, we have that $|z_i| > 2 + \delta_1$ for some $i \in \{1, \dots, k-j+1\}$. Without loss of generality, let us assume that $|z_1| > 2 + \delta_1$. Hence $B_{z_1}(\delta_1/2) \cap B_O(2) = \varnothing$. Further, since $Q(\mathbf{y})$ is an intersection of at most $d+1$ balls of radius 2 whose centers (one of which is z_1) are exactly at distance two apart from each other, we have that $|B_{z_1}(\delta_1/2) \cap Q(\mathbf{y})| \ge \beta > 0$ for some $\beta > 0$. Indeed, $B_{z'}(\delta_1/8) \subset B_{z_1}(\delta_1/2) \cap Q(\mathbf{y})$ for $z' = (1 - 3\delta_1/8)z_1$. Thus, we get that $|Q(\mathbf{y}) \setminus B_O(2)| \ge \beta$ for some $\beta > 0$. We have illustrated this argument in the case $d = 2$ and $j = 1$ in Figure 5. Next, consider the case when $k < d$. Since we are interested in minimizing $|Q(\mathbf{y}) \setminus B_O(2)|$, we can assume that $|z_i| = 2$ for all $i \in \{1, \dots, k+1\}$. Further, if $k < d - 1$ choose additional points $\zeta_1, \dots, \zeta_{d-k-1}$ all on the boundary of $B_O(2)$ so that (O, ζ) forms a d-simplex with side lengths 2 where $\zeta = (x_{k-j+2}, \dots, x_{k+1}, z_1, \dots z_{k-j+1}, \zeta_1, \dots, \zeta_{d-k-1})$. When $k = d - 1$, we set $\zeta = \mathbf{y} = (x_{k-j+2}, \dots, x_{k+1}, z_1, \dots, z_{k-j+1})$ and observe that it still holds that (O, ζ) forms a d-simplex with side lengths 2. From Claim 2, we can choose $x \in \mathbb{R}^d$ such that $|x| \ge 2 + \delta_1$ and (ζ, x) forms a d-simplex with side lengths 2. Since $x \in Q(\mathbf{y})$, we can argue as in the case $k = d$ in the above paragraph that $|Q(\mathbf{y}) \setminus B_O(2)| \ge |B_x(\delta_1/2) \cap Q(\mathbf{y})| \ge \beta$ for some $\beta > 0$. We have illustrated this in Figure 6.

This completes the proof except the two claims which will be proven next.

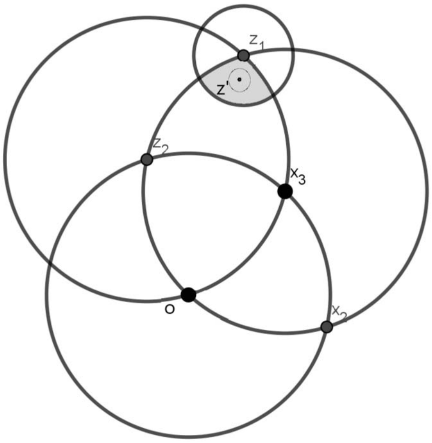

Fig. 5. Illustration of Lemma 3.2 for $d = k = 2$, $j = 1$, $\delta = 0$: Here, $\mathbf{x} = (x_2, x_3)$ and $\mathbf{y} = (x_3, z_1, z_2)$. Further, $h(O, \mathbf{x}) = h(\mathbf{y}) = 1$, $|x_2| = |x_3| = |x_3 - z_1| = |x_3 - z_2| = |z_1 - z_2| = 2$ and $|z_1|, |z_2| \geq 2$. Claim 1 in the proof below implies that $|z_1| \geq 2 + \delta_1$ for some $\delta_1 > 0$ and we have that $B_{z'}(\delta_1/8) \subset B_{z_1}(\delta_1/2) \cap Q(\mathbf{y})$ for $z' = (1 - 3\delta_1/8)z_1$.

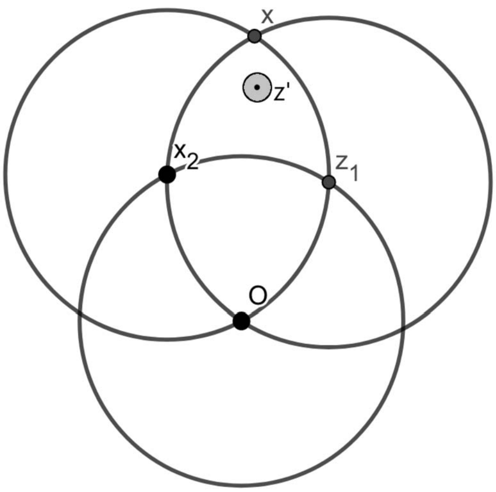

Fig. 6. Illustration of Lemma 3.2 for $d = 2$, $k = j = 1$, $\delta = 0$: Here, $\mathbf{x} = (x_2)$ and $\mathbf{y} = \zeta = (x_2, z_1)$. Further, $h(O, \mathbf{x}) = h(\mathbf{y}) = 1$, $|x_2| = |x_2 - z_1| = 2$ and $|z_1| \geq 2$. By Claim 2, there exists $|x| \geq 2 + \delta_1$ for some $\delta_1 > 0$ and consequently, we have that $B_{z'}(\delta_1/8) \subset B_{z_1}(\delta_1/2) \cap Q(\mathbf{y})$.

Proof of Claim 1: Without loss of generality, we can choose $u_1 = O$. We can further assume that $|u_i| \leq 3$ for all $i = 2, \dots, d+1$ as minimum will be attained by such a configuration of points. Since $\max_{i \neq \ell} |u_i - u_\ell|$ is a continuous function of $u_2, \dots u_{d+1}$ on $B_O(3)^d$, the minimum δ_1 is attained. If $\delta_1 = 0$, we have a contradiction that there is a configuration of $(d+2)$ points which form a $(d+1)$-simplex in $\mathbb{R}^d$ with side-lengths 2.

Proof of Claim 2: To show this, we will make a specific choice of $\mathbf{u} = (u_1, \dots, u_{d+1})$ and show existence of u. Any other choice will be a rotation and translation of this configuration. To simplify notation we relabel $\mathbf{u}$ to be $(O, u_1, \dots, u_d)$. Let e_i, $i = 1, \dots, d$ be the unit vectors along the coordinate axes. Let $\mathbf{v} = \sum_{i=1}^d e_i$ and write for $i = 1, \dots, d$, $u_i = ae_i + b\mathbf{v}$ for some $a, b \in \mathbb{R}$ to be chosen later. Since $\mathbf{u}$ forms a simplex with side lengths two, the constants a, b must satisfy the following two conditions: (i) $|u_i| = 2$, $i = 1, \dots, d$ and (ii) $|u_i - u_\ell| = 2$, $i, \ell = 1, 2, \dots, d$, $i \neq \ell$. In other words, to obtain u_i we start with the vector of length two along e_i and rotate it towards $\mathbf{v}$.

From condition (ii) above, it follows that $a = \sqrt{2}$ and from (i) b is the positive solution of the quadratic equation $(a+b)^2 + (d-1)b^2 = 4$ and thus

$$b = \frac{-\sqrt{2} + \sqrt{2}\sqrt{1+d}}{d} \quad \Rightarrow \quad \frac{a+db}{\sqrt{d}} = \sqrt{2}\sqrt{\frac{1+d}{d}} > 1. \tag{3.44}$$

Denote by $C = \frac{\sum_{i=1}^d u_i}{d} = (\frac{a}{d} + b)\mathbf{v}$ the centroid of the points in $u_1, \dots, u_d$. Choose $u' = 2C$. Thus C is on the hyperplane containing the points in $u_1, \dots, u_d$ and O, C and u' are collinear with C being the mid point of the line segment joining

O and u'. From (3.44) we obtain

$$|u'| = 2\left(\frac{a}{d} + b\right)\sqrt{d} = 2\frac{a+db}{\sqrt{d}} > 2.$$

O is at a distance two from all the points $u_1, \ldots, u_d$ and by symmetry, the point u' is also at a distance two from all the points in $u_1, \ldots, u_d$. Thus, $O, u_1, \ldots, u_d, u'$ satisfy the assumptions of Claim 1 and so $|u'| > 2 + \delta_1$ by Claim 1. $\square$

Proof of Theorem 2.8. Throughout this proof, we take $p = \mathcal{R}$, $q = \mathcal{U}$ and fix $0 \le k \le d$. So, we omit the superscripts and subscripts as usual and denote $r_n^{\mathcal{R},\mathcal{U}}(c_n)$ by r_n and $J_{n,k}^{\mathcal{R},\mathcal{U}}(r_n^{\mathcal{R},\mathcal{U}}(c_n))$ by J_n. Recall that the sequence $\{c_n\}$ satisfies

$$\beta_n = \mathbb{E}\big[J_n\big(r_n(c_n)\big)\big] \to e^{-\alpha} \quad \text{as } n \to \infty. \tag{3.45}$$

Recall that $Q^{\mathcal{R},\mathcal{U}}(\mathbf{x}, r_n) = \bigcap_{i=1}^{k+1} B_{x_i}(2r_n)$ and let $m = m^{\mathcal{R},\mathcal{U}}$. Also, we set $h_n(\mathbf{y}) = h(\mathbf{y}, r_n)$, $Q_n(\mathbf{y}) = Q(\mathbf{y}, r_n)$, $\tilde{Q}_n(\mathbf{y}) = Q_n(\mathbf{y}) \setminus \{\mathbf{y}\}$. The proof follows by verifying the criterion given in Theorem 3.1. To invoke this criterion, take $\eta = \mathcal{P}_n$, $f(\mathbf{y}, \eta) = h_n(\mathbf{y})1[\eta(Q_n(\mathbf{y})) = 0]$. So, $W_n = F(\mathcal{P}_n)$ is the number of isolated Rips k-simplices or equivalently isolated vertices in the graph $G_k(\mathcal{P}_n, r_n)$ and more explicitly,

$$W_n = J_n\big(r_n(c_n)\big) = F(\mathcal{P}_n) = \sum_{\mathbf{y} \in \mathcal{P}_n^{(k+1)}} f(\mathbf{y}, \mathcal{P}_n \setminus \mathbf{y}) = \sum_{\mathbf{y} \in \mathcal{P}_n^{(k+1)}} h_n(\mathbf{y})1\big[\mathcal{P}_n\big(\tilde{Q}_n(\mathbf{y})\big) = 0\big]. \tag{3.46}$$

For any $\mathbf{x} \in (\mathbb{R}^d)^{k+1}$, set $\mathcal{P}_n^{\mathbf{x}} = (\mathcal{P}_n \cap Q_n(\mathbf{x})^c) \cup \{\mathbf{x}\}$. Set $U_n^{\mathbf{x}} = W_n$ and define $V_n^{\mathbf{x}}$ as

$$V_n^{\mathbf{x}} = \sum_{\substack{\mathbf{y} \in (\mathcal{P}_n^{\mathbf{x}})^{(k+1)} \\ \mathbf{y} \ne \mathbf{x}}} f\big(\mathbf{y}, \mathcal{P}_n^{\mathbf{x}}\big) = \sum_{\substack{\mathbf{y} \in (\mathcal{P}_n^{\mathbf{x}})^{(k+1)} \\ \mathbf{y} \ne \mathbf{x}}} h_n(\mathbf{y})1\big[\mathcal{P}_n^{\mathbf{x}}\big(\tilde{Q}_n(\mathbf{y})\big) = 0\big],$$

where $\mathbf{y} \ne \mathbf{x}$ denotes that $\mathbf{y}$ differs from $\mathbf{x}$ in at least one co-ordinate. Let $\mathbf{x}$ be such that $p_n(\mathbf{x}) := \mathbb{E}[f(\mathbf{x}, \mathcal{P}_n)] > 0$. In particular, this implies $h_n(\mathbf{x}) = 1$. Further, we have that

$$1 + V_n^{\mathbf{x}} = f\big(\mathbf{x}, \mathcal{P}_n^{\mathbf{x}} \setminus \mathbf{x}\big) + V_n^{\mathbf{x}} = F\big(\mathcal{P}_n^{\mathbf{x}}\big) \stackrel{d}{=} F(\mathcal{P}_n \cup \mathbf{x})|\big\{f(\mathbf{x}, \mathcal{P}_n) = 1\big\}.$$

The first equality follows because $h_n(\mathbf{x}) = 1$ and $\mathcal{P}_n^{\mathbf{x}}(\tilde{Q}_n(\mathbf{x})) = 0$, the second equality follows from definition of $F(\mathcal{P}_n^{\mathbf{x}})$ (see (3.46)) and the third equality follows because $\mathcal{P}_n \cup \{\mathbf{x}\}|\{f(\mathbf{x}, \mathcal{P}_n) = 1\} = \mathcal{P}_n \cup \{\mathbf{x}\}|\{\mathcal{P}_n(\tilde{Q}_n(\mathbf{x})) = 0\} \stackrel{d}{=} \mathcal{P}_n^{\mathbf{x}}$. We define

$$\begin{aligned} W_n^{(1)}(\mathbf{x}) &:= \sum_{\substack{\mathbf{y} \in \mathcal{P}_n^{(k+1)} \\ Q_n(\mathbf{y}) \cap Q_n(\mathbf{x}) \ne \varnothing}} f(\mathbf{y}, \mathcal{P}_n), \qquad W_n^{(2)}(\mathbf{x}) := \sum_{\substack{\mathbf{y} \in (\mathcal{P}_n^{\mathbf{x}})^{(k+1)}, \mathbf{y} \ne \mathbf{x} \\ Q_n(\mathbf{y}) \cap Q_n(\mathbf{x}) \ne \varnothing}} f\big(\mathbf{y}, \mathcal{P}_n^{\mathbf{x}}\big), \\ W_n^{(3)}(\mathbf{x}) &:= \sum_{\substack{\mathbf{y} \in \mathcal{P}_n^{(k+1)} \\ Q_n(\mathbf{y}) \cap Q_n(\mathbf{x}) = \varnothing}} f(\mathbf{y}, \mathcal{P}_n) = \sum_{\substack{\mathbf{y} \in (\mathcal{P}_n^{\mathbf{x}})^{(k+1)} \\ Q_n(\mathbf{y}) \cap Q_n(\mathbf{x}) = \varnothing}} f\big(\mathbf{y}, \mathcal{P}_n^{\mathbf{x}}\big), \end{aligned} \tag{3.47}$$

where the last equality follows by observing that $\mathbf{y} \ne \mathbf{x}$ and $f(\mathbf{y}, \mathcal{P}_n^{\mathbf{x}}) = f(\mathbf{y}, \mathcal{P}_n)$ if $Q_n(\mathbf{y}) \cap Q_n(\mathbf{x}) = \varnothing$. Now, we can write $U_n^{\mathbf{x}} = W_n^{(1)}(\mathbf{x}) + W_n^{(3)}(\mathbf{x})$ and $V_n^{\mathbf{x}} = W_n^{(2)}(\mathbf{x}) + W_n^{(3)}(\mathbf{x})$. This yields $|U(\mathbf{x}) - V(\mathbf{x})| \le W_1(\mathbf{x}) + W_2(\mathbf{x})$.

We let $w_n(\mathbf{x}) = w_n^{(1)}(\mathbf{x}) + w_n^{(2)}(\mathbf{x})$ with $w_n^{(i)}(\mathbf{x}) = \mathbb{E}[W_n^{(i)}(\mathbf{x})]$, $i = 1, 2$. Then applying Theorem 3.1, we obtain

$$d_{\mathrm{TV}}\big(J_n, \mathrm{Poi}(\beta_n)\big) \le \frac{1 \wedge \beta_n^{-1}}{(k+1)!}(I_1 + I_2), \tag{3.48}$$

where

$$I_i = n^{k+1} \int_{U^{k+1}} w_n^{(i)}(\mathbf{x}) p_n(\mathbf{x})\, d\mathbf{x}, \quad i = 1, 2. \tag{3.49}$$

The result now follows from (3.45) and (3.48) provided we show that $I_i \to 0$ as $n \to \infty$ for $i = 1, 2$. Recall that

$$p_n(\mathbf{x}) = \mathbb{E}\big[f(\mathbf{x}, \mathcal{P}_n)\big] = h_n(\mathbf{x}) e^{-n|Q_n(\mathbf{x})|}.$$

By the Campbell–Mecke formula applied to $w_1(\mathbf{x})$ as in (3.1), (3.2) and noting that $Q_n(\mathbf{y})\cap Q_n(\mathbf{x})\neq\varnothing$ as well as setting $Q(O,y):=Q((O,\mathbf{y}),1)$, we obtain

$$\begin{aligned}
I_1 &= C_1 n^{k+1}\int_{U^{k+1}} d\mathbf{x}\, p_n(\mathbf{x}) n^{k+1}\int_{\{\mathbf{y}:Q_n(\mathbf{y})\cap Q_n(\mathbf{x})\neq\varnothing\}} p_n(\mathbf{y})\,d\mathbf{y}\\
&\le C_1 n^{2(k+1)}\int_{\mathbf{x}\in U^{k+1}}\int_{\mathbf{y}\in B_{x_1}(6r_n)\times U^k} h_n(\mathbf{x})h_n(\mathbf{y})e^{-n(|Q_n(\mathbf{x})|+|Q_n(\mathbf{y})|)}\,d\mathbf{x}\,d\mathbf{y}\\
&\le C_2 r_n^d\biggl(n\bigl(nr_n^d\bigr)^k\int_{(r_n^{-1}U)^k} h(O,\mathbf{y})e^{-nr_n^d|Q(O,\mathbf{y})|}\,d\mathbf{y}\biggr)^2 = C_2 r_n^d\beta_n^2\to 0,
\end{aligned}\tag{3.50}$$

from (3.3), (3.45) and the fact that $r_n\to 0$ as $n\to\infty$.

To analyse I_2, we will write it as a sum depending on the number of coordinates common to $\mathbf{x}$ and $\mathbf{y}$.

$$I_2=\sum_{j=0}^{k} I_{2j},$$

where

$$I_{2j}:=\binom{k+1}{j} n^{2(k+1)-j}\int_{\mathbf{x}\in U^{k+1}}\int_{\mathbf{z}\in(U\setminus Q(\mathbf{x},r_n))^{k-j+1}} h_n(\mathbf{x})h_n(\mathbf{y})e^{-n|Q_n(\mathbf{x})\cup Q_n(\mathbf{y})|}\,d\mathbf{x}\,d\mathbf{z},\tag{3.51}$$

$\mathbf{y}=(x_{k-j+2},\dots,x_{k+1},z_1,\dots z_{k-j+1})$ and $\mathbf{z}=(z_1,\dots z_{k-j+1})$. Note that $\mathbf{x}$, $\mathbf{y}$ have j coordinates $(x_{k-j+2},\dots,x_{k+1})$ in common. Since the metric is toroidal, for any $x_1\in U$ the integration with respect to the remaining variables yields a function that does not depend on x_1. Hence we can fix the first variable to be the origin O. Bounding (3.51) as in (3.50), we obtain

$$I_{2j}\le C_2 n\bigl(nr_n^d\bigr)^{2k-j+1}\int_{B_O(2)^k} d\mathbf{x}\int_{(B_O(6)\setminus Q(O,\mathbf{x}))^{k-j+1}} d\mathbf{z}\, h\bigl((O,\mathbf{x})\bigr)h(\mathbf{y})e^{-nr_n^d|Q(O,\mathbf{x})\cup Q(\mathbf{y})|}=L_{1j}+L_{2j},\tag{3.52}$$

where $\mathbf{x}=(x_2,\dots,x_{k+1})$, $\mathbf{y}$, $\mathbf{z}$ are as above and

$$\begin{aligned}
L_{1j}&:=C_2 n\bigl(nr_n^d\bigr)^{2k-j+1}\int_{B_O(2)^k} d\mathbf{x}\int_{(B_O(6)\setminus Q(O,\mathbf{x}))^{k-j+1}} d\mathbf{z}\mathbf{1}\bigl(\bigl|Q(O,\mathbf{x})\bigr|\vee\bigl|Q(\mathbf{y})\bigr|>m+\epsilon\bigr)\\
&\quad\times h\bigl((O,\mathbf{x})\bigr)h(\mathbf{y})e^{-nr_n^d|Q(O,\mathbf{x})\cup Q(\mathbf{y})|},
\end{aligned}\tag{3.53}$$

$$\begin{aligned}
L_{2j}&:=C_2 n\bigl(nr_n^d\bigr)^{2k-j+1}\int_{B_O(2)^k} d\mathbf{x}\int_{(B_O(6)\setminus Q(O,\mathbf{x}))^{k-j+1}} d\mathbf{z}\,\mathbf{1}\bigl(\bigl|Q(O,\mathbf{x})\bigr|\vee\bigl|Q(\mathbf{y})\bigr|\le m+\epsilon\bigr)\\
&\quad\times h\bigl((O,\mathbf{x})\bigr)h(\mathbf{y})e^{-nr_n^d|Q(O,\mathbf{x})\cup Q(\mathbf{y})|}
\end{aligned}\tag{3.54}$$

and $\epsilon>0$ is arbitrary. Using the restriction $|Q(O,\mathbf{x})|\vee|Q(\mathbf{y})|>m+\epsilon$ and substituting for r_n in (3.53) yields the bound

$$L_{1j}\le C_3 n\bigl(nr_n^d\bigr)^{2k-j+1}e^{-\frac{m+\epsilon}{c_n}(\log n+(k-1)\log\log n+\alpha)}.\tag{3.55}$$

Since $c_n\to m$, we can choose $\eta\in(0,\frac{\epsilon}{m})$ such that $m+\epsilon>(1+\eta)c_n$ for all n sufficiently large. Using this in (3.55) we obtain that as $n\to\infty$

$$L_{1j}\le C_4\frac{n(\log n)^{2k-j+1}}{n^{1+\eta}}\to 0.\tag{3.56}$$

It remains to show that $L_{2j}\to 0$. Denote by

$$\begin{aligned}
\tilde{D}_{j,\epsilon}=\bigl\{&(\mathbf{x},\mathbf{z})\in B_O(2)^k\times\bigl(B_O(6)\setminus Q(O,\mathbf{x})\bigr)^{k-j+1}:\\
&\bigl|Q(O,\mathbf{x})\bigr|\vee\bigl|Q(\mathbf{y})\bigr|\le m+\epsilon, h\bigl((O,\mathbf{x})\bigr)=1, h(\mathbf{y})=1\bigr\},
\end{aligned}\tag{3.57}$$

the region of integration in (3.54), where $\mathbf{x}=(x_2,\dots,x_{k+1})\in\mathbb{R}^{dk}$, $\mathbf{z}=(z_1,\dots,z_{k-j+1})\in\mathbb{R}^{d(k-j+1)}$, and $\mathbf{y}=(x_{k-j+2},\dots,x_{k+1},z_1,\dots,z_{k-j+1})\in\mathbb{R}^{d(k+1)}$.

Since $|Q(O,\mathbf{x})| \vee |Q(\mathbf{y})|$ is continuous in $\tilde{D}_{j,\epsilon}$ and the minimum m is achieved in $D_{j,\delta}$ for any $\delta > 0$ (see Lemma 3.2 for definition), we can choose an ϵ small enough such that there exists a δ_0 small enough with $\tilde{D}_{j,\epsilon} \subset D_{j,\delta_0}$. It then follows from Lemma 3.2, (3.54), (3.45) and (2.7) that for $\epsilon > 0$ sufficiently small,

$$\begin{aligned}
L_{2j} &\le C_2 n\big(nr_n^d\big)^{2k-j+1} \int_{\tilde{D}_{j,\epsilon}} d\mathbf{x}\, d\mathbf{z}\, e^{-nr_n^d(|Q(O,\mathbf{x})|+|Q(\mathbf{y})\setminus Q(O,\mathbf{x})|)} \\
&\le C_5 n\big(nr_n^d\big)^{2k-j+1} \int_{\tilde{D}_{j,\epsilon}} dx\, d\mathbf{z}\, e^{-nr_n^d(|Q(O,\mathbf{x})|+\beta)} \\
&\le C_6 \beta_n \big(nr_n^d\big)^{k-j+1} e^{-nr_n^d\beta} \le C_7 (\log n)^{k-j+1} n^{-\beta/c_n} \to 0,
\end{aligned} \tag{3.58}$$

as $c_n \to m > 0$. This completes the proof of Theorem 2.8. □

Proof of Theorem 2.9. We shall again fix $1 \le k \le d$, $p = \mathcal{R}$, $q = \mathcal{U}$ and omit these subscripts and superscstipts. Further, let $r_n = r_n^{\mathcal{R},\mathcal{U}}(c_n)$. By Slutksy's lemma [26, Chapter 6, Theorem 6.5] and Theorem 2.8, it suffices to show that for any $L \ge 2$, $\mathbb{E}[J_n(r_n, L)] \to 0$. Now, fix $L \ge 2$ and let $\Gamma \in \Delta(L, M)$. (Recall the notation in (3.42) and (3.43) from the proof of Theorem 1.3.) We shall show that

$$\mathbb{E}\big[J_n(r_n,\Gamma)\big] \to 0.$$

Deriving as in the proof of Theorem 1.3 (and using the same notation), we have the following bound:

$$\mathbb{E}\big[J_n(r_n,\Gamma)\big] \le \frac{n(n(r_n)^d)^{M-1}}{M!} \int_{(B_O(K))^{M-1}} \prod_{i=1}^{L} h\big(\mathbf{x}^i\big) e^{-n(r_n)^d|\cup_{i=0}^{L} Q(\mathbf{x}^i)|}\, dx_2 \dots dx_M,$$

where for all $1 \le i \le L$, $\mathbf{x}^i$ is a $(k+1)$-subset of $\{x_1, \dots, x_M\}$ where we have set $x_1 = O$. Now, we shall break the proof into three cases. Fix $\epsilon > 0$ which will be chosen later. We shall break the integral into three cases i.e., define

$$\begin{aligned}
A_1 &:= \Big\{\mathbf{x} \in \big(B_O(K)\big)^{M-1} : \max_{i=1,\dots,L} \big|Q\big(\mathbf{x}^i\big)\big| > m + \epsilon\Big\}, \\
A_2 &:= \Big\{\mathbf{x} \in \big(B_O(K)\big)^{M-1} : \max_{i=1,\dots,L} \big|Q\big(\mathbf{x}^i\big)\big| \le m + \epsilon,\ \max_{1\le i<j\le M} |x_i - x_j| > 2\Big\}, \\
A_3 &:= \Big\{\mathbf{x} \in \big(B_O(K)\big)^{M-1} : \max_{i=1,\dots,L} \big|Q\big(\mathbf{x}^i\big)\big| \le m + \epsilon,\ \max_{1\le i<j\le M} |x_i - x_j| \le 2\Big\}.
\end{aligned}$$

Thus, we write $\mathbb{E}[J_n(r_n,\Gamma)] \le I_1 + I_2 + I_3$, where I_i's are defined as

$$I_i := \frac{n(n(r_n^*)^d)^{M-1}}{M!} \int_{A_i} \prod_{i=1}^{L} h\big(\mathbf{x}^i\big) e^{-n(r_n^*)^d|\bigcup_{i=0}^{L} Q(\mathbf{x}^i)|}\, d\mathbf{x}_2 \dots d\mathbf{x}_M.$$

Now we shall show that $I_i \to 0$ as $n \to \infty$ for $i = 1, 2, 3$ and thus complete the proof.

First consider I_1. Here we have that $\max_{i=1,\dots,L} |Q(\mathbf{x}^i)| > m + \epsilon$. This is the easiest of the three cases. Here using a bound similar to L_{1j} in (3.55), we can show that I_1 converges to 0 as in (3.56).

The analysis of the remaining two cases will follow along similar lines to the bounds obtained for L_{2j} in the proof of Theorem 2.8 using Lemma 3.2.

Next consider I_2. Here we have that $\max_{i=1,\dots,L} |Q(\mathbf{x}^i)| \le m + \epsilon$ but $\max_{1\le i<j\le M} |x_i - x_j| > 2$.

Without loss of generality, re-write $\mathbf{x}^1 = (O, \mathbf{x}) = (O, x_2, \dots, x_{k+1})$, $\mathbf{x}^2 = \mathbf{y} = (x_{k-j+2}, \dots, x_{k+1}, z_1, \dots, z_{k-j+1})$ for some $j \ge 1$ with $|z_1| > 2$. Setting $\mathbf{z} = (z_1, \dots, z_{k-j+1})$, we have that $(\mathbf{x}, \mathbf{z}) \in \tilde{D}_{j,\epsilon}$, where $\tilde{D}_{j,\epsilon}$ is as defined in (3.57). Now, as we argued below (3.57) using continuity of $Q(.)$ as well as minimum being achieved in $D_{j,\delta}$, we again have that for ϵ small enough, there exists $\delta_0 > 0$ such that $\tilde{D}_{j,\epsilon} \subset D_{j,\delta_0}$ and hence from the geometric Lemma 3.2, we have that the following inequality holds for some $\beta > 0$:

$$\big|Q\big(\mathbf{x}^1\big) \cup Q\big(\mathbf{x}^2\big)\big| \ge \big|Q\big(\mathbf{x}^1\big)\big| + \beta. \tag{3.59}$$

Thus we have that for some constant C,

$$I_2 \le C\big(n\big(r_n^*\big)^d\big)^{M-k-1} e^{-\beta n(r_n^*)^d} \times n\big(n\big(r_n^*\big)^d\big)^k \int_{(B_O(K))^k} e^{-n(r_n^*)^d|Q(O,x_2,\dots,x_{k+1})|}\, dx_2 \dots dx_{k+1}.$$

Since the convergence of the latter term on the RHS follows by Proposition 2.4 and the first term converges to 0, we have that $I_2 \to 0$.

Finally, consider I_3 and here we have that $\max_{i=1,\dots,L} |Q(\mathbf{x}^i)| \le m + \epsilon$ but $\max_{1 \le i < j \le M} |x_i - x_j| \le 2$.

Firstly note that this means that $M \le d + 1$ but because Γ is a component of order at least two, $M \ge k + 2$. Thus, $A_3 = \varnothing$ unless we assume that $k + 2 \le M \le d + 1$. Further, $L = \binom{M}{k+1}$ since all $(k+1)$-tuples will form k-simplices.

Since for all k-simplices $\mathbf{x}^i$ we have that $|Q(\mathbf{x}^i)| \le m + \epsilon$ and by continuity of $|Q(\mathbf{x}^i)|$, there exists $\delta > 0$ (depending on ϵ) such that $2 - \delta \le |x_i - x_j| \le 2$ for all $1 \le i < j \le M$. Again using geometric Lemma 3.2, we have that for ϵ sufficiently small, there exists $\beta > 0$ such that (3.59) holds. Now, by proceeding as in case of I_2, we conclude that $I_3 \to 0$ as $n \to \infty$. □

Appendix: Simplicial homology and Morse critical points

In this appendix, to help the reader to understand the connections to other random topology literature referenced in the introduction, we shall quickly introduce some basic notions from simplicial homology. For more details, refer to [19,41].

We shall use the definitions and notations from Section 2.1. Most results in random topology assume that the coefficients are from a field $\mathbb{F}$. For ease of exposition, we shall further assume that $\mathbb{F} = \mathbb{Z}_2 = \{0, 1\}$. Let $\mathcal{K}$ be a finite simplicial complex. We assume $k \ge 0$. A *simplicial k-chain* is a formal sum $\sum_i c_i \sigma_i$ with $\sigma_i \in S_k(\mathcal{K})$ and $c_i \in \mathbb{F}$. The collection of all simplicial k-chains is the free abelian group C_k and we set $C_{-1} = \{0\}$. Now we define the *k-th boundary operator* $\partial_k : C_k \to C_{k-1}$ by defining it on simplices as follows and then extending it linearly:

$$\partial_k\big([v_1, \dots, v_{k+1}]\big) := \sum_{j=1}^{k+1} [v_1, \dots, \hat{v_j}, \dots, v_{k+1}],$$

where $[v_1, \dots, \hat{v_j}, \dots, v_{k+1}]$ is to denote that the jth vertex has been deleted from the simplex. The key property of the boundary operators is that $\partial_k \partial_{k+1} = 0$. We now define the *cycle group* as $Z_k := \ker \partial_k$ and *boundary group* $B_k := im\, \partial_{k+1}$ where ker and im denote kernel and image respectively. Finally, we define the kth homology group as $H_k := Z_k / B_k$ and it is well-defined because of the property of boundary operators. Since our coefficients are from a field, all the above-defined groups are actually vector spaces and hence we can define the kth Betti number as $\beta_k := \mathrm{rank}(H_k) = \mathrm{rank}(Z_k) - \mathrm{rank}(B_k)$. Intuitively, $\mathrm{rank}(Z_k)$ is the number of 'linearly independent' k-cycles and β_k is the number of 'linearly independent' non-trivial k-cycles or more geometrically, the number of non-trivial holes.

Given a locally finite-set of points $\mathcal{X}$ and $r \ge 0$, we have defined the Čech complex $\mathcal{C}(\mathcal{X}, r)$ in Definition 2.2. Further, we also have the Euclidean subset $\cup_{x \in \mathcal{X}} B_x(r)$. The homology groups for the latter are defined via the so-called singular homology which we shall not introduce here. However, from the important *nerve theorem* [6, Theorem 10.7], we have that the homology groups of $\bigcup_{x \in \mathcal{X}} B_x(r)$ and the corresponding Čech complex are isomorphic. Indeed, they are shown to be homotpy equivalent to each other but for our purposes isomorphism of homology groups suffices. Since $\cup_{x \in \mathcal{X}} B_x(r) \subset \mathbb{R}^d$, we have that $H_k(\mathcal{C}(\mathcal{X}, r)) = 0$ for $k \ge d$. However, in our case, we consider $\mathcal{X} \subset [0, 1]^d$ and with the toroidal metric. Hence $\bigcup_{x \in \mathcal{X}} B_x(r)$ is actually embedded in $\mathbb{R}^{d+1}$ and so $H_k(\mathcal{C}(\mathcal{X}, r)) = 0$ for $k > d$.

A very important tool in understanding topology of geometric complexes has been *discrete Morse theory*. Classically Morse theory has dealt with critical points of smooth functions and in discrete Morse theory, this is extended to non-smooth functions. The specific approach for Čech complexes is discussed in detail in [9] and we sketch the same here. For a finite point process $\mathcal{X}$, define the distance function $d_{\mathcal{X}} : \mathbb{R}^d \to \mathbb{R}_+$ as

$$d_{\mathcal{X}}(x) := \min_{X \in \mathcal{X}} \|x - X\|, \quad x \in \mathbb{R}^d.$$

Though this is non-smooth, the critical points are defined as follows. Index 0 critical points are the points where $d_{\mathcal{X}} = 0$ (local and global minima) which are nothing but points of $\mathcal{X}$. We will need a little notation before defining higher index critical points. The points $x_i, 1 \le i \le k+1$ are said to lie in general position if the points do not lie in a $(k-1)$-dimensional affine space. In such a case, $C(\{x_1, \dots, x_{k+1}\})$ denote the center of the unique $k-1$ dimensional sphere, $R(\{x_1, \dots, x_{k+1}\})$ be the radius of this unique ball and $conv^o(\{x_1, \dots, x_{k+1}\})$ denotes the interior of the convex hull formed by the points of $\{x_1, \dots, x_{k+1}\}$. For higher indices $1 \le k \le d$, a point $c \in \mathbb{R}^d$ is said to be an *index k critical point* if there exists a $\{x_1, \dots, x_{k+1}\} \subset \mathcal{X}$ which lie in general position such that $C(\{x_1, \dots, x_{k+1}\}) \in conv^o(\{x_1, \dots, x_{k+1}\})$ and $\mathcal{X}(B_{C(\{x_1,\dots,x_{k+1}\})}(R(\{x_1, \dots, x_{k+1}\}))) = 0$. We shall also be interested in critical points c which are at most at a distance r from $\mathcal{X}$ i.e., $d_{\mathcal{X}}(c) \le r$ and denote the number of such critical points as $N_k(\mathcal{X}, r)$. One has the trivial Morse inequality that $\beta_k(\mathcal{C}(\mathcal{X}, r)) \le N_k(\mathcal{X}, r)$ and further if $N_k(\mathcal{X}, r) = N_k(\mathcal{X}, s)$ and $N_{k+1}(\mathcal{X}, r) = N_{k+1}(\mathcal{X}, s)$ for $r > s$, then $\beta_k(\mathcal{C}(\mathcal{X}, r)) = \beta_k(\mathcal{C}(\mathcal{X}, s))$. These two facts are crucially used in understanding Betti numbers of random Čech complexes via Morse critical points in [9,12,13].

Acknowledgements

The authors are thankful to Omer Bobrowski for many discussions as well as sharing drafts of his preprints. The authors also wish to thank an anonymous referee for comments leading to an improved exposition.

References

[1] M. J. B. Appel and R. P. Russo. The connectivity of a graph on uniform points on $[0,1]^d$. *Statist. Probab. Lett.* **60** (2002) 351–357. MR1947174 https://doi.org/10.1016/S0167-7152(02)00233-X

[2] R. Atkin. An algebra for patterns on a complex, I. *Int. J. Man-Mach. Stud.* **6** (3) (1974) 285–307. MR0424293

[3] R. Atkin. An algebra for patterns on a complex, II. *Int. J. Man-Mach. Stud.* **8** (5) (1976) 483–498. MR0574079 https://doi.org/10.1016/s0020-7373(76)80015-6

[4] H. Barcelo, X. Kramer, R. Laubenbacher and C. Weaver. Foundations of a connectivity theory for simplicial complexes. *Adv. in Appl. Math.* **26** (2) (2001) 97–128. MR1808443 https://doi.org/10.1006/aama.2000.0710

[5] H. Barcelo and R. Laubenbacher. Perspectives on A-homotopy theory and its applications. *Discrete Math.* **298** (1–3) (2005) 39–61. MR2163440 https://doi.org/10.1016/j.disc.2004.03.016

[6] A. Björner. Topological methods. In *Handbook of Combinatorics* 1819–1872, **2**, 1995.

[7] B. Błaszczyszyn and D. Yogeshwaran. Clustering and percolation of point processes. *Electron. J. Probab.* **18** (2013) 72. MR3091718 https://doi.org/10.1214/EJP.v18-2468

[8] O. Bobrowski. Homological connectivity in Čech complexes, 2019. Available at arXiv:1906.04861.

[9] O. Bobrowski and R. J. Adler. Distance functions, critical points, and the topology of random Čech complexes. *Homology, Homotopy Appl.* **16** (2) (2014) 311–344. MR3280987 https://doi.org/10.4310/HHA.2014.v16.n2.a18

[10] O. Bobrowski and M. Kahle. Topology of random geometric complexes: A survey. *J. Appl. Comput. Top.* (2014) 1–34. MR3975557 https://doi.org/10.1007/s41468-017-0010-0

[11] O. Bobrowski and S. Mukherjee. The topology of probability distributions on manifolds. *Probab. Theory Related Fields* **161** (3) (2015) 651–686. MR3334278 https://doi.org/10.1007/s00440-014-0556-x

[12] O. Bobrowski and G. Oliveira. Random Čech complexes on Riemannian manifolds. *Random Structures Algorithms* **54** (3) (2019) 373–412. MR3938773 https://doi.org/10.1002/rsa.20800

[13] O. Bobrowski and S. Weinberger. On the vanishing of homology in random Čech complexes. *Random Structures Algorithms* **51** (1) (2017) 14–51. MR3668845 https://doi.org/10.1002/rsa.20697

[14] B. Bollobás and O. Riordan. Clique percolation. *Random Structures Algorithms* **35** (3) (2009) 294–322. MR2548516 https://doi.org/10.1002/rsa.20270

[15] G. Carlsson. Topological pattern recognition for point cloud data. *Acta Numer.* **23** (2014) 289–368. MR3202240 https://doi.org/10.1017/S0962492914000051

[16] A. E. Costa, M. Farber and T. Kappeler. Topics of stochastic algebraic topology. In *Proceedings of the Workshop on Geometric and Topological Methods in Computer Science (GETCO)* 53–70. *Electronic Notes in Theoretical Computer Science* **283**, 2012. MR2955097 https://doi.org/10.1016/j.entcs.2012.05.005

[17] H.-L. de Kergorlay, U. Tillmann and O. Vipond. Random Čech complexes on manifolds with boundary, 2019. Available at arXiv:1906.07626.

[18] I. Derényi, G. Palla and T. Vicsek. Clique percolation in random networks. *Phys. Rev. Lett.* **94** (16) (2005). https://doi.org/10.1103/PhysRevLett.94.160202

[19] H. Edelsbrunner and J. L. Harer. *Computational Topology, an Introduction*. American Mathematical Society, Providence, RI, 2010. MR2572029

[20] P. Erdös and A. Rényi. On random graphs. I. *Publ. Math. Debrecen* **6** (1959) 290–297. MR0120167

[21] L. Flatto and D. J. Newman. Random coverings. *Acta Math.* **138** (1) (1977) 241–264. MR0440651 https://doi.org/10.1007/BF02392317

[22] E. N. Gilbert. Random plane networks. *SIAM J. Appl. Math.* **9** (1961) 533–543. MR0132566

[23] L. Goldstein and M. D. Penrose. Normal approximation for coverage models over binomial point processes. *Ann. Appl. Probab.* **20** (2010) 696–721. MR2650046 https://doi.org/10.1214/09-AAP634

[24] A. Gundert and U. Wagner. On Laplacians of random complexes. In *Proceedings of the Twenty-Eighth Annual Symposium on Computational Geometry* 151–160. ACM, New York, 2012. MR3024710 https://doi.org/10.1145/2261250.2261272

[25] B. Gupta and S. K. Iyer. Criticality of the exponential rate of decay for the largest nearest-neighbor link in random geometric graphs. *Adv. in Appl. Probab.* **42** (3) (2010) 631–658. MR2779553 https://doi.org/10.1239/aap/1282924057

[26] A. Gut. *An Intermediate Course in Probability*. Springer, Berlin, 2009. MR2528081 https://doi.org/10.1007/978-1-4419-0162-0

[27] P. Hall. *Introduction to the Theory of Coverage Processes*. Wiley, New York, 1988. MR0973404 https://doi.org/10.1016/0167-0115(88)90159-0

[28] Y. Hiraoka and T. Shirai. Minimum spanning acycle and lifetime of persistent homology in the Linial–Meshulam process. *Random Structures Algorithms* **51** (2) (2017) 315–340. MR3683365 https://doi.org/10.1002/rsa.20718

[29] D. Horak and J. Horst. Spectra of combinatorial Laplace operators on simplicial complexes. *Adv. Math.* **244** (2013) 303–336. MR3077874 https://doi.org/10.1016/j.aim.2013.05.007

[30] S. K. Iyer and D. Thacker. Nonuniform random geometric graphs with location-dependent radii. *Ann. Appl. Probab.* **22** (5) (2012) 2048–2066. MR3025688 https://doi.org/10.1214/11-AAP823

[31] M. Kahle. Topology of random clique complexes. *Discrete Math.* **309** (6) (2009) 1658–1671. MR2510573 https://doi.org/10.1016/j.disc.2008.02.037

[32] M. Kahle. Random geometric complexes. *Discrete Comput. Geom.* **45** (3) (2011) 553–573. MR2770552 https://doi.org/10.1007/s00454-010-9319-3

[33] M. Kahle. Sharp vanishing thresholds for cohomology of random flag complexes. *Ann. of Math.* **179** (2014) 1085–1107. MR3171759 https://doi.org/10.4007/annals.2014.179.3.5

[34] M. Kahle. Topology of random simplicial complexes: A survey. *Contemp. Math.– Am. Math. Soc.* **620** (2014) 201–222. MR3290093 https://doi.org/10.1090/conm/620/12367

[35] M. Kahle and B. Pittel. Inside the critical window for cohomology of random k-complexes *Rand. Struct. Alg.* **48** (1) (2014) 102–124. MR3432573 https://doi.org/10.1002/rsa.20577

[36] M. Kraetzl, R. Laubenbacher and M. E. Gaston. Combinatorial and algebraic approaches to network analysis. DSTO Internal Report, 2001.

[37] N. Linial and R. Meshulam. Homological connectivity of random 2-complexes. *Combinatorica* **26** (4) (2006) 475–487. MR2260850 https://doi.org/10.1007/s00493-006-0027-9

[38] R. Meshulam and N. Wallach. Homological connectivity of random k-dimensional complexes. *Random Structures Algorithms* **34** (3) (2009) 408–417. MR2504405 https://doi.org/10.1002/rsa.20238

[39] P. A. P. Moran. The random volume of interpenentrating spheres in space. *J. Appl. Probab.* **10** (1973) 837–846. MR0353412 https://doi.org/10.2307/3212370

[40] S. Mukherjee and J. Steenbergen. Random walks on simplicial complexes and harmonics. *Random Structures Algorithms* **49** (2) (2016) 379–405. MR3536545 https://doi.org/10.1002/rsa.20645

[41] J. R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, Reading, 1984. MR0755006

[42] P. Niyogi, S. Smale and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** (1) (2008) 419–441. MR2383768 https://doi.org/10.1007/s00454-008-9053-2

[43] T. Owada. Limit theorems for the sum of persistence barcodes. *Ann. Appl. Probab.* **28** (5) (2018) 2814–2854. MR3847974 https://doi.org/10.1214/17-AAP1375

[44] T. Owada and R. J. Adler. Limit theorems for point processes under geometric constraints (and topological crackle). *Ann. Probab.* **45** (3) (2017) 2004–2055. MR3650420 https://doi.org/10.1214/16-AOP1106

[45] G. Palla, D. Ábel, I. Farkas, P. Pollner, I. Derényi and T. Vicsek. k-clique percolation and clustering. In *Handbook of Large-Scale Random Networks* 369–408. B. Bollobás, R. Kozma and D. Miklós (Eds). Springer, Berlin Heidelberg, 2008. MR2582394 https://doi.org/10.1007/978-3-540-69395-6_9

[46] O. Parzanchevski and R. Rosenthal. Simplicial complexes: Spectrum, homology and random walks. *Random Structures Algorithms* **50** (2) (2017) 225–261. MR3607124 https://doi.org/10.1002/rsa.20657

[47] M. D. Penrose. The longest edge of the random minimal spanning tree. *Ann. Appl. Probab.* **7** (1997) 340–361. MR1442317 https://doi.org/10.1214/aoap/1034625335

[48] M. D. Penrose. *Random Geometric Graphs*. Oxford University Press, New York, 2003. MR1986198 https://doi.org/10.1093/acprof:oso/9780198506263.001.0001

[49] M. D. Penrose. Inhomogeneous random graphs, isolated vertices, and Poisson approximation. *J. Appl. Probab.* **55** (1) (2018) 112–136. MR3780386 https://doi.org/10.1017/jpr.2018.9

[50] R. Schneider and W. Weil. *Stochastic and Integral Geometry*. Springer, Berlin, 2008. MR2455326 https://doi.org/10.1007/978-3-540-78859-1

[51] P. Skraba, G. Thoppe and D. Yogeshwaran. Randomly weighted d-complexes: Minimal spanning acycles and persistence diagrams. *Elec. J. Comb.* **27**(2) (2020). https://doi.org/10.37236/8679

[52] D. Yogeshwaran and R. J. Adler. On the topology of random complexes built over stationary point processes. *Ann. Appl. Probab.* **25** (6) (2015) 3338–3380. MR3404638 https://doi.org/10.1214/14-AAP1075