The Annals of Applied Probability
2015, Vol. 25, No. 6, 3338–3380
DOI: 10.1214/14-AAP1075

ON THE TOPOLOGY OF RANDOM COMPLEXES BUILT OVER STATIONARY POINT PROCESSES[1]

BY D. YOGESHWARAN AND ROBERT J. ADLER

Indian Statistical Institute, Bangalore and Technion—Israel Institute of Technology

There has been considerable recent interest, primarily motivated by problems in applied algebraic topology, in the homology of random simplicial complexes. We consider the scenario in which the vertices of the simplices are the points of a random point process in $\mathbb{R}^d$, and the edges and faces are determined according to some deterministic rule, typically leading to Čech and Vietoris–Rips complexes. In particular, we obtain results about homology, as measured via the growth of Betti numbers, when the vertices are the points of a general stationary point process. This significantly extends earlier results in which the points were either i.i.d. observations or the points of a Poisson process. In dealing with general point processes, in which the points exhibit dependence such as attraction or repulsion, we find phenomena quantitatively different from those observed in the i.i.d. and Poisson cases. From the point of view of topological data analysis, our results seriously impact considerations of model (non)robustness for statistical inference. Our proofs rely on analysis of subgraph and component counts of stationary point processes, which are of independent interest in stochastic geometry.

1. Introduction. There has been considerable recent interest, primarily motivated by problems in applied algebraic topology, in the homology of random simplicial complexes. Two main scenarios have been considered. In the geometric model, the vertices of the simplices are a random point set, and the edges and faces are determined according to some deterministic rule, typically related to the distance between pairs, or general subsets, of vertices. This has lead, for example, to the Čech and Vietoris–Rips complexes on random Euclidean point sets, studied in papers such as [22, 24], with an extension to the manifold setting in [7].

Another approach has been to consider random subgraphs of complete graphs, leading to a number of papers dealing with the topology of random complexes generalising Erdős–Rényi graphs, as in, for example, [1, 14, 21, 27, 29]. Also, see the recent survey [23] for progress in this direction.

Received November 2012; revised September 2014.

[1]Supported in part by Israel Science Foundation 853/10, AFOSR FA8655-11-1-3039 and FP7-ICT-318493-STREP.

MSC2010 subject classifications. Primary 60G55, 60D05, 05E45; secondary 05C10, 55U10, 58K05.

Key words and phrases. Point process, random geometric complex, Čech, Vietoris–Rips, component counts, Betti numbers, Morse critical points, negative association.

The current paper is concerned with the first of these approaches, although with a novel and—from the point of view of both theory and applications—important change of emphasis. Previous papers on simplicial complexes built over random point sets have always assumed that the points were either independent, identically distributed (i.i.d.) observations from some underlying distribution on $\mathbb{R}^d$, or points of a (typically nonhomogeneous) Poisson point process. Our aim in this paper is to investigate situations in which the points are chosen from a general point process, in which the points exhibit dependence such as attraction or repulsion. From the point of view of topological data analysis (TDA) our results, which show that local dependencies can have a major effect on the growth rates of topological quantifiers such as Betti numbers, impact on considerations of model (non)robustness for statistical inference in TDA. We shall not address these issues here, however, beyond a few comments in Section 1.3 below.

To start being a little more specific, given a point process (i.e., locally finite random counting measure) Φ on $\mathbb{R}^d$, recall that the random geometric graph $G(\Phi, r)$, for $r > 0$, is defined as the graph with vertex set Φ and (undirected) edge set $\{(X, Y) \in \Phi^2 : \|X - Y\| \le r\}$. The properties of random geometric graphs when Φ is a Poisson point process or a point process of i.i.d. points have been analysed in detail (cf. [35]), and recently interest has turned to the richer topic of random simplicial complexes built over these point sets.

A nonempty family $\mathcal{K}$ of finite subsets of a finite set V (called *vertices*) is *an abstract simplicial* complex if $\mathcal{X} \in \mathcal{K}$ and $\mathcal{Y} \subset \mathcal{X}$ implies that $\mathcal{Y} \in \mathcal{K}$. Elements of $\mathcal{K}$ are called *faces* or *simplices*, and the *dimension of a face* is defined as its cardinality minus 1. We shall be concerned with two specific complexes (we shall omit the prefix "abstract simplicial" from now on), the Čech and Vietoris–Rips complexes. Let $B_x(\varepsilon)$ denote the ball of radius ε around x, and $\Phi = \{x_1, x_2, \dots, x_m\}$ be a finite collection of points in $\mathbb{R}^d$.

DEFINITION 1.1 (Čech complexes). The complex $C(\Phi, \varepsilon)$, constructed according to the following rules, is called the Čech complex associated to Φ and ε:

(1) the 0-simplices of $C(\Phi, \varepsilon)$ are the points in Φ;

(2) an n-simplex or n-dimensional "face" $\sigma = [x_{i_0}, \dots, x_{i_n}]$ is in $C(\Phi, \varepsilon)$ if $\bigcap_{k=0}^{n} B_{x_{i_k}}(\varepsilon/2) \neq \emptyset$.

DEFINITION 1.2 (Vietoris–Rips complexes). The complex $R(\Phi, \varepsilon)$, constructed according to the following rules, is called the Vietoris–Rips complex associated to Φ and ε:

(1) the 0-simplices of $R(\Phi, \varepsilon)$ are the points in Φ;

(2) an n-simplex or n-dimensional "face" $\sigma = [x_{i_0}, \dots, x_{i_n}]$ is in $R(\Phi, \varepsilon)$ if $B_{x_{i_k}}(\varepsilon/2) \cap B_{x_{i_m}}(\varepsilon/2) \neq \emptyset$ for every $0 \le k < m \le n$.

The collection of all faces of dimension at most k is called the k-skeleton of a complex. Observe that the 1-skeletons of both the Čech and Vietoris–Rips complexes are equal and the same as the geometric graph on Φ with radius ε. More information on these complexes will be given in Section 4.1 when it is needed. Both of these (related) complexes are important in their own right, with the Čech complex being of particular interest since it is known to be homotopy equivalent to the random Boolean set $\bigcup_{x \in \Phi} B_x(\varepsilon)$, which appears in integral geometry (e.g., [36]) and continuum percolation (e.g., [28]). This homotopy equivalence follows from the nerve theorem [2], Theorem 10.7. We shall concentrate in this paper on the ranks of the homology groups—that is, the Betti numbers—of these complexes in the random scenario. At a heuristic level, the kth Betti number β_k measures the number of k-dimensional cycles or "holes" in the complex. As a consequence of the nerve theorem, $\beta_k = 0$ for $k \geq d$ for the Čech complex, and this is one of the distinguishing features of the Čech complex from that of the Vietoris–Rips complex.

A complementary approach to studying the topological structure of simplicial complexes is via (nonsmooth) Morse theory, and here results for Poisson process generated complexes are given in [6] via results on the Morse theory of the distance function. Contrasted with this is discrete Morse theory [16], which has also been used to study random complexes in [21, 22]. In fact, the local structure of Morse critical points (both nonsmooth and discrete) is often more amenable to computation than the global structure of the Betti numbers. Thus we shall also take this route in parts of this paper.

There are some recurring themes and techniques in the analysis of Betti numbers and Morse critical points, which are intimately related to the subgraph and component counts of the corresponding random geometric graph. Thus, from the purely technical side, much of this paper will be concerned with the intrinsically interesting task of extending the results of [35], Chapter 3, on subgraph and component counts of Poisson point processes to more general stationary point processes.

Subgraph counts of a random geometric graph are an example of U-statistics of point processes. Hence, apart from their applications in this article, our techniques to study subgraph counts of random geometric graph over general stationary point processes could be useful to derive asymptotics for many other translation and scale invariant U-statistics of point processes (e.g., the number of k-simplices in a Čech or Vietoris–Rips complexes). Also, the results on subgraph counts are used to derive results about clique numbers, maximum degree and chromatic number of the random geometric graph on Poisson or i.i.d. point process ([35], Chapter 6) and with a similar approach, our results can be used to derive asymptotics for clique numbers, maximum degree and chromatic number of random geometric graphs over general stationary point processes; see [3], Section 4.3.1.

Analysis of subgraph counts will take up all of Section 3, the longest section of the paper. From these results, we shall be able to extract results about Betti numbers (via combinatorial topology) as well as the numbers of Morse critical points.

In formulating results, we shall relate the topological features of the random simplicial complexes to known, inherent properties of the underlying point processes, including joint densities, void probabilities or Palm void probabilities. The first two of these properties, along with association properties, are known to be useful in studying measures of clustering, and their impact on percolation of random geometric graphs was studied in [4]. Since our asymptotic results help quantify the impact of clustering measures such as sub-Poisson and negative association on topological features of point processes, they provide additional applications of the tools of Błaszczyszyn and Yogeshwaran [5] as measures of clustering.

A sampler of some of our main results follows a little necessary notation.

1.1. *Some notation.* We use $|\cdot|$ to denote Lebesgue measure and $\|\cdot\|$ for the Euclidean norm on $\mathbb{R}^d$. Depending on context, $|\cdot|$ will also denote the cardinality of a set. As above, we denote the ball of radius r centred at $x \in \mathbb{R}^d$ by $B_x(r)$. For $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^{dk}$, let $B_{\mathbf{x}}(r) = \bigcup_{i=1}^k B_{x_i}(r)$, $h(\mathbf{x}) = h(x_1, \dots, x_k)$ for $h : \mathbb{R}^{dk} \to \mathbb{R}$ and $d\mathbf{x} = dx_1 \cdots dx_k$. Let $\mathbf{1} = (1, \dots, 1)$. We also use the standard Bachman–Landau notation for asymptotics[2] and say that a sequence of events $A_n, n \geq 1$ occurs *with high probability* (*w.h.p.*) if $\mathbb{P}\{A_n\} \to 1$ as $n \to \infty$.

1.2. *A result sampler.* We shall now describe, without (sometimes important) precise technical conditions, some of our main results. Full details are given in the main body of the paper. We start with Φ, a unit intensity, stationary point process on $\mathbb{R}^d$, and set[3]

$$\Phi_n = \Phi \cap \left[\frac{-n^{1/d}}{2}, \frac{n^{1/d}}{2}\right]^d. \tag{1.1}$$

Let

$$\beta_k\big(C(\Phi_n, r)\big), \qquad \beta_k\big(R(\Phi_n, r)\big),$$

[2]That is, for sequences a_n and b_n of positive numbers, we write

$$\begin{aligned}
a_n = o(b_n) &\iff \text{for any } c > 0 \text{, there is a } n_0 \text{ such that } a_n < cb_n \text{ for all } n > n_0;\\
a_n = O(b_n) &\iff \text{there exists a } c > 0 \text{ and a } n_0 \text{ such that } a_n < cb_n \text{ for all } n > n_0;\\
a_n = \omega(b_n) &\iff \text{for any } c > 0 \text{, there is a } n_0 \text{ such that } a_n > cb_n \text{ for all } n > n_0;\\
a_n = \Omega(b_n) &\iff \text{there exists a } c > 0 \text{ and a } n_0 \text{ such that } a_n > cb_n \text{ for all } n > n_0;\\
a_n = \Theta(b_n) &\iff a_n = O(b_n) \text{ and } a_n = \Omega(b_n).
\end{aligned}$$

[3]Note that our basic setup is a little different from that of all the earlier papers mentioned above. To compare our results with existing ones on Poisson or i.i.d. point processes, note that r_n^d in our results typically corresponds to nr_n^d elsewhere. For a general (non-Poisson) point process, (1.1) provides a more natural setting.

respectively, denote the kth Betti numbers of the Čech and Vietoris–Rips complexes based on Φ_n. Note that the β_k of a complex depends on the $(k+1)$ skeleton of the complex alone, and since the 1-skeletons are the same for both Čech and Vietoris–Rips complexes, we have that $\beta_0(C(\Phi_n,r)) = \beta_0(R(\Phi_n,r))$.

In addition, let $\mathcal{M}_k(\Phi_n)$ denote the set of Morse critical points (to be defined in Section 5.1) of index $k \in \{0,\dots,d\}$ for the distance function

$$d_n(x) = \min_{X \in \Phi_n} \|x - X\|,$$

and set

$$N_k(\Phi_n, r) = \bigl|\{c \in \mathcal{M}_k(\Phi_n) : d_n(c) \le r\}\bigr|.$$

The importance of the critical points stems from the Morse inequalities, which imply, in particular, that every index k critical point contributing to $N_k(\Phi_n,r)$ either increases $\beta_k(C(\Phi_n,r))$ by 1 or decreases $\beta_{k-1}(C(\Phi_n,r))$ by 1. In particular, this implies that $\beta_k(C(\Phi_n,r)) \le N_k(\Phi_n,r)$.

This paper is concerned with the behavior, as $n \to \infty$, of $\beta_k(C(\Phi_n,r_n))$, $\beta_k(R(\Phi_n,r_n))$, $N_k(\Phi_n,r_n)$ and $\chi(C(\Phi_n,r_n))$, where χ denotes the Euler characteristic. In particular, we shall provide closed form expressions for the asymptotic, normalized first moments of these variables, along with bounds for second moments for most of them.

Throughout the remainder of this subsection we shall assume that Φ is stationary, unit mean and negatively associated (defined rigorously in Section 2.2). Additional side conditions may also need to hold, but we shall not state them here. Two simple examples for which everything works are provided by the Ginibre point process and the simple perturbed lattice. Many of the results hold for various other sub-classes of point processes as well, but our nonspecific blanket assumptions allow for ease of exposition. We divide the results into three classes, depending on the behavior of r_n.

I. SPARSE REGIME: $r_n \to 0$. Note that since the points of Φ only generate edges and faces of the Čech and Vietoris–Rips complexes $C(\Phi_n,r)$ and $R(\Phi_n,r)$ when they are distance less than r apart, and since Φ has, on, average, only one point per unit cube, if r is small we expect that both of these complexes will be made up primarily of the isolated points of Φ. We describe this fact by calling this the "sparse" regime.

Since the β_0's are equal for the two complexes, in this setting,

$$\mathbb{E}\{\beta_0(C(\Phi_n,r_n))\} = \mathbb{E}\{\beta_0(R(\Phi_n,r_n))\} = \Theta(n),$$

and for $k \ge 1$, there exist functions $f^k \equiv 1$ [i.e., $f^k(r) = 1$, $\forall r$] or $f^k(r) \to 0$, as $r \to 0$, depending on the precise distribution of Φ and on the index k, such that

$$\begin{aligned} &\mathbb{E}\{\beta_k(C(\Phi_n,r_n))\} = \Theta(n r_n^{d(k+1)} f^{k+2}(r_n)), \qquad k \in \{0,\dots,d-1\},\\ (1.2)\qquad &\mathbb{E}\{\beta_k(R(\Phi_n,r_n))\} = \Theta(n r_n^{d(2k+1)} f^{2k+2}(r_n)), \qquad k \ge 1,\\ &\mathbb{E}\{N_k(\Phi_n,r_n)\} = \Theta(n r_n^{dk} f^{k+1}(r_n)), \qquad k \in \{0,\dots,d-1\}, \end{aligned}$$

and $\mathsf{Var}(N_k(\Phi_n, r_n)) = O(\mathbb{E}\{N_k(\Phi_n, r_n)\})$, where $\mathsf{Var}(X)$ is the variance of X. In addition, $\mathbb{E}\{n^{-1}\chi(C(\Phi_n, r_n))\} \to 1$.

In the classical Poisson case, studied in the references given above, it is known that the same results hold with $f^k \equiv 1$.

Using stochastic ordering techniques, we shall also show that clustering of point processes increases the functions $f^k(r)$ and consequently the mean of the β_k and N_k as well. Also, we know that for the Ginibre point process and for the zeroes of Gaussian entire functions, $f^k(r) = r^{k(k-1)}$. Thus there is a systematic difference between the scaling limits for Poisson and at least some negatively associated point processes.

II. THERMODYNAMIC REGIME: $r_n^d \to \beta \in (0, \infty)$. In this regime an edge between two points in Φ, which are, in a rough sense, an average distance of one unit apart, will be formed if they manage to get within a distance $\beta^{1/d}$ of one another. Since, in most scenarios, there should be a reasonable probability of this happening, we expect to see quite a few edges and, in fact, simplices and homologies up to dimension $d-1$. Indeed, this is the case, and the main result in this regime is that topological complexity grows at a rate proportional to the number of points, in the sense that

$$\mathbb{E}\{\beta_k(C(\Phi_n, r_n))\} = \Theta(n), \qquad k \in \{0, \dots, d-1\},$$

with identical results for $\mathbb{E}\{\beta_k(R(\Phi_n, r_n))\}$ and $\mathbb{E}\{N_k(\Phi_n, r_n)\}$ for the appropriate k. In addition, $\mathsf{Var}(N_k(\Phi_n, r_n)) = O(\mathbb{E}\{N_k(\Phi_n, r_n)\})$ and

$$\mathbb{E}\{n^{-1}\chi(C(\Phi_n, r_n))\} \to 1 + \sum_{k=1}^{d} (-1)^k \nu_k(\Phi, \beta),$$

where the $\nu_k(\Phi, \beta)$ are defined in Theorem 5.2. Since there is no appearance in these results of an analogue to the f of (1.2), the normalizations here have the same orders as in the Poisson and i.i.d. cases.

III. CONNECTIVITY REGIME: $r_n^d = \Theta(\log n)$. Clearly, if r_n is large enough, there comes a point (which we call the *contractibility radius*) beyond which each point of Φ_n will be connected to the others, and the Čech complex will become contractible to a single point, while the Vietoris–Rips complex will become topologically k-connected. (This is certainly the case if $r_n = \sqrt{d}n^{1/d}$.) The question then is "how large is large enough?"

It turns out that in the current scenario of negative association there exist case dependent constants C such that for $r_n \geq C(\log n)^{1/d}$, $C(\Phi_n, r_n)$ is contractible w.h.p. as $n \to \infty$. In the specific cases of the Ginibre process or zeroes of Gaussian entire functions, this happens earlier, and $r_n = \Theta((\log n)^{1/4})$ is the radius for contractibility of the Čech complex. As a trivial corollary, it follows that, w.h.p. $\chi(C(\Phi_n, r_n)) = 1$ when r_n is the radius of contractibility. Further, for the Ginibre process, $r_n = \Theta((\log n)^{1/4})$ is also the critical radius for k-connectedness of the Vietoris–Rips complex.

1.3. *Some implications for topological data analysis.* Perhaps the core tool of TDA is persistent homology, as visualized through barcodes and persistence diagrams; cf. [11, 15, 19, 42]. See also the very accessible recent survey [12]. While here is not the place to go into the details of persistent homology, it can be described reasonably simply in the setting of this paper. For a given n, and a collection of points Φ_n, consider the collections of Čech (or Vietoris–Rips) complexes $C(\Phi_n, r)$ built over these points, as r grows. Initially, $C(\Phi_n, 0)$ will contain only the points of Φ_n. However, as r increases, different homological entities (cycles of differing degree) will appear and, eventually, disappear. If to each such phenomenon we assign an interval starting at the birth time and ending at the death time, then the collection of all of these intervals is a representation of the persistent homology generated by Φ_n and is known as its *barcode*. The individual intervals are referred to as *bars*. The Betti numbers $\beta_k(C(\Phi_n, r))$ therefore count the number of bars related to k-cycles active at "connection distance" r.

Heuristics and simulations[4] (see Figures 1, 2, 3) indicate that as the points are more regularly distributed in a point process, the bars start later and vanish earlier than those for Poisson point process. In the three figures, all point processes are of unit intensity, and we observe that the hypergeometric perturbed lattice has more regularly distributed points than the Poisson point process, which in turn has more regularly distributed points than the negative binomial perturbed lattice. We can see that the corresponding bars start earlier and end later as we go from hypergeometric perturbed lattice to the Poisson point process to the negative binomial perturbed

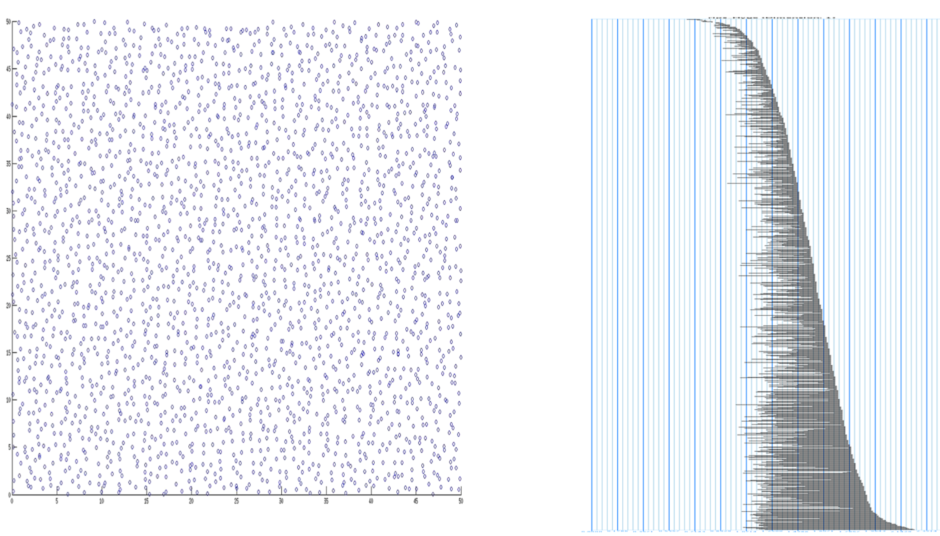

FIG. 1. *Hypergeometric perturbed lattice. H_1 barcodes of the Rips complex.*

[4]These barcodes were simulated using the easy-to-use and open access package javaPlex [39].

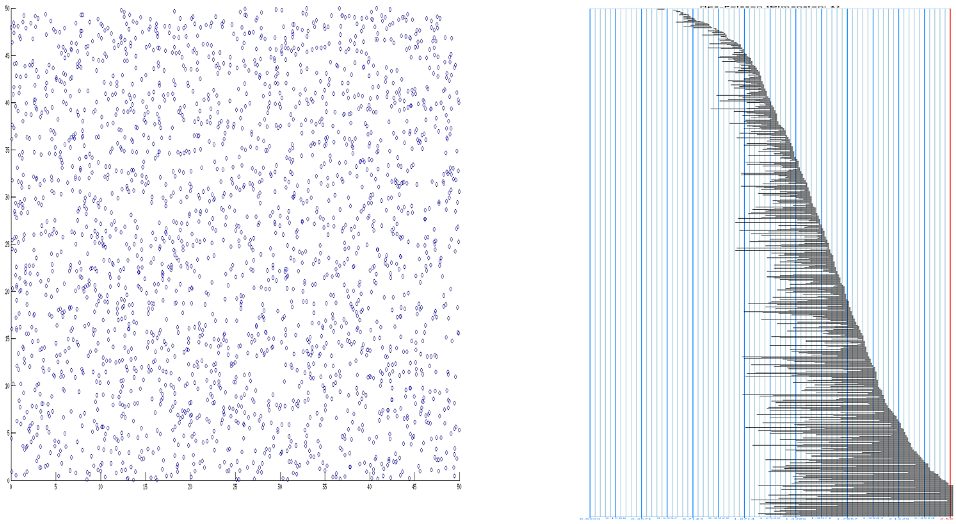

FIG. 2. *Poisson point process. H_1 barcodes of the Rips complex.*

lattice. Some of our results confirm this heuristic. For example, using the results above it is easy to see that nontrivial homology groups of Čech and Vietoris–Rips complexes start to appear once r_n satisfies $r_n^{d(k+1)} f^{k+2}(r_n) = \omega(n^{-1})$. For the Poisson case this requires only $r_n = \omega(n^{-1/d(k+1)})$. Since, typically, $f(r) \to 0$ as $r \to 0$, we therefore generally need larger radii for nontrivial homology (and hence for bars) to appear. The disappearance of homology is harder, however, and

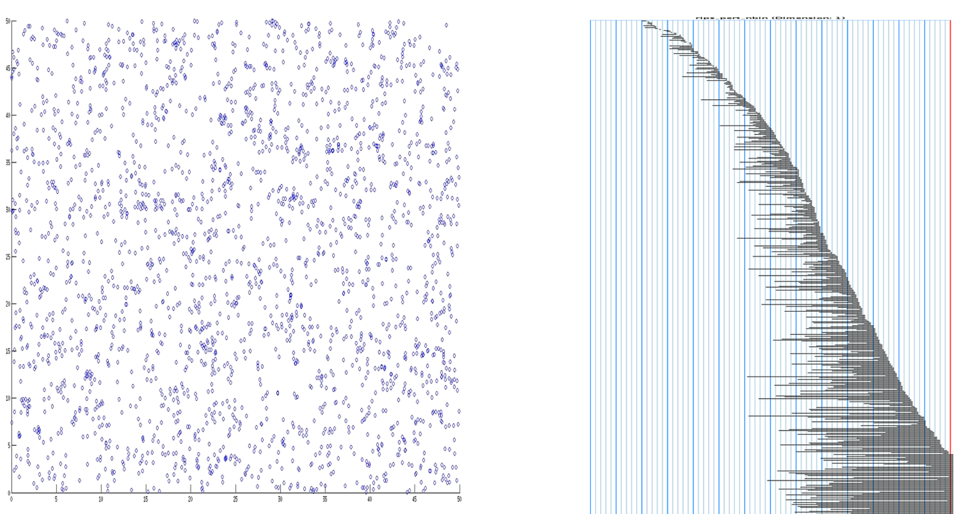

FIG. 3. *Negative Binomial perturbed lattice. H_1 barcodes of the Rips complex.*

in general, our results on connectivity cannot confirm the heuristic. However, for the Ginibre point process and zeroes of GEF in $\mathbb{R}^2$, they do show that nontrivial topology vanishes at $r_n = \omega((\log n)^{1/4})$ as opposed to $\omega((\log n)^{1/2})$ for a two-dimensional Poisson point process of constant intensity.

As for implications to TDA, applied topologists are beginning to appreciate the fact that stochasticity underlies their data as a consequence of sampling, and are beginning to build statistical models to allow parameter estimation and inference (e.g., [8, 9, 13, 30, 40]). The results of this paper show that small changes in model structure (such as the introduction of attraction and repulsion between points in a data cloud) can have measurable effects on topological behavior.

The remainder of the paper is organized as follows: in the following section, we shall summarize some facts needed from the theory of point processes. Sections 3, 4 and 5 contain the core technical results on component and subgraph counts, Betti numbers and Morse critical points, respectively. We shall give careful proofs for all the results of Section 3 barring the results on extension to subcomplex counts in Section 3.6 since these mimic earlier proofs. The results of the Sections 4 and 5 are either easy corollaries of earlier results or can be proved by using similar techniques, and so there we shall give less detail. Appendix contains a technical result regarding Palm void probabilities of the Ginibre process which Manjunath Krishnapur proved for us.

2. Point processes. Our aim in this section is to set up some general definitions related to point processes, give some background on those of main interest to us and to prove two technical results, of some independent interest, which we shall need later.

2.1. *Point processes and Palm measures.* A point process Φ in $\mathbb{R}^d$ is a $\mathcal{N}$-valued random variable, where $\mathcal{N}$ is the space of locally finite (Radon) counting measures in $\mathbb{R}^d$ equipped with the canonical σ-algebra; cf. [25, 36, 38]. We can represent Φ as either a random measure, $\Phi(\cdot) = \sum_i \delta_{X_i}(\cdot)$ or as a random point set $\Phi = \{X_i\}_{i \ge 1}$, where, in both cases, the X_i are the "points" of the process.

The factorial moment measure $\alpha^{(k)}$ of a point process Φ is defined by

$$\alpha^{(k)}\left(\prod_{i=1}^{n} B_i\right) = \mathbb{E}\left\{\prod_{i=1}^{n} \Phi(B_i)\right\},$$

for disjoint bounded Borel subsets $B_1, \dots, B_n$. When $k = 1$, $\alpha := \alpha^{(1)}$ is called the *intensity* or *mean measure*, and $\alpha^{(k)}$ also serves as the intensity measure of the point process

$$\Phi^{(k)} := \{(X_1, \dots, X_k) \in \Phi^k : X_i \neq X_j, \forall i \neq j\}.$$

The *kth joint intensity*, $\rho^{(k)} : (\mathbb{R}^d)^k \to [0, \infty)$ is the density (if it exists) of α^k with respect to (in this paper) Lebesgue measure. The $\rho^{(k)}$ characterize a simple point process just as moments characterize a random variable. A sufficient condition for joint intensities (when they exist) to characterize a simple point process is $\rho^{(k)}(\cdot) \leq C^k$ for some constant C and for all $k \geq 1$; cf. [20], Lemma 4.2.6 and Remark 1.2.4. Throughout, we shall restrict ourselves to simple stationary point processes of unit intensity; namely, $\alpha(B) = |B|$ for all bounded, Borel B. We also shall assume that all the joint intensities $\rho^{(k)}(\cdot)$ exist for the point processes under consideration in this article.

For a point process Φ whose probability distribution is $\mathbb{P}$, its reduced *Palm probability distribution* $\mathbb{P}_x^!$ at $x \in \mathbb{R}^d$ is defined as the probability measure that satisfies the following disintegration formula for any bounded measurable function $u : \mathcal{N} \times \mathbb{R}^d \to \mathbb{R}_+$ with compact support in the second co-ordinate:

$$\int_{\mathcal{N}} \mathbb{P}(d\phi) \int_{\mathbb{R}^d} \phi(dx) u(\phi, x) = \int_{\mathbb{R}^d} dx \int_{\mathcal{N}} \mathbb{P}_x^!(d\phi) u\big(\phi \cup \{x\}, x\big).$$

As a consequence of the above definition, for the corresponding Palm expectation $\mathbb{E}_x^!$ with the function u satisfying assumptions as above, we get the well-known *refined Campbell theorem* (cf. [38], page 119, [36], Theorem 3.3),

$$\mathbb{E}\Bigg\{\sum_{X \in \Phi} u(\Phi, X)\Bigg\} = \int_{x \in \mathbb{R}^d} \mathbb{E}_x^! \big\{u\big(\Phi \cup \{x\}, x\big)\big\}\, dx. \tag{2.1}$$

If the point process is not stationary or has unit intensity, one can still define Palm probability distribution by replacing dx on the RHS of the above two equations with the intensity measure of the point process. In particular, the definition of Palm probability gives us that $\mathbb{P}_x^!\{\Phi(x) = 0\} = 1$. Intuitively, $\mathbb{P}_x^!$ is the distribution of the remainder of the point process, conditioned on there having been a point at x.

2.2. *Some special cases*. We shall assume the reader is familiar with stationary Poisson point processes, determined, for example, by $\rho^k \equiv 1$ for all k, and use this as a basis for comparison in a quick tour through some non-Poisson cases that will provide examples for the theorems of the remaining sections.

Associated point processes. A point process Φ is called *associated* (or *positively associated*) if for any finite collection of disjoint bounded Borel subsets $B_1, \ldots, B_k \subset \mathbb{R}^d$ and f, g continuous and increasing functions taking values in $[0, 1]$,

$$\mathsf{Cov}\big(f\big(\Phi(B_1), \ldots, \Phi(B_k)\big), g\big(\Phi(B_1), \ldots, \Phi(B_k)\big)\big) \geq 0; \tag{2.2}$$

cf. [10]. The referenced article gives many examples of associated processes. We call a point process Φ *negatively associated* if

$$\mathsf{Cov}\big(f\big(\Phi(B_1), \ldots, \Phi(B_k)\big), g\big(\Phi(B_{k+1}), \ldots, \Phi(B_l)\big)\big) \leq 0, \tag{2.3}$$

for any finite collection of bounded Borel subsets $B_1, \dots, B_l \subset \mathbb{R}^d$ such that $(B_1 \cup \cdots \cup B_k) \cap (B_{k+1} \cup \cdots \cup B_l) = \emptyset$ and f, g increasing bounded continuous functions.

In general, the literature contains fewer examples of negatively associated processes than their positively associated counterparts, a phenomenon that occurs even in simpler situations; cf. [33]. We shall give two examples of negatively associated point processes below (determinantal point processes and the simple perturbed lattice) as these are of more interest to us in this article, but we refer the reader to [3, 10] for many examples of positively associated point processes. The stationary Poisson point process is both negatively and positively associated. Finite independent unions of negatively associated point processes is negatively associated as well, and this can be used to construct many examples of negatively associated point processes from a few simple examples. Just to reiterate the earlier point about scarcity of negatively associated point processes, not many "natural" examples, apart from the two presented below and binomial point process, are known. This is in contrast to the situation for positively associated point processes. Here are three other point processes of interest to us:

Determinantal processes. A simple point process Φ on $\mathbb{R}^d$ is said to be determinantal with kernel $K : (\mathbb{R}^d)^2 \to \mathbb{C}$ if its joint intensities satisfy the following equality for all $k \geq 1$ and for all $x_1, \dots, x_k \in \mathbb{R}^d$:

$$\rho^k(x_1, \dots, x_k) = \det\bigl(K(x_i, x_j)_{1 \leq i, j \leq k}\bigr), \tag{2.4}$$

where det indicates a determinant of a matrix.

Stationary determinantal point processes with continuous kernels are negatively associated [18], Corollary 6.3. For examples of stationary determinantal point processes, see [26], Section 5. A determinantal process of particular interest is the unit intensity *Ginibre process* ([20], Section 4.3.7), which has the continuous kernel

$$K(z, w) = \exp\bigl(-\tfrac{1}{2}\bigl(\|z\|^2 + \|w\|^2\bigr) + z\overline{w}\bigr), \qquad z, w \in \mathbb{C}.$$

In [31], the authors have introduced a family of determinantal point processes called the α-Ginibre point processes in the context of modeling cellular networks with $\alpha = 1$ corresponding to the Ginibre point process, and as $\alpha \to 0$, the α-Ginibre point processes converges to the appropriate Poisson point process. This class of point processes gives a continuous family of point processes between the Poisson point process and the Ginibre point process.

A counterpart to determinantal point processes are permanental point processes, which can be defined by replacing the determinant in (2.4) by a matrix permanent.

Perturbed lattices. Let $N_z : z \in \mathbb{Z}^d$ be i.i.d. integer valued random variables distributed as N, and $X_{iz}, i \geq 1, z \in \mathbb{Z}^d$ be i.i.d. $\mathbb{R}^d$ valued random variables distributed as X. A perturbed lattice is defined as

$$\Phi(N, X) := \bigcup_{z \in \mathbb{Z}^d} \bigcup_{i=1}^{N_z} \{z + X_{iz}\},$$

provided that $\Phi(N, X)$ is a simple point process. N is called the *replication kernel*, and X is called the *perturbation kernel*. Though the point process is stationary with respect to lattice translations only, we can make it stationary with respect to $\mathbb{R}^d$ translations by shifting the origin uniformly within $[0, 1]^d$; that is, $\bigcup_{z\in\mathbb{Z}^d}\bigcup_{i=1}^{N_z}\{U + z + X_{iz}\}$ is stationary if U is uniformly distributed in $[0, 1]^d$. The point process for which $N \equiv 1$ and X is uniform on the unit cube is known as the *simple perturbed lattice* and is negatively associated. For more details, see below and especially [5].

Zeroes of a Gaussian entire function. (Normalized) Gaussian entire functions are defined on the complex plane $\mathbb{C}$ via the a.s. convergent expansion $f(z) = \sum_{n=0}^{\infty}\xi_n z^n/\sqrt{n!}$, where the ξ_n are i.i.d. standard complex Gaussians. The zeros of f (when considered as a point process in $\mathbb{R}^2$ and called as zeros of GEF), while neither negatively associated nor determinantal, share many properties with the Ginibre process that make them interesting and tractable; cf. [20] for more background.

Sub- and super-Poisson processes. At times, weaker notions than association, based only on factorial moment measures, suffice to establish interesting results.

We say that a point process Φ_1 is *α-weaker* than Φ_2 (written $\Phi_1 \leq_{\alpha-w} \Phi_2$) if $\alpha_1^{(k)}(B) \leq \alpha_2^{(k)}(B)$ for all $k \geq 1$ and bounded Borel $B \subset (\mathbb{R}^d)^k$. We call a point process *α-negatively associated* (*associated*) if $\alpha^{(k+l)}(B_1 \times B_2) \leq (\geq)$ $\alpha^{(k)}(B_1)\alpha^{(l)}(B_2)$ for all $k, l \geq 1$ and bounded Borel $B_1 \times B_2 \subset (\mathbb{R}^d)^k \times (\mathbb{R}^d)^l$.

Negative association (association) implies α-negative association (association) which in turn implies α-weaker ordering with respect to the Poisson process with intensity measure α.

Even weaker notions of association come from looking at void probabilities, and we say that a point process Φ_1 is *ν-weaker* than Φ_2 (denoted by $\Phi_1 \leq_{\nu-w} \Phi_2$) if

$$\nu_1(B) = \mathbb{P}\{\Phi_1(B) = \varnothing\} \leq \mathbb{P}\{\Phi_2(B) = \varnothing\} = \nu_2(B)$$

for all B bounded Borel subsets.

Finally, we call a point process *α-sub-Poisson* (*super-Poisson*) if it is α-weaker (stronger) than the Poisson point process and similarly for *ν-sub-Poisson* (*super-Poisson*). A point process is *weakly sub-Poisson* (*super-Poisson*) if it is both α- and ν-sub-Poisson (super-Poisson).

It is known that negative association (association) implies the weak sub-Poisson (super-Poisson) property. Other examples come from perturbed lattices. For example, if the replication kernel N is hypergeometric or binomial and X uniform, then the resulting perturbed lattice $\Phi(N, X)$ is a weakly sub-Poisson point process. One can also construct a sequence of perturbed lattices $\Phi(N_n, X), n \geq 1$ whose joint intensities and void probabilities monotonically increase to that of the

Poisson point process by choosing the replication kernels N_n to be distributed as $\mathrm{Bin}(n, \frac{1}{n})$. On the other hand, negative binomial and geometric perturbation kernels lead to weakly super-Poisson processes. Permanental point processes are also weakly sub-Poisson. See [5] for proofs and more about stochastic ordering of point processes.

2.3. *Two technical lemmas.* We shall state some general results about Palm measures of these point processes that we need later. The first lemma shows that negatively associated point processes are "stochastically stronger" than their Palm versions. This can also be viewed as a justification for the usage of negative association as the defining property of sparse point processes. The second shows that Palm versions of negatively associated point processes also exhibit negative association. We state the results in more generality than we need, since they seem to be of independent interest.

LEMMA 2.1. *Let Φ be a negatively associated stationary point process in $\mathbb{R}^d$ of unit intensity and $F:\mathbb{R}^{dn} \to \mathbb{R}_+$ an increasing bounded continuous function. Then for $B_1, \dots, B_n$ disjoint bounded Borel subsets and almost every $\mathbf{x} \in \mathbb{R}^{dk}$,*

$$\mathbb{E}^!_{x_1,\dots,x_k}\big(F\big(\Phi(B_1), \dots, \Phi(B_n)\big)\big) \le \mathbb{E}\big\{F\big(\Phi(B_1), \dots, \Phi(B_n)\big)\big\}. \tag{2.5}$$

The above inequality will be reversed for an associated point process.

PROOF. For $0 < \epsilon < r$ we have

$$\begin{aligned}
&\mathbb{E}\big\{F\big(\Phi\big(B_1 \setminus B_{\mathbf{x}}(r)\big), \dots, \Phi\big(B_n \setminus B_{\mathbf{x}}(r)\big)\big)|\Phi\big(B_{x_i}(\epsilon)\big) \ge 1, 1 \le i \le k\big\} \\
&\quad = \frac{\mathbb{E}\{F(\Phi(B_1 \setminus B_{\mathbf{x}}(r)), \dots, \Phi(B_n \setminus B_{\mathbf{x}}(r))) \prod_{i=1}^k \mathbb{1}[\Phi(B_{x_i}(\epsilon)) \ge 1]\}}{\mathbb{P}\{\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le k\}} \\
&\quad \le \frac{\mathbb{E}\{F(\Phi(B_1 \setminus B_{\mathbf{x}}(r)), \dots, \Phi(B_n \setminus B_{\mathbf{x}}(r)))\}\mathbb{E}\{\prod_{i=1}^k \mathbb{1}[\Phi(B_{x_i}(\epsilon)) \ge 1]\}}{\mathbb{P}\{\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le k\}} \\
&\quad = \mathbb{E}\big\{F\big(\Phi\big(B_1 \setminus B_{\mathbf{x}}(r)\big), \dots, \Phi\big(B_n \setminus B_{\mathbf{x}}(r)\big)\big)\big\},
\end{aligned}$$

where the inequality is due to the negative association property of Φ.

Sending first $\epsilon \to 0$ and then $r \to 0$, (2.5) follows immediately from [37], Lemma 6.3, and monotone convergence. □

LEMMA 2.2. *Let Φ be a negatively associated stationary point process in $\mathbb{R}^d$ of unit intensity. We shall also assume the existence of all the joint intensities of the point process. Let $F:\mathbb{R}^{dn} \to \mathbb{R}_+$ and $G:\mathbb{R}^{dm} \to \mathbb{R}_+$ be increasing bounded continuous functions. Then for $B_1, \dots, B_{m+n}$ disjoint bounded Borel subsets and* *almost every* $\mathbf{x} \in \mathbb{R}^{d(k+l)}$,

$$\begin{aligned}
&\mathbb{E}^!_{\mathbf{x}}\{F(\Phi(B_1), \dots, \Phi(B_n))G(\Phi(B_{n+1}), \dots, \Phi(B_{m+n}))\}\rho^{(k+l)}(\mathbf{x}) \\
&\quad \le \mathbb{E}^!_{x_1,\dots,x_k}\{F(\Phi(B_1), \dots, \Phi(B_n))\} \\
&\qquad \times \mathbb{E}^!_{x_{k+1},\dots,x_{k+l}}\{G(\Phi(B_{n+1}), \dots, \Phi(B_{m+n}))\} \\
&\qquad \times \rho^{(k)}(x_1, \dots, x_k)\rho^{(l)}(x_{k+1}, \dots, x_{k+l}).
\end{aligned}$$

The above inequality will be reversed for an associated point process Φ.

PROOF. As in the proof of Lemma 2.1, take $0 < \epsilon < r$. For notational simplicity, set $B^* = B \setminus B_{\mathbf{x}}(r)$ for bounded Borel set B:

$$\begin{aligned}
&\mathbb{E}\{F(\Phi(B_1^*), \dots, \Phi(B_n^*))G(\Phi(B_{n+1}^*), \dots, \Phi(B_{m+n}^*)) \\
&\qquad\qquad\qquad |\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le (k+l)\} \\
&\quad \times \mathbb{P}\{\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le (k+l)\} \\
&\quad = \mathbb{E}\Bigg\{F(\Phi(B_1^*), \dots, \Phi(B_n^*)) \prod_{i=1}^k \mathbb{1}[\Phi(B_{x_i}(\epsilon)) \ge 1] \\
&\qquad\qquad \times G(\Phi(B_{n+1}^*), \dots, \Phi(B_{m+n}^*)) \prod_{i=1}^l \mathbb{1}[\Phi(B_{x_{k+i}}(\epsilon)) \ge 1]\Bigg\} \\
&\quad \le \frac{\mathbb{E}\{F(\Phi(B_1^*), \dots, \Phi(B_n^*)) \prod_{i=1}^k \mathbb{1}[\Phi(B_{x_i}(\epsilon)) \ge 1]\}}{\mathbb{P}\{\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le k\}} \\
&\qquad \times \frac{\mathbb{E}\{G(\Phi(B_{n+1}^*), \dots, \Phi(B_{m+n}^*)) \prod_{i=1}^l \mathbb{1}[\Phi(B_{x_{k+i}}(\epsilon)) \ge 1]\}}{\mathbb{P}\{\Phi(B_{x_{k+i}}(\epsilon)) \ge 1, 1 \le i \le l\}} \\
&\qquad \times \mathbb{P}\{\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le k\}\mathbb{P}\{\Phi(B_{x_{k+i}}(\epsilon)) \ge 1, 1 \le i \le l\} \\
&\quad = \mathbb{E}\{F(\Phi(B_1^*), \dots, \Phi(B_n^*))|\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le k\} \\
&\qquad \times \mathbb{E}\{G(\Phi(B_{n+1}^*), \dots, \Phi(B_{m+n}^*))|\Phi(B_{x_{k+i}}(\epsilon)) \ge 1, 1 \le i \le l\} \\
&\qquad \times \mathbb{P}\{\Phi(B_{x_i}(\epsilon)) \ge 1, 1 \le i \le k\}\mathbb{P}\{\Phi(B_{x_{k+i}}(\epsilon)) \ge 1, 1 \le i \le l\},
\end{aligned}$$

where the inequality is due to the negative association of Φ. As in the previous proof, the conditional expectations in the first and last expressions converge to the respective Palm expectations as $\epsilon \to 0$. Since Φ is a simple point process, after dividing by $|B_0(\epsilon)|^{k+l}$ on both sides, the product of the probability terms in the last line converges to $\rho^{(k)}(x_1, \dots, x_k)\rho^{(l)}(x_{k+1}, \dots, x_{k+l})$ while the probability term in the first line converges to $\rho^{(k+l)}(\mathbf{x})$ as $\epsilon \to 0$. Complete the proof by sending $r \to 0$. $\square$

3. Subgraph and component counts in random geometric graphs. Recall that for a point set Φ and radius $r > 0$, the geometric graph $G(\Phi, r)$ is defined as the graph with vertex set Φ and edge-set $\{(X, Y) : \|X - Y\| \leq r\}$. We shall work with restrictions of Φ to a sequence of increasing windows $W_n = [\frac{-n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$, along with a radius regime $\{r_n > 0\}_{n\geq 1}$, setting $\Phi_n := \Phi \cap W_n$. The choice of the radius regime will impact on the asymptotic properties of the geometric graph when the points of Φ are those of a point process.

Let Γ be a connected graph on k vertices. In this section we shall be interested in how often Γ appears (up to graph isomorphisms) in a sequence of geometric graphs $G_n = G(\Phi_n, r_n)$, and how often among such appearances it is actually isomorphic to a component of G_n; namely, it is a Γ-component of G_n. For graphs built over Poisson and i.i.d. processes, we know from [35], Chapters 3, 13, that no Γ-components exist when $n(r_n^d)^{k-1} \to 0$ ($|\Gamma| = k$), but that they do appear when $n(r_n^d)^{k-1} \to \infty$. The Γ-components continue to exist even when $r_n^d = o(\log n)$ and vanish when $r_n^d = \omega(\log n)$, which is the threshold for connectivity of the graph.

In this section, we shall show, among other things, that the threshold for formation of Γ-components for negatively associated processes with $r_n \to 0$ is $n(r_n^d)^{k-1} f^k(r_n) \to \infty$, for functions f^k which typically satisfy $f^k(r) \to 0$ as $r \to 0$, and so is higher than in the Poisson case. These components continue to exist even when $r_n^d \to \beta > 0$. The threshold for the vanishing of components will be treated in the next section.

The reader should try to keep this broader picture in mind as she wades through the various limits of this section.

3.1. *Some notation and a start.* As above, let Γ be a connected graph on k vertices, $k \geq 1$ and $\{x_1, \ldots, x_k\}$ a collection of k points in $\mathbb{R}^d$. Introduce the (indicator) function $h_\Gamma : \mathbb{R}^{dk} \times \mathbb{R}_+ \to \{0, 1\}$ by

$$h_\Gamma(\mathbf{x}, r) := \mathbb{1}\big[G(\{x_1, \ldots, x_k\}, r) \simeq \Gamma\big], \tag{3.1}$$

where $\simeq$ denotes graph isomorphism and $\mathbb{1}$ is the usual indicator function. For a fixed sequence $\{r_n\}$ set

$$h_{\Gamma,n}(\mathbf{x}) := h_\Gamma(\mathbf{x}, r_n), \tag{3.2}$$

and, for $r = 1$, write

$$h_\Gamma(\mathbf{x}) := h_\Gamma(\mathbf{x}, 1). \tag{3.3}$$

Moving now to the random setting, in which Φ is a simple point process with kth intensities $\rho^{(k)}$, we say that Γ is *a feasible subgraph* of Φ if

$$\int_{(\mathbb{R}^d)^k} h_\Gamma(\mathbf{x}) \rho^{(k)}(\mathbf{x})\, d\mathbf{x} > 0.$$

Thus Γ is a feasible subgraph of Φ if the $\alpha^{(k)}$ measure of finding a copy of it (up to graph isomorphism) in $G(\Phi, 1)$ is positive. For many of our examples of point processes and graphs Γ, feasibility will hold because $\rho^{(k)}(\mathbf{x}) > 0$ a.e. or at least on a large enough set.

We shall be interested in the the number of Γ-subgraphs, $G_n(\Phi, \Gamma)$, and number of Γ-components, $J_n(\Phi, \Gamma)$, of Φ_n, which are defined as follows:

$$\begin{aligned} G_n(\Phi, \Gamma) &:= \frac{1}{k!} \sum_{X \in \Phi_n^{(k)}} h_{\Gamma,n}(X), \\ J_n(\Phi, \Gamma) &:= \frac{1}{k!} \sum_{X \in \Phi_n^{(k)}} h_{\Gamma,n}(X) \mathbb{1}\big[\Phi_n\big(B_X(r_n)\big) = k\big]. \end{aligned} \tag{3.4}$$

We shall now make a small digression to clarify the terminology. In the terminology of [35], $G_n(\Phi, \Gamma)$ is referred to as the number of induced Γ-subgraphs of $G(\Phi, r)$ and not the number of Γ-isomorphic subgraphs. However, it is easy to see that the latter is a finite linear combination of the number of induced subgraphs of the same order. We shall be considering only induced subgraphs in this article and hence shall chose to omit the adjective *induced*.

Note that J_n considers graphs based on vertices in Φ_n only, namely, all vertices that lie in W_n. Such a graph, however, may have vertices in the complement of W_n, provided the points are within distance a r_n of W_n, and so actually be part of something larger. To account for this boundary effect, we introduce an additional variable, which does not count such "boundary crossing" graphs. This is given by

$$\tilde{J}_n(\Phi, \Gamma) := \frac{1}{k!} \sum_{X \in \Phi_n^{(k)}} h_{\Gamma,n}(X) \mathbb{1}\big[\Phi\big(B_X(r_n)\big) = k\big]. \tag{3.5}$$

We shall see later that in the sparse and thermodynamic regimes, the differences between J_n and $\tilde{J}_n$ disappear in asymptotic results. Nevertheless, both are needed for the proofs.

The key ingredient in obtaining asymptotics for sub-graph counts and component counts are the following closed-form expressions, which are immediate consequences of the Campbell–Mecke formula:

$$\mathbb{E}\big\{G_n(\Phi, \Gamma)\big\} = \frac{1}{k!} \int_{W_n^k} h_{\Gamma,n}(\mathbf{x}) \rho^{(k)}(\mathbf{x})\, d\mathbf{x}, \tag{3.6}$$

$$\mathbb{E}\big\{J_n(\Phi, \Gamma)\big\} = \frac{1}{k!} \int_{W_n^k} h_{\Gamma,n}(\mathbf{x}) \mathbb{P}^!_{\mathbf{x}}\big\{\Phi_n\big(B_{\mathbf{x}}(r_n)\big) = 0\big\} \rho^{(k)}(\mathbf{x})\, d\mathbf{x}. \tag{3.7}$$

Much of the remainder of this section is based on obtaining asymptotic expressions for these integrals in terms of basic point process parameters in the sparse and thermodynamic regimes, as well as looking at bounds on variances. We shall consider the connectivity regime only in the following section on Betti numbers. Our results here extend those of [35], Chapter 3, for Poisson and i.i.d. processes, and the general approach of the proofs is thus similar.

3.2. *Sparse regime*: $r_n \to 0$. The intuition behind the following theorem is that in the sparse regime it is difficult to find Γ-subgraphs in a random geometric graph, and even more unlikely that any such subgraph will have another point of the point process near it. This implies that almost all [in the sense made precise by (3.8)] such subgraphs will actually be a component of the full graph, disconnected from other components.

THEOREM 3.1. *Let Φ be a stationary point process in $\mathbb{R}^d$ of unit intensity and Γ be a feasible connected graph of Φ on k vertices. Let $\rho^{(k)}$ be almost everywhere continuous. Assume that $\rho^{(k)}(0,\dots,0)=0$, and that there exist functions $f_\rho^k:\mathbb{R}_+\to\mathbb{R}_+$ and $g_\rho^k:(B_0(k))^k\to\mathbb{R}_+$ such that*

$$\rho^{(k)}(r\mathbf{y})=\Theta\big(f_\rho^k(r)\big) \quad \textit{and} \quad \lim_{r\to 0}\frac{\rho^{(k)}(r\mathbf{y})}{f_\rho^k(r)}=g_\rho^k(\mathbf{y}),$$

for all $\mathbf{y}$ of the form $\mathbf{y}=(0,y_2,\dots,y_k)$. Further, assume that $f^{k+1}=O(f^k)$ as $r\to 0$ and g_ρ^k is almost everywhere continuous. Let $r_n\to 0$. Then

$$\begin{aligned}\lim_{n\to\infty}\frac{\mathbb{E}\{G_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)}f^k(r_n)} &= \lim_{n\to\infty}\frac{\mathbb{E}\{J_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)}f^k(r_n)}\\ &= \mu_0(\Phi,\Gamma) \qquad (3.8)\\ &:= \begin{cases}1, & k=1,\\ \dfrac{1}{k!}\displaystyle\int_{\mathbb{R}^{d(k-1)}}h_\Gamma(\mathbf{y})g_\rho^k(\mathbf{y})\,d\mathbf{y}, & k\ge 1.\end{cases}\end{aligned}$$

If $\rho^{(k)}(0,\dots,0)>0$, then the same result holds with $f_\rho^k\equiv 1$ and $g_\rho^k\equiv\rho^{(k)}(0,\dots,0)$.

Before turning to the proof of the theorem, we shall make a few points about its conditions, and provide some examples. As before, we are assuming that all point processes are normalized to have unit intensity.

REMARK 3.2. (1) Note that the theorem does not guarantee the positivity of $\mu_0(\Phi,\Gamma)$.

(2) $f^1(r)\equiv 1$ for all stationary point processes of unit intensity since, in this case, $\rho^{(1)}\equiv 1$.

(3) It is easy to check that if Φ is α-negatively associated or α-super-Poisson, then the condition $f^{k+1}=O(f^k)$ as $r\to 0$ is satisfied.

(4) In the case $\rho^k(0,\dots,0)=0$ for $k\ge 2$, even if we cannot find appropriate f^k or g_ρ^k, it is still true that $\mathbb{E}\{G_n(\Phi,\Gamma)\}=o(nr_n^{d(k-1)})$.

(5) If Φ is only $\mathbb{Z}^d$-stationary (as is the case with perturbed lattices), then it will be clear from the proof that (3.8) still holds, but with

$$\mu_0(\Phi,\Gamma):=\frac{1}{k!}\int_{[0,1]^d}\int_{\mathbb{R}^{d(k-1)}}h_\Gamma(x,\mathbf{y})g_\rho^k(x,\mathbf{y})\,dx\,d\mathbf{y}.$$

(6) For a homogeneous Poisson point process, the theorem holds with $f^k \equiv 1$ and $g_\rho^k \equiv 1$, recovering [35], Proposition 3.1.

(7) If $\Phi \geq_{\alpha-w} \Phi_{(1)}$, then for all $k \geq 1$, $\rho^{(k)} \geq \rho_{(1)}^{(k)} \equiv 1$ and hence $f^k \equiv 1$ and also $\mu_0(\Phi, \Gamma) > 0$. Examples of point processes in this class are all super-Poisson perturbed lattices and permanental point processes.

(8) For a perturbed lattice Φ with perturbation kernel $N \in \{0, \dots, K\}$ a.s., $\rho^{(k)}(0, \dots, 0) > 0$ if and only if $k \leq K$. In this case, $\mu_0(\Phi, \Gamma) > 0$ for a connected graph Γ on k vertices. For connected graphs Γ on k vertices with $k > K$, $nr_n^{-d(k-1)}\mathbb{E}\{G_n(\Phi, \Gamma)\} \to 0$. For sub-Poisson perturbed lattices, the existence of f^k depends on the perturbation kernel. However, for high values of k, it is clear that the scaling for sub-Poisson perturbed lattices will differ significantly from that of the Poisson case.

(9) From [32], Theorem 1.1, for the zeroes of Gaussian entire function and calculations similar to [20], Theorem 4.3.10, for the Ginibre point process, one can check that in both cases

$$\rho^k(x_1, \dots, x_k) = \Theta\biggl(\prod_{i<j} \|x_i - x_j\|^2\biggr).$$

Hence $f^k(r) = \Theta(r^{k(k-1)})$ for these processes.

PROOF OF THEOREM 3.1. We shall prove the theorem for $k \geq 2$. The case $k = 1$ follows easily by making a few notational changes to the general case. We start with the convergence of $\mathbb{E}\{G_n(\Phi, \Gamma)\}$. In the expression for $\mathbb{E}\{G_n(\Phi, \Gamma)\}$ in (3.6), make the change of variable $x_i = x_1 + r_n y_i$ for $i \geq 2$ and then use stationarity of the point process to obtain

$$\begin{aligned}
&\mathbb{E}\{G_n(\Phi, \Gamma)\} \\
&\quad = \frac{r_n^{d(k-1)}}{k!} \int_{W_n} \int_{(r_n^{-1}(W_n - x))^{k-1}} h_{\Gamma,n}(x\mathbf{1} + r_n\mathbf{y})\rho^{(k)}(x\mathbf{1} + r_n\mathbf{y})\, dx \cdots dy_k \\
&\quad = \frac{r_n^{d(k-1)}}{k!} \int_{W_n} \int_{(r_n^{-1}(W_n - x))^{k-1}} h_{\Gamma,n}(r_n\mathbf{y})\rho^{(k)}(r_n\mathbf{y})\, dx \cdots dy_k \\
&\quad \leq \frac{r_n^{d(k-1)}}{k!} \int_{W_n} \int_{\mathbb{R}^{d(k-1)}} h_\Gamma(\mathbf{y})\rho^{(k)}(r_n\mathbf{y})\, dx \cdots dy_k \\
&\quad = \frac{nr_n^{d(k-1)}}{k!} \int_{\mathbb{R}^{d(k-1)}} h_\Gamma(\mathbf{y})\rho^{(k)}(r_n\mathbf{y})\, d\mathbf{y}.
\end{aligned}$$

Since Γ is a connected graph, $h_\Gamma \equiv 0$ outside $(B_0(k))^{k-1}$, and hence the preceding integral is finite. Further for all $x \in W_{(n^{1/d}-2k)^d}$, it follows that

$$B_O(k) \subset (W_n - x) \subset r_n^{-1}(W_n - x)$$

for large n. Hence for large enough n,

$$\mathbb{E}\{G_n(\Phi,\Gamma)\} \geq \frac{r_n^{d(k-1)}}{k!}\int_{W_{(n^{1/d}-2k)^d}}\int_{(r_n^{-1}(W_n-x))^{k-1}} h_\Gamma(\mathbf{y})\rho^{(k)}(r_n\mathbf{y})\,dx\cdots dy_k$$
$$= \frac{r_n^{d(k-1)}}{k!}\int_{W_{(n^{1/d}-2k)^d}}\int_{\mathbb{R}^{d(k-1)}} h_\Gamma(\mathbf{y})\rho^{(k)}(r_n\mathbf{y})\,dx\cdots dy_k$$
$$= \frac{((n^{1/d}-2k)^d)r_n^{d(k-1)}}{k!}\int_{\mathbb{R}^{d(k-1)}} h_\Gamma(\mathbf{y})\rho^{(k)}(r_n\mathbf{y})\,d\mathbf{y}.$$

Thus, as $n\to\infty$,

$$\frac{\mathbb{E}\{G_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)}} \sim \frac{1}{k!}\int_{\mathbb{R}^{d(k-1)}} h_\Gamma(\mathbf{y})\rho^{(k)}(r_n\mathbf{y})\,d\mathbf{y}.$$

Note that we can restrict the range of integration in the above equation to $B_0(k)$. Since $\rho^{(k)}(r_n\mathbf{y})/f^k(r_n) = g_\rho^k(\mathbf{y})$ a.e. in $B_0(k)$, and g_ρ^k is bounded (as it is continuous) in $B_0(k)$, we can use the Lebesgue dominated convergence theorem to show that, as $n\to\infty$,

$$\frac{\mathbb{E}\{G_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)}f^k(r_n)} \to \mu_0(\Phi,\Gamma).$$

This proves the convergence of expected number of Γ-subgraphs.

We shall now show that the normalized expected numbers of components and subgraphs are asymptotically equivalent for small enough radii. This will complete the proof of the theorem.

Using the lower bound of $1-\Phi(B_X(r_n))$ for the void term in J_n [see (3.4)], we obtain the following lower bound for J_n:

$$J_n(\Phi,\Gamma) \geq G_n(\Phi,\Gamma) - \frac{1}{k!}\sum_{X\in\Phi_n^{(k)}} h_{\Gamma,n}(X)\Phi\bigl(B_X(r_n)\bigr)$$
$$= G_n(\Phi,\Gamma) - E_n(\Phi,\Gamma).$$

Since $J_n \leq G_n$, we only need to show that $\frac{\mathbb{E}\{E_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)}f^k(r_n)} \to 0$. From the Campbell–Mecke formula, we have

$$\mathbb{E}\{E_n(\Phi,\Gamma)\} = \frac{1}{k!}\int_{W_n^k} h_{\Gamma,n}(\mathbf{x})\mathbb{E}_{\mathbf{x}}^!\{\Phi\bigl(B_{\mathbf{x}}(r_n)\bigr)\}\rho^{(k)}(\mathbf{x})\,d\mathbf{x}.$$

From [37], Lemma 6.4, we know that $\rho_{\mathbf{x}}^{!(1)}(y) = \frac{\rho^{(k+1)}(\mathbf{x},y)}{\rho^{(k)}(\mathbf{x})}$. Now applying the Campbell–Mecke formula for $\mathbb{E}_{\mathbf{x}}^!$ in the above equation, we find that

$$\mathbb{E}\{E_n(\Phi,\Gamma)\} = \frac{1}{k!}\int_{W_n^k\times B_{\mathbf{x}}(r_n)} h_{\Gamma,n}(\mathbf{x})\rho^{(k+1)}(\mathbf{x},y)\,d\mathbf{x}\,dy.$$

Now apply the change of variables $x_i = x_1 + r_n y_i$ for $i \geq 2$, $y = r_n y$ and proceed as in the case of $\mathbb{E}\{G_n\}$ to see that, for large enough n,

$$\mathbb{E}\{E_n(\Phi,\Gamma)\} \leq \frac{nr_n^{dk}}{k!}\int_{\mathbb{R}^{d(k-1)}\times B_0(k+1)} h_\Gamma(\mathbf{y})\rho^{(k+1)}(r_n\mathbf{y}, r_n y)\,d\mathbf{y}\,dy,$$

where the additional factor of r_n^d is due to the y variable. Dividing by $nr_n^{d(k-1)} f^k(r_n)$ and bounding h_Γ by 1, we have

$$\frac{\mathbb{E}\{E_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)} f^k(r_n)} \leq \frac{r_n^d f^{k+1}(r_n)}{k! f^k(r_n)} \int_{B_0(k)^{k-1}\times B_0(k+1)} \frac{\rho^{(k+1)}(r_n\mathbf{y}, r_n y)}{f^{k+1}(r_n)}\,dy\,d\mathbf{y}.$$

Since $f^{k+1}(r) = O(f^k(r))$ by assumption, $\frac{\mathbb{E}\{E_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)} f^k(r_n)} \to 0$ and hence

$$\frac{\mathbb{E}\{J_n(\Phi,\Gamma)\}}{nr_n^{d(k-1)} f^k(r_n)} \to \mu_0(\Phi,\Gamma),$$

as required. $\square$

The following corollary follows easily from the ordering of the joint intensities of the point processes.

COROLLARY 3.3. *Let $\Phi_i, i = 1, 2$, be two stationary point processes and $f^k_{\rho_i}, g^k_{\rho_i}$ correspond to the functions of Theorem 3.1. If $\Phi_1 \leq_{\alpha-w} \Phi_2$, then $f^k_{\rho_1} \leq f^k_{\rho_2}$. If $f^k_{\rho_1} \equiv f^k_{\rho_2}$, then $g^k_{\rho_1} \leq g^k_{\rho_2}$, and hence $\mu_0(\Phi_1,\Gamma) \leq \mu_0(\Phi_2,\Gamma)$ for a connected graph Γ that is feasible for both Φ_1 and Φ_2.*

3.3. *Thermodynamic regime*: $r_n^d \to \beta$.

THEOREM 3.4. *Let Φ be a stationary point process in $\mathbb{R}^d$ of unit intensity and Γ be a feasible connected graph of Φ on k vertices. Assume that $\rho^{(k)}$ is almost everywhere continuous, and let $r_n^d \to \beta > 0$ and $\mathbf{y} = (0, y_2, \dots, y_k)$. Then*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{G_n(\Phi,\Gamma)\}}{n} = \mu_\beta(\Phi,\Gamma) := \begin{cases} 1, & k = 1, \\ \dfrac{\beta^{k-1}}{k!}\displaystyle\int_{\mathbb{R}^{d(k-1)}} h_\Gamma(\mathbf{y})\rho^{(k)}(\beta^{1/d}\mathbf{y})\,d\mathbf{y}, & k \geq 2, \end{cases} \tag{3.9}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}\{J_n(\Phi,\Gamma)\}}{n} = \gamma_\beta(\Phi,\Gamma) \tag{3.10}$$

$$:= \begin{cases} \mathbb{P}^!_O\{\Phi(B_O(\beta^{1/d}))=0\}, & k=1, \\ \dfrac{\beta^{k-1}}{k!}\displaystyle\int_{\mathbb{R}^{d(k-1)}} h_\Gamma(\mathbf{y})\rho^{(k)}(\beta^{1/d}\mathbf{y}), & \\ \qquad\qquad \times \mathbb{P}^!_{\beta^{1/d}\mathbf{y}}\{\Phi(B_{\beta^{1/d}\mathbf{y}}(\beta^{1/d}))=0\}\,d\mathbf{y}, & k\geq 2. \end{cases}$$

If Φ is a negatively associated point process with $\mathbb{P}\{\Phi(B_{\mathbf{x}}(\beta^{1/d}))=0\}>0$ for almost every $\mathbf{x}\in B_0(\beta^{1/d}k)^k$, then $\gamma_\beta(\Phi,\Gamma)>0$.

Again, before turning to the proof, we make some observations about the theorem:

(1) The positivity of $\gamma_\beta(\Phi,\Gamma)$ is not immediate. For an example in which this does not hold, let Φ_0 be a Poisson point process of unit intensity in $\mathbb{R}^d$, $\Phi_i, i\geq 1$ i.i.d. copies of the point process of 4 i.i.d. uniformly distributed points in $B_O(\beta^{1/d}/2)$, and define the Cox point process,

$$\Phi := \bigcup_{X_i\in\Phi_0}\{X_i+\Phi_i\}.$$

Clearly, for all $X\in\Phi$, $\mathbb{P}\{\Phi(B_X(\beta^{1/d}))\geq 4\}=1$.

Now take $r_n^d\equiv\beta$ and Γ a triangle, and note that $J_n(\Phi,\Gamma)=0$ for all $n\geq 1$ and so $\gamma_\beta(\Phi,\Gamma)=0$, even though all the assumptions of Theorem 3.4 are satisfied.

(2) As in Corollary 3.3, $\Phi_1\leq_{\alpha-w}\Phi_2$ implies that $\mu_\beta(\Phi_1,\Gamma)\leq\mu_\beta(\Phi_2,\Gamma)$. However, as the previous example shows, the situation for $\gamma_\beta(\Phi,\Gamma)$ is somewhat more complicated.

(3) If $|\Gamma|=1$, then $J_n(\Phi,\Gamma)$ is the number of isolated nodes in the Boolean model of balls of radii β centered on the points of Φ. The Palm measure of a determinantal point process is also determinantal and in particular, for the Ginibre process, $\rho^{!(1)}(z)=1-e^{-\|z\|^2}$. Using this explicit structure, it can be shown that, for small enough β,

$$\gamma_\beta(\Phi_{\mathrm{Gin}},\Gamma)\geq 1-\pi\beta^2+\pi(1-e^{-\beta^2})>1-\pi\beta^2+O(\pi^2\beta^4)=\gamma_\beta(\Phi_{\mathrm{Poi}},\Gamma),$$

and hence the inequality for the γ_β could be reversed in the thermodynamic regime for even negatively associated point processes as compared to the sparse regime.

PROOF OF THEOREM 3.4. Since the proof here is similar to the preceding one, we shall not give all the details, and again, we shall only bother with the case $k\geq 2$. Starting with (3.6) and (3.7), the proof follows similar lines to that of Theorem 3.1. The difference is that $r_n^{d(k-1)}\to\beta^{k-1}$ and $\rho^{(k)}(r_n\mathbf{y})\to\rho^{(k)}(\beta^{1/d}\mathbf{y})$, and so there is no need for additional scaling. For the convergence of J_n, one first shows the convergence of $\tilde{J}_n$ using similar techniques to those in the proof of Theorem 3.1. Then note that

$$\tilde{J}_n(\Phi,\Gamma)\leq J_n(\Phi,\Gamma)\leq\tilde{J}_n(\Phi,\Gamma)+G_n(\Phi_n/\Phi_{(n^{1/d}-(k+1)r_n)^d},\Gamma).$$

The rightmost term in the upper bound accounts for the boundary effects, and by arguments similar to those in the proof of Theorem 3.1, it is easy to see that

$$\mathbb{E}\{G_n(\Phi_n/\Phi_{(n^{1/d}-(k+1)r_n)}, \Gamma)\} = O(|W_n/W_{(n^{1/d}-(k+1)r_n)^d}|) = O(n^{(d-1)/d}).$$

More importantly for us, this expectation is $o(n)$, and so of lower order than $\tilde{J}_n(\Phi, \Gamma)$. Thus $\mathbb{E}\{J_n(\Phi, \Gamma)\}/n$ also converges to $\gamma_\beta(\Phi, \Gamma)$. Since $r_n^d \to \beta > 0$, the void probability term in J_n is not necessarily degenerate.

The positivity of $\gamma_\beta(\Phi, \Gamma)$ for negatively associated point processes is an easy corollary of Lemma 2.1. We need only note that

$$F(\Phi(B)) = \mathbb{1}[\Phi(B) = 0] = (1 - \Phi(B)) \vee 0$$

is a decreasing bounded continuous function and hence

$$\mathbb{P}^!_{\mathbf{x}}(\Phi(B_{\mathbf{x}}(\beta^{1/d})) = 0) \geq \mathbb{P}\{\Phi(B_{\mathbf{x}}(\beta^{1/d})) = 0\} > 0$$

for a.e. $\mathbf{x} \in B_0(k)^k$. This completes the proof. $\square$

3.4. *Variance bounds for the sparse and thermodynamic regimes.* The crux of the second moment bounds lies in the fact that, up to constants, variances are essentially bounded above (below) by expectations for negatively associated (associated) point processes. [It is simple to check that $\mathsf{Var}(\cdot) = \Theta(\mathbb{E}\{\cdot\})$ for the Poisson process, which is both negatively associated and associated; cf. [35], Chapter 3.] We, however, shall need to extend these inequalities to graph variables, and this is the content of this section.

THEOREM 3.5 (Covariance bounds in sparse regime). *Let Γ and Γ_0 be two feasible connected graphs on k and l ($k \geq l \geq 2$) vertices, respectively, for a stationary point process Φ with almost everywhere continuous joint densities. Let Φ satisfy the assumptions of Theorem 3.1 and assume that the f^j and g^j_ρ exist for all $j \leq k + l$. Further, let $r_n \to 0$ and $\mu_0(\Phi, \Gamma) > 0$.*

(1) *If Φ is α-negatively associated, then*

$$\mathsf{Cov}(G_n(\Phi, \Gamma), G_n(\Phi, \Gamma_0)) = O(\mathbb{E}\{G_n(\Phi, \Gamma)\}).$$

(2) *If Φ is α-associated, then*

$$\mathsf{Cov}(G_n(\Phi, \Gamma), G_n(\Phi, \Gamma_0)) = \Omega(\mathbb{E}\{G_n(\Phi, \Gamma)\}).$$

PROOF. We shall prove the result for α-negatively associated processes and $k \geq 2$. The α-associated case follows by reversing the inequality sign in (3.11) below, and the case $k = 1$ needs a few simple notational changes. We shall again use the Campbell–Mecke formula to obtain closed-form expressions for the second moments and then perform a similar analysis as in the proof of Theorem 3.1 to obtain the asymptotics.

For $j \le l$ and $\mathbf{x} = (x_1, \dots, x_{k+l-j})$, in analogy to (3.1)–(3.3), define

$$h_{\Gamma,\Gamma_0,j}(\mathbf{x}) := h_\Gamma(x_1, \dots, x_k) h_{\Gamma_0}(x_1, \dots, x_j, x_{k+1}, \dots, x_{k+l-j}),$$
$$h_{\Gamma,\Gamma_0,j,n}(\mathbf{x}) := h_{\Gamma,n}(x_1, \dots, x_k) h_{\Gamma_0,n}(x_1, \dots, x_j, x_{k+1}, \dots, x_{k+l-j}).$$

Then

$$\begin{aligned}
&\mathbb{E}\{G_n(\Phi,\Gamma)G_n(\Phi,\Gamma_0)\} \\
&\quad = \mathbb{E}\Bigg\{\sum_{\mathcal{X},\mathcal{Y}\subset\Phi, |\mathcal{X}|=k, |\mathcal{Y}|=l} h_{\Gamma,n}(\mathcal{X}) h_{\Gamma_0,n}(\mathcal{Y})\Bigg\} \\
&\quad = \sum_{j=0}^{l} \mathbb{E}\Bigg\{\sum_{\mathcal{X},\mathcal{Y}\subset\Phi, |\mathcal{X}|=k, |\mathcal{Y}|=l, |\mathcal{X}\cap\mathcal{Y}|=j} h_{\Gamma,n}(\mathcal{X}) h_{\Gamma_0,n}(\mathcal{Y})\Bigg\} \\
&\quad = \sum_{j=0}^{l} \frac{1}{j!(k-j)!(l-j)!} \int_{W_n^{k+l-j}} h_{\Gamma,\Gamma_0,j,n}(\mathbf{x}) \rho^{(k+l-j)}(\mathbf{x})\, d\mathbf{x} \\
&\quad \le \sum_{j=1}^{l} \frac{1}{j!(k-j)!(l-j)!} \int_{W_n^{k+l-j}} h_{\Gamma,\Gamma_0,j,n}(\mathbf{x}) \rho^{(k+l-j)}(\mathbf{x})\, d\mathbf{x} \qquad (3.11)\\
&\qquad + \frac{1}{k!l!} \int_{W_n^k} \int_{W_n^l} h_{\Gamma,\Gamma_0,0,n}(\mathbf{x}) \rho^{(k)}(x_1, \dots, x_k) \\
&\qquad\qquad \times \rho^{(l)}(x_{k+1}, \dots, x_{k+l})\, dx_1 \cdots dx_{k+l} \\
&\quad = \sum_{j=1}^{l} \frac{1}{j!(k-j)!(l-j)!} \int_{W_n^{k+l-j}} h_{\Gamma,\Gamma_0,j,n}(\mathbf{x}) \rho^{(k+l-j)}(\mathbf{x})\, d\mathbf{x} \\
&\qquad + \mathbb{E}\{G_n(\Phi,\Gamma)\}\mathbb{E}\{G_n(\Phi,\Gamma_0)\},
\end{aligned}$$

where the inequality is due to the α-negative association property. Thus using similar arguments as in the proof of Theorem 3.1 and setting $\mathbf{y} = (0, y_2, \dots, y_{k+l-j})$, we have

$$\begin{aligned}
&\mathsf{Cov}(G_n(\Phi,\Gamma), G_n(\Phi,\Gamma_0)) \\
&\quad \le \sum_{j=1}^{l} \frac{1}{j!(k-j)!(l-j)!} \int_{W_n^{k+l-j}} h_{\Gamma,\Gamma_0,j,n}(\mathbf{x}) \rho^{(k+l-j)}(\mathbf{x})\, d\mathbf{x} \\
&\quad \sim \sum_{j=1}^{l} \frac{n r_n^{d(k+l-j-1)} f^{k+l-j}(r_n)}{j!(k-j)!(l-j)!} \int_{\mathbb{R}^{d(k+l-j-1)}} h_{\Gamma,\Gamma_0,j}(\mathbf{y}) g_\rho^{(k+l-j)}(\mathbf{y})\, d\mathbf{y} \qquad (3.12)\\
&\quad = O\big(n r_n^{d(k-1)} f^k(r_n)\big) \\
&\quad = O\big(\mathbb{E}\{G_n(\Phi,\Gamma)\}\big),
\end{aligned}$$

which is what we needed to show. □

Unlike the sparse regime, subgraph counts and component counts have different limits in the thermodynamic regime and hence we need variance bounds on component counts in the thermodynamic regime.

THEOREM 3.6 (Variance bounds in the thermodynamic regime). *Let Φ be a negatively associated stationary point process in $\mathbb{R}^d$ of unit intensity and Γ be a feasible connected graph of Φ on k vertices. Assume that $\rho^{(k)}$ is almost everywhere continuous. Let $r_n^d \to \beta > 0$ and $\gamma_\beta(\Phi, \Gamma) > 0$. Then we have that*

$$\mathsf{Var}(\tilde{J}_n(\Phi, \Gamma)) = O(\mathbb{E}\{J_n(\Phi, \Gamma)\}).$$

PROOF. First, write

$$\tilde{J}_n(\Phi, \Gamma)^2 = \tilde{J}_n(\Phi, \Gamma) + \sum_{X,Y \subset \Phi_n, |X|=|Y|=k} h_{\Gamma,n}(X) h_{\Gamma,n}(Y) \times \mathbb{1}[\Phi(B_X(r_n)) = \Phi(B_Y(r_n)) = 0].$$

By the Campbell–Mecke formula,

$$\begin{aligned}
&\mathbb{E}\{\tilde{J}_n(\Phi, \Gamma)^2\} \\
&\quad = \mathbb{E}\{\tilde{J}_n(\Phi, \Gamma)\} \\
&\qquad + \frac{1}{(k!)^2} \int_{W_n^k \times W_n^k} h_{\Gamma,n}(\mathbf{x}) h_{\Gamma,n}(\mathbf{y}) \mathbb{1}[G(\{\mathbf{x}, \mathbf{y}\}; r_n) \text{ is disconnected}] \\
&\qquad\qquad \times \mathbb{P}^!_{\mathbf{x},\mathbf{y}}\{\Phi(B_{\mathbf{x}}(r_n)) = \Phi(B_{\mathbf{y}}(r_n)) = 0\} \rho^{(2k)}(\mathbf{x}, \mathbf{y})\, d\mathbf{x}\, d\mathbf{y}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathsf{Var}(\tilde{J}_n(\Phi, \Gamma)) &= \mathbb{E}\{\tilde{J}_n(\Phi, \Gamma)\} \\
&\quad + \frac{1}{(k!)^2} \int_{W_n^k \times W_n^k} h_{\Gamma,n}(\mathbf{x}) h_{\Gamma,n}(\mathbf{y}) Q_n(\mathbf{x}, \mathbf{y})\, d\mathbf{x}\, d\mathbf{y},
\end{aligned}$$

where

$$\begin{aligned}
Q_n(\mathbf{x}, \mathbf{y}) := {} & \mathbb{1}[G(\{\mathbf{x}, \mathbf{y}\}; r_n) \text{ is disconnected}] \\
& \times \mathbb{P}^!_{\mathbf{x},\mathbf{y}}\{\Phi(B_{\mathbf{x}}(r_n)) = \Phi(B_{\mathbf{y}}(r_n)) = 0\} \\
& \times \rho^{(2k)}(\mathbf{x}, \mathbf{y}) - \mathbb{P}^!_{\mathbf{x}}\{\Phi(B_{\mathbf{x}}(r_n)) = 0\} \\
& \times \mathbb{P}^!_{\mathbf{y}}\{\Phi(B_{\mathbf{y}}(r_n)) = 0\} \rho^{(k)}(\mathbf{x}) \rho^{(k)}(\mathbf{y}).
\end{aligned}$$

Choose n large enough so that $r_n \le \beta^{1/d} + \frac{1}{4}$. For such an n and negatively associated Φ, we know from Lemma 2.2 that $Q_n(\mathbf{x}, \mathbf{y}) \le 0$ for all $\mathbf{x}, \mathbf{y}$ such that the set distance $d_S(\mathbf{x},\mathbf{y}) := \inf_{i,j} \|x_i - y_j\| > 3\beta^{1/d}$. Thus, we have that

$$\begin{aligned}\mathsf{Var}(\tilde{J}_n(\Phi,\Gamma)) \leq \mathbb{E}\{\tilde{J}_n(\Phi,\Gamma)\} &+ \frac{1}{(k!)^2}\int_{W_n^k \times W_n^k} h_{\Gamma,n}(\mathbf{x}) h_{\Gamma,n}(\mathbf{y}) \\ &\times Q_n(\mathbf{x},\mathbf{y}) \mathbb{1}[d(\mathbf{x},\mathbf{y}) \leq 3\beta^{1/d}]\, d\mathbf{x}\, d\mathbf{y}.\end{aligned}$$

From Theorem 3.4, we know that $\mathbb{E}\{\tilde{J}_n(\Phi,\Gamma)\} = \Theta(n)$, and using similar methods as in that theorem, one can show that the latter term in the above equation is of $O(n)$. Combining the two, we get that $\mathsf{Var}(\tilde{J}_n(\Phi,\Gamma)) = O(n)$. $\square$

3.5. *Phase transitions in the sparse and thermodynamic regimes.* So far, we have concentrated on the asymptotic behavior of the expectations of the numbers of different types of subgraphs that appear in the random graph associated with a point process. In this section we shall combine expectations on first and second moment to obtain results about these numbers themselves, looking at probabilities that they are nonzero, as well as L_2 and almost sure results about growth and decay rates. The main theorem of this section is the following:

THEOREM 3.7. *Let Φ be a stationary point process with almost everywhere continuous joint densities and Γ a feasible connected graph for Φ on k vertices.*

(1) *Let Φ satisfy the assumptions of Theorem 3.1 with $\mu_0(\Phi,\Gamma) > 0$. Let $r_n \to 0$.*

(a) *If $nr_n^{d(k-1)} f^k(r_n) \to 0$,*[5] *then $\mathbb{P}\{G_n(\Phi,\Gamma) \geq 1\} \to 0$.*

(b) *If Φ is α-negatively associated and $nr_n^{d(k-1)} f^k(r_n) \to \beta$ for some $0 < \beta < \infty$, there there exists a finite C (dependent on the process but not on Γ) for which*

$$\lim_{n\to\infty} \mathbb{P}\{J_n(\Phi,\Gamma) \geq 1\} \geq \left[1 + \frac{C}{\beta\mu_0(\Phi,\Gamma)}\right]^{-1}.$$

(c) *If Φ is α-negatively associated and $nr_n^{d(k-1)} f^k(r_n) \to \infty$, then*

$$\frac{J_n(\Phi,\Gamma)}{nr_n^{d(k-1)} f^k(r_n)} \xrightarrow{L_2} \mu_0(\Phi,\Gamma).$$

(2) *Let Φ be a negatively associated point process satisfying the assumptions of Theorem 3.4 with $\gamma_\beta(\Phi,\Gamma) > 0$. Let $r_n^d \to \beta$. Then*

$$\frac{J_n(\Phi,\Gamma)}{n} \xrightarrow{L_2} \gamma_\beta(\Phi,\Gamma). \tag{3.13}$$

[5]Note that neither this assumption nor the one in (1)(b) can hold for $k = 1$, as $f^1(r) \equiv 1$. Hence the statements do not say anything in these two cases.

PROOF. The proof for part (1)(a) follows from Markov's inequality and Theorem 3.1. The proof of (1)(b) is based on the following second moment bound:

$$\begin{aligned}\mathbb{P}\{J_n(\Phi,\Gamma)\geq 1\} &\geq \frac{(\mathbb{E}\{J_n(\Phi,\Gamma)\})^2}{\mathbb{E}\{J_n(\Phi,\Gamma)^2\}}\\ &\geq \frac{(\mathbb{E}\{J_n(\Phi,\Gamma)\})^2}{\mathbb{E}\{G_n(\Phi,\Gamma)^2\}}\\ &\geq \left[\frac{(\mathbb{E}\{G_n(\Phi,\Gamma)\})^2}{(\mathbb{E}\{J_n(\Phi,\Gamma)\})^2}+\frac{\mathsf{Var}(G_n(\Phi,\Gamma))}{(\mathbb{E}\{J_n(\Phi,\Gamma)\})^2}\right]^{-1}.\end{aligned}$$

Now, by applying Theorem 3.5, we obtain that there exists a $C>0$ for which

$$\mathbb{P}\{J_n(\Phi,\Gamma)\geq 1\}\geq\left[\frac{(\mathbb{E}\{G_n(\Phi,\Gamma)\})^2}{(\mathbb{E}\{J_n(\Phi,\Gamma)\})^2}+\frac{C}{\mathbb{E}\{G_n(\Phi,\Gamma)\}}\right]^{-1}.$$

Under the assumptions of (1)(b), $\mathbb{E}\{G_n(\Phi,\Gamma)\}$ converges to $\beta\mu_0(\Phi,\Gamma)$, while the first term in the square brackets converges to 1 by Theorem 3.1.

For (1)(c), observe that

$$\begin{aligned}\mathsf{Var}\big(J_n(\Phi,\Gamma)\big)&\leq \mathsf{Var}\big(G_n(\Phi,\Gamma)\big)\\ &\quad+2\mathbb{E}\{G_n(\Phi,\Gamma)\}\mathbb{E}\{E_n(\Phi,\Gamma)\}-\big(\mathbb{E}\{E_n(\Phi,\Gamma)\}\big)^2,\end{aligned}$$

where $E_n(\Phi,\Gamma)$ is as defined in the proof of Theorem 3.1. From the proofs of Theorems 3.1 and 3.5, it follows that

$$\mathsf{Var}\big(J_n(\Phi,\Gamma)\big)=O\big(\mathsf{Var}\big(G_n(\Phi,\Gamma)\big)\big)=O\big(nr_n^{d(k-1)}f^k(r_n)\big),$$

which completes the proof for this case.

We now prove part 2 in a similar fashion. In fact, it follows easily from Theorem 3.6 and the relation between J_n and $\tilde{J}_n$ noted in the proof of Theorem 3.4. More specifically, as $n\to\infty$,

$$\mathsf{Var}\left(\frac{\tilde{J}_n(\Phi,\Gamma)}{n}\right)\to 0$$

and

$$\mathbb{E}\left\{\frac{\|J_n(\Phi,\Gamma)-\tilde{J}_n(\Phi,\Gamma)\|}{n}\right\}=\frac{\mathbb{E}\{G_n(\Phi_n/\Phi_{(n^{1/d}-(k+1)r_n)^d},\Gamma)\}}{n}\to 0.$$

Thus, we have that $\frac{\tilde{J}_n(\Phi,\Gamma)}{n}\xrightarrow{P}\gamma_\beta(\Phi,\Gamma)$ and $\frac{\|J_n(\Phi,\Gamma)-\tilde{J}_n(\Phi,\Gamma)\|}{n}\xrightarrow{P}0$ as $n\to\infty$. $\square$

Since J_n is a K_d-Lipschitz functional of counting measures for a constant K_d depending only on the dimension d (see [35], Proof of Theorem 3.15), the result in (3.13) above can be strengthened to a concentration inequality for stationary determinantal point processes by using the concentration inequality in [34], Theorem 3.6. Further, (3.13) can also be extended to a strong law for ergodic point processes via the methods used in [41], Lemma 3.2.

3.6. *Extension to subcomplex counts*. The earlier section was concerned about subgraph and component counts but, as will be seen later, the techniques can be adapted to the analysis of wider classes of functionals. One specific class of functionals for which we shall explicitly state the asymptotics are subcomplex counts. These will be used in the next section. While asymptotics for Vietoris–Rips complexes can be derived using those of subgraph counts, we shall need the results of this section to derive the corresponding asymptotics for Čech complexes. We shall need a few definitions before stating these results.

Let $\mathcal{K}$ and $\mathcal{L}$ be two complexes with vertex-sets V_1 and V_2, respectively. A function $f : V_1 \to V_2$ is called a *simplicial map* if $[f(v_1), \dots, f(v_k)]$ is a face of $\mathcal{L}$ whenever $[v_1, \dots, v_k]$ is a k-face of $\mathcal{K}$. If f is a bijection and f^{-1} is also a simplicial map, f is said to be a *simplicial isomorphism*. If there exists a simplicial isomorphism between two complexes $\mathcal{K}$ and $\mathcal{L}$, then we write $\mathcal{K} \simeq \mathcal{L}$.

Let Δ be a complex on k vertices ($k \geq 1$) such that its 1-dimensional skeleton (i.e., the underlying graph) is connected (as a graph), and let $\{x_1, \dots, x_k\}$ be a collection of k points in $\mathbb{R}^d$. As in the graph case, introduce the (indicator) function $\tilde{h}_\Delta : \mathbb{R}^{dk} \times \mathbb{R}_+ \to \{0, 1\}$ defined by

$$\tilde{h}_\Delta(\mathbf{x}, r) := \mathbb{1}\big[C\big(\{x_1, \dots, x_k\}, r\big) \simeq \Delta\big], \tag{3.14}$$

where $\simeq$ denotes simplicial isomorphism, and C was defined in Definition 1.1. Let Φ be a simple stationary point process and $r_n, n \geq 1$ be a sequence of radii. As before, setting $\tilde{h}_\Delta(\mathbf{x}) := \tilde{h}_\Delta(\mathbf{x}, 1)$, we call Δ a *feasible subcomplex* of Φ if

$$\int_{(\mathbb{R}^d)^k} \tilde{h}_\Delta(\mathbf{x}) \rho^{(k)}(\mathbf{x})\, d\mathbf{x} > 0.$$

We can define an (induced) subcomplex count for the Čech complex on the point process Φ_n as follows:

$$\tilde{C}_n(\Phi, \Delta) := \frac{1}{k!} \sum_{X \in \Phi_n^{(k)}} \tilde{h}_\Delta(X, r_n).$$

Also of interest is the number of isolated Δ subcomplexes of the Čech complex on the point process Φ_n, defined as follows:

$$\tilde{C}_n^*(\Phi, \Delta) := \frac{1}{k!} \sum_{X \in \Phi_n^{(k)}} \tilde{h}_\Delta(X, r_n) \mathbb{1}\big[\Phi_n\big(B_X(r_n)\big) = k\big].$$

For the sake of brevity and to avoid repetition, we shall not provide the proofs of the following two theorems, as they are a simple extension of the proofs of Theorems 3.1, 3.4 and 3.7; see also the explanation before (5.1) in Section 5.

THEOREM 3.8. *Let Φ be a stationary point process satisfying the assumptions of Theorem 3.1 and Δ_k be a feasible connected complexes of Φ on k vertices. Let*

$r_n \to 0$. *Then*

$$\begin{aligned}\lim_{n\to\infty} \frac{\mathbb{E}\{\tilde{C}_n(\Phi, \Delta_k)\}}{n r_n^{d(k-1)} f^k(r_n)} &= \lim_{n\to\infty} \frac{\mathbb{E}\{\tilde{C}_n^*(\Phi, \Delta_k)\}}{n r_n^{d(k-1)} f^k(r_n)} \\ &= \tilde{\mu}_0(\Phi, \Delta_k) \\ &:= \begin{cases} 1, & k = 1, \\ \dfrac{1}{k!} \displaystyle\int_{\mathbb{R}^{d(k-1)}} h_{\Delta_k}(\mathbf{y}) g_\rho^k(\mathbf{y})\, d\mathbf{y}, & k \geq 2. \end{cases}\end{aligned} \tag{3.15}$$

If $\rho^{(k)}(0, \dots, 0) > 0$, *then the same result holds with* $f_\rho^k \equiv 1$ *and* $g_\rho^k \equiv \rho^{(k)}(0, \dots, 0)$.

If Φ *is* α*-negatively associated*, $\tilde{\mu}_0(\Phi, \Delta_k) > 0$ *and* $n r_n^{d(k-1)} f^k(r_n) \to \infty$, *then*

$$\frac{\tilde{C}_n^*(\Phi, \Delta_k)}{n r_n^{d(k-1)} f^k(r_n)} \xrightarrow{L_2} \tilde{\mu}_0(\Phi, \Delta_k).$$

THEOREM 3.9. *Let* Φ *be a stationary point process in* $\mathbb{R}^d$ *of unit intensity and* Δ *be a feasible connected complex of* Φ *on* k *vertices. Assume that* $\rho^{(k)}$ *is almost everywhere continuous, and let* $r_n^d \to \beta > 0$ *and* $\mathbf{y} = (0, y_2, \dots, y_k)$. *Then*

$$\begin{aligned}&\lim_{n\to\infty} \frac{\mathbb{E}\{\tilde{C}_n(\Phi, \Delta)\}}{n} \\ &\quad = \tilde{\mu}_\beta(\Phi, \Delta) \\ &\quad := \begin{cases} 1, & k = 1, \\ \dfrac{\beta^{k-1}}{k!} \displaystyle\int_{\mathbb{R}^{d(k-1)}} \tilde{h}_\Delta(\mathbf{y}) \rho^{(k)}(\beta^{1/d}\mathbf{y})\, d\mathbf{y}, & k \geq 2, \end{cases}\end{aligned} \tag{3.16}$$

$$\begin{aligned}&\lim_{n\to\infty} \frac{\mathbb{E}\{\tilde{C}_n^*(\Phi, \Delta)\}}{n} \\ &\quad = \tilde{\gamma}_\beta(\Phi, \Delta) \\ &\quad := \begin{cases} \mathbb{P}_O^!\{\Phi(B_O(\beta^{1/d})) = 0\}, & k = 1, \\ \dfrac{\beta^{k-1}}{k!} \displaystyle\int_{\mathbb{R}^{d(k-1)}} \tilde{h}_\Delta(\mathbf{y}) \rho^{(k)}(\beta^{1/d}\mathbf{y}) & \\ \qquad\qquad \times \mathbb{P}_{\beta^{1/d}\mathbf{y}}^!\{\Phi(B_{\beta^{1/d}\mathbf{y}}(\beta^{1/d})) = 0\}\, d\mathbf{y}, & k \geq 2. \end{cases}\end{aligned} \tag{3.17}$$

If Φ *is a negatively associated point process and* $\tilde{\gamma}_\beta(\Phi, \Delta) > 0$, *then*

$$\frac{\tilde{C}_n^*(\Phi, \Delta_k)}{n} \xrightarrow{L_2} \tilde{\gamma}_\beta(\Phi, \Delta_k).$$

Further, if Φ *is a negatively associated point process such that for almost every* $\mathbf{x} = (x_1, \dots, x_k) \in B_0(\beta^{1/d} k)^k$, $\mathbb{P}\{\Phi(B_\mathbf{x}(\beta^{1/d})) = 0\} > 0$, *then* $\tilde{\gamma}_\beta(\Phi, \Delta) > 0$.

4. Betti numbers of random geometric complexes. This is really the main section of the paper, giving, as it does, results about the homology of random geometric complexes through their Betti numbers. Despite this, it will turn out that, as mentioned earlier, the hard work for the proofs has already been done in the previous section.

We shall start with a review of the basic topological notions needed to formulate our results, along with an explanation of the connections between Betti numbers of random complexes, component numbers of random geometric graphs and subcomplex counts. This connection was established and exploited in [22, 23] to extract theorems for Betti numbers from those for the component counts of random geometric graphs and subcomplex counts.

4.1. *Topological preliminaries.* Recall that Čech and Vietoris–Rips complexes and their faces were already defined at Definitions 1.1 and 1.2 in the Introduction, and that the dimension of a face σ is $|\sigma| - 1$. Recall also that the edges of the random geometric graph $G(\Phi, r)$ are the 1-dimensional faces of $C(\Phi, r)$ or $R(\Phi, r)$.

Now, however, we require some additional terminology. The Vietoris–Rips complex $R(\Phi_n, r)$ is also called the *clique complex* (or *flag complex*) of $G(\Phi_n, r)$, as the faces are cliques (complete subgraphs) of the 1-dimensional faces. Let $H_k(C(\Phi_n, r))$ and $H_k(R(\Phi_n, r))$, respectively, denote the kth simplicial homology groups of the random Čech and Vietoris–Rips complexes. (We shall take our homologies over the field $\mathbb{Z}_2$, but this will not be important.) In this section we shall be concerned with asymptotics for the Betti numbers $\beta_k(C(\Phi_n, r))$ and $\beta_k(R(\Phi_n, r))$, (i.e., the ranks of the homologies) and through them the appearance and disappearance of homology groups.

Next, let P_k be the $(k+1)$-dimensional cross-polytope in $\mathbb{R}^{k+1}$, containing the origin, and defined to be the convex hull of the $2k+2$ points $\{\pm e_i\}$, where $e_1, \dots, e_{k+1}$ are the standard basis vectors of $\mathbb{R}^{k+1}$. The boundary of P_k, which we denote by $\tilde{O}_k$, is a k-dimensional simplicial complex, homotopic to a k-dimensional sphere. Let O_k be the 1-skeleton of $\tilde{O}_k$ that is, the clique complex of the graph O_k is $\tilde{O}_k$. In terms of simplicial homology of the random Vietoris–Rips complexes, the existence of subgraphs isomorphic to O_k is the key to understanding k-cycles, and so the kth homology. In fact, from [21], Lemma 5.3, we know that, because the Vietoris–Rips complex is a clique complex, any nontrivial element of the k-dimensional homology $H_k(R(\Phi_n, r))$ arises from a subcomplex on at least $2k+2$ vertices. If it has only $2k+2$ vertices, then it will be isomorphic to $\tilde{O}_k$ and the corresponding 1-skeleton will be isomorphic to O_k.

Now let Γ_k^j, $j = 1, \dots, n_k$ ($n_k < \infty$) be an ordering of the different graphs that arise when extending a $(k+1)$-clique (i.e., a k-dimensional face) to a minimal (in terms of the number of edges) connected subgraph on $2k+3$ vertices. Thus the Γ_k^j are all graphs on $2k+3$ vertices, having $\binom{k+1}{2} + k + 2$ edges.

Finally, for a given finite graph Γ, let $\tilde{G}(\Phi_n, \Gamma)$ denote the number of subgraphs of $G(\Phi_n, r_n)$ that are isomorphic to Γ. However, as explained in the discussion after (3.4), $\tilde{G}(\Phi_n, \Gamma)$ is a finite linear combination of $G_n(\Phi, \Gamma')$'s with Γ''s being of the same order as Γ.

Then [21], Lemma 5.3, and a dimension bound in [22], equation (3.1), imply the following crucial inequality linking Betti numbers to component and subgraph counts in Vietoris–Rips complexes for $k \geq 1$ and for all $n \geq 1$:

$$J_n(\Phi_n, O_k) \leq \beta_k\big(R(\Phi_n, r_n)\big) \leq J_n(\Phi, O_k) + \sum_{j=1}^{n_k} \tilde{G}_n\big(\Phi, \Gamma_k^j\big). \tag{4.1}$$

A related inequality holds for Čech complexes. Let $\tilde{\Gamma}_k$ be the complex on k vertices such that any $k-1$ vertices form a $(k-1)$-face, but $\tilde{\Gamma}_k$ is not a k-face. Any collection of vertices X for which $G(X, r) \simeq \tilde{\Gamma}_k$ is said to form an *empty* $(k-1)$-*simplex*. Let $\tilde{\Gamma}_k'$ be the complex of a $(k-1)$-face with an extra edge attached to two vertices and $\tilde{\Gamma}_k''$ be the graph of a $(k-1)$-face with a path of length 2 attached to one of the vertices. Both $\tilde{\Gamma}_k'$ and $\tilde{\Gamma}_k''$ are complexes of order $k+1$. The we have the following combinatorial inequality from [23], equation (5), for $k \in \{0, \dots, d-1\}$ and for all $n \geq 1$:

$$\begin{aligned} \tilde{C}_n^*(\Phi, \tilde{\Gamma}_{k+2}) &\leq \beta_k\big(C(\Phi_n, r_n)\big) \\ &\leq \tilde{C}_n^*(\Phi, \tilde{\Gamma}_{k+2}) + \tilde{C}_n\big(\Phi, \tilde{\Gamma}_{k+2}'\big) + \tilde{C}_n\big(\Phi, \tilde{\Gamma}_{k+2}''\big). \end{aligned} \tag{4.2}$$

With these combinatorial inequalities in hand, we are now ready to develop limit theorems for the Betti numbers of the random Čech and Vietoris–Rips complexes (Section 4.2) as well as find thresholds for vanishing and nonvanishing of homology groups (Section 4.3).

4.2. *Expectations of Betti numbers*. We return now to the setting of a stationary point process Φ in $\mathbb{R}^d$ and the sequence of finite point processes Φ_n. Our results all follow quite easily from the corresponding limit theorems in Section 3, and we continue to use the notation of that section without further comment.

The underlying heuristic is that in the sparse regime the order is determined by the order of the minimal structure involved in forming homology groups, which is O_k for the random Vietoris–Rips complex and Γ_k for the random Čech complex. Using Theorem 3.1 for the Vietoris–Rips complexes and Theorem 3.8 for the Čech complexes, it is easy to see that these are the leading order terms and that the G and $\tilde{G}$ terms in both (4.1) and (4.2) are, asymptotically, irrelevant. Hence, we have the following result.

THEOREM 4.1 (Sparse regime: $r_n \to 0$). *Let Φ be a stationary point process in $\mathbb{R}^d$ satisfying the assumptions in Theorem 3.1 for all $k \geq 1$. Let $r_n \to 0$. Further,* *assume that* $\tilde{\mu}_0(\Phi, \tilde{\Gamma}_{k+2}) > 0$ *for all* $k \in \{0, \dots, d-1\}$ *and* $\mu_0(\Phi, O_k) > 0$ *for all* $k \geq 1$. *Then*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{\beta_k(C(\Phi_n, r_n))\}}{n r_n^{d(k+1)} f^{k+2}(r_n)} = \tilde{\mu}_0(\Phi, \tilde{\Gamma}_{k+2}), \qquad k \in \{0, \dots, d-1\},$$

$$\lim_{n\to\infty} \frac{\mathbb{E}\{\beta_k(R(\Phi_n, r_n))\}}{n r_n^{d(2k+1)} f^{2k+2}(r_n)} = \mu_0(\Phi, O_k), \qquad k \geq 1.$$

For $k = 0$, *we have that*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{\beta_0(C(\Phi_n, r_n))\}}{n} = \lim_{n\to\infty} \frac{\mathbb{E}\{\beta_0(R(\Phi_n, r_n))\}}{n} = 1.$$

PROOF. We start with the case $k \geq 1$ and $k \in \{0, \dots, d-1\}$ for the Vietoris–Rips and Čech complexes, respectively. From Theorems 3.1 and 3.8, the orders of magnitude of the terms in (4.1) and (4.2) are as follows:

$$\begin{aligned}
\mathbb{E}\{\tilde{C}_n^*(\Phi, \tilde{\Gamma}_{k+2})\} &= \Theta(n r_n^{d(k+1)} f^{k+2}(r_n)),\\
\mathbb{E}\{\tilde{C}_n(\Phi, \tilde{\Gamma}'_{k+2})\} &= \Theta(n r_n^{d(k+2)} f^{k+3}(r_n)),\\
\mathbb{E}\{\tilde{C}_n(\Phi, \tilde{\Gamma}''_{k+2})\} &= \Theta(n r_n^{d(k+2)} f^{k+3}(r_n)),\\
\mathbb{E}\{J_n(\Phi, O_k)\} &= \Theta(n r_n^{d(2k+1)} f^{2k+2}(r_n)),\\
\mathbb{E}\{\tilde{G}(\Phi, \Gamma_k^j)\} &= \Theta(n r_n^{d(2k+2)} f^{2k+3}(r_n)), \qquad 1 \leq j \leq n_k.
\end{aligned}$$

Substituting these into (4.1) and (4.2), and using the fact that the limits of $\mathbb{E}\{\tilde{C}_n^*(\Phi, \tilde{\Gamma}_{k+2})\}$ and $\mathbb{E}\{J_n(\Phi, O_k)\}$ are explicitly known from Theorems 3.1 and 3.8, completes the proof of the theorem.

For the case $k = 0$, the bounds similar to (4.1) on β_0 and a similar argument will give the right asymptotics. □

Turning now to the thermodynamic regime, and applying the same arguments as in the previous proof, but using Theorems 3.4 and 3.9 in place of Theorems 3.1 and 3.8, we find that all the terms in (4.1) and (4.2) are of order $\Theta(n)$. This leads to the following result.

THEOREM 4.2 (Thermodynamic regime: $r_n^d \to \beta$). *Let* Φ *be a stationary point process in* $\mathbb{R}^d$ *satisfying the assumptions in Theorem* 3.4 *for all* $k \geq 1$. *Let* $r_n^d \to \beta \in (0, \infty)$. *Further, assume that* $\tilde{\gamma}_\beta(\Phi, \tilde{\Gamma}_k) > 0$ *for all* $k \in \{0, \dots, d-1\}$ *and* $\gamma_\beta(\Phi, O_k) > 0$ *for all* $k \geq 1$. *Then, for all* $k \geq 0$,

$$\mathbb{E}\{\beta_k(R(\Phi_n, r_n))\} = \Theta(n),$$

and for all $k \in \{0, \dots, d-1\}$,

$$\mathbb{E}\{\beta_k(C(\Phi_n, r_n))\} = \Theta(n).$$

The above asymptotics have been strengthened to convergence and strong laws in the recent preprint [41]. Further, we note without proof that one can obtain ordering results for Betti numbers of $\alpha - w$ ordered point processes in the sparse regime analogous to Corollary 3.3 but not in the thermodynamic regime.

4.3. *Thresholds for homology groups*. Our aim in this subsection is to establish results about the conditions under which different homology groups appear and disappear in the homology of random complexes. We shall need to treat Čech and Vietoris–Rips complexes separately, and start with results on the contractibility of these. We follow these with the key results of the section, on thresholds for the appearance and disappearance of homology groups. These results also show that γ-weakly sub-Poisson point processes have lower vanishing thresholds for given Γ-components. As a corollary to the results on Čech complexes, we also obtain an asymptotic result on the behavior of the Euler characteristic $\chi(C(\Phi, r))$.

Recall that there are a number of equivalent definitions for the Euler characteristic. However, the most natural for us at this point is

$$\chi\bigl(C(\Phi, r)\bigr) := \sum_{k\geq 0}(-1)^k \beta_k\bigl(C(\Phi, r)\bigr). \tag{4.3}$$

THEOREM 4.3 (Contractibility of Čech complexes). *Let Φ be a stationary γ-weakly sub-Poisson point process. Then there exists a $C_d > 0$ such that for $r_n \geq C_d(\log n)^{1/d}$, w.h.p. $C(\Phi_n, r_n)$ is contractible and $\chi(C(\Phi_n, r_n)) = 1$.*

PROOF. We start with a proof of contractibility and then show that $\chi(C(\Phi_n, r_n)) = 1$, w.h.p. As in the proof of contractibility for Poisson Čech complexes in [22], Theorem 6.1, we shall show that, for our choice of r_n, the set $\bigcup_{X\in\Phi_n} B_X(r_n/2)$ covers W_n w.h.p. Then the nerve theorem of [2], Theorem 10.7, implies that the Čech complex is contractible w.h.p. Let $\mathbb{Z}^d$ be the d-dimensional lattice, and let Q_{z_i}, $1 \leq i \leq N_n$ be an enumeration of the cubes of the scaled lattice $\frac{r_n}{4\sqrt{d}}\mathbb{Z}^d$ that are fully contained within W_n. If every cube contains a point of Φ, then $\bigcup_{X\in\Phi_n} B_X(r_n/2)$ covers W_n. By the union bound,

$$\begin{aligned}
\mathbb{P}\Bigl\{W_n \not\subseteq \bigcup_{X\in\Phi_n} B_X(r_n/2)\Bigr\} &\leq \sum_{i=1}^{N_n}\mathbb{P}\{\Phi(Q_{z_i}) = 0\} \\
&\leq N_n \mathbb{P}\Bigl\{\Phi_{(1)}\Bigl(B_O\Bigl(\frac{r_n}{8\sqrt{d}}\Bigr)\Bigr) = 0\Bigr\} \\
&\leq \frac{(4\sqrt{d})^d n}{r_n^d} e^{-(r_n/(8\sqrt{d}))^d},
\end{aligned}$$

where $\Phi_{(1)}$ is the Poisson point process of unit intensity. All that remains is to choose an appropriate $C_d > 0$ to complete the proof of contractibility for general stationary γ-weakly sub-Poisson point processes.

As for the proof of the statement about the Euler characteristic, the following obvious bound suffices:

$$\begin{aligned}\mathbb{P}\Big\{W_n \subset \bigcup_{X\in\Phi_n} B_X(r_n/2)\Big\}\\ &\le \mathbb{P}\{\beta_0(C(\Phi_n, r_n)) = 1, \beta_k(C(\Phi_n, r_n)) = 0, k \ge 1\}\\ &\le \mathbb{P}\{\chi(C(\Phi_n, r_n)) = 1\}.\end{aligned}$$ □

With these results in hand, we can now use bounds (4.1) and (4.2) along with L_2 convergence results of Theorems 3.7, 3.8 and 3.9 to complete the picture about vanishing and nonvanishing of homology groups of Čech complexes and Vietoris–Rips complexes.

THEOREM 4.4 (Thresholds for Čech complexes). *Let Φ be a stationary point process satisfying the assumptions on its joint intensities $\rho^{(k)}$ as in Theorems 3.1 and 3.4 for all $k \ge 1$. Then the following statements hold*:

(1) *Let Φ be a γ-weakly sub-Poisson point process.*

(a) *If*

$$r_n^{d(k+1)} f^{k+2}(r_n) = o(n^{-1}) \quad or \quad r_n^d = \omega(\log_n),$$

then $\beta_k(C(\Phi_n, r_n)) = 0$, $k \in \{0, \ldots, d-1\}$, w.h.p.

(b) *If $r_n^d = \omega(\log_n)$, then $\beta_0(C(\Phi_n, r_n)) = 1$, w.h.p.*

(2) *Let Φ be a negatively associated point process. Further assume that $\tilde{\mu}_0(\Phi, \Gamma_k) > 0$ and $\tilde{\gamma}_\beta(\Phi, \Gamma_k) > 0$, both for all $k \in \{0, \ldots, d-1\}$ and all $\beta > 0$.*

(a) *If*

$$r_n^{d(k+1)} f^{k+2}(r_n) = \omega(n^{-1}) \quad and \quad r_n^d = O(1),$$

then $\beta_k(C(\Phi_n, r_n)) \ne 0$, $k \in \{0, \ldots, d-1\}$, w.h.p.

(b) *If $r_n^d = O(1)$, then $\beta_0(C(\Phi_n, r_n)) \ne 0$, w.h.p.*

In the absence of a contractibility result for the Vietoris–Rips complex, we are unable to estimate the second thresholds, where the homology groups vanish. Thus we have the following less complete picture for the Vietoris–Rips complex. Since $H_0(C(\Phi_n, r_n)) = H_0(R(\Phi_n, r_n))$, we shall restrict ourselves to only $H_k(R(\Phi_n, r_n))$, $k \ge 1$, in the following theorem.

THEOREM 4.5 (Thresholds for Vietoris–Rips complexes). *Let Φ be a stationary point process satisfying the assumptions on its joint intensities $\rho^{(k)}$ as in Theorems 3.1 and 3.4 for all $k \ge 1$. Then the following statements hold for $k \ge 1$:*

(1) *If*

$$r_n^{d(2k+1)} f^{2k+2}(r_n) = o(n^{-1}),$$

then $\beta_k(R(\Phi_n, r_n)) = 0$, *w.h.p.*

(2) *Let* Φ *be a negatively associated point process. Further assume that* $\mu_0(\Phi, O_k) > 0$ *and* $\gamma_\beta(\Phi, O_k) > 0$, *both for all* $k \geq 1$ *and all* $\beta > 0$. *If*

$$r_n^{d(2k+1)} f^{2k+1}(r_n) = \omega(n^{-1}) \quad \text{and} \quad r_n^d = O(1),$$

then $\beta_k(R(\Phi_n, r_n)) \neq 0$, *w.h.p.*

4.4. *Further results for the Ginibre process.* Using the special structure of the Ginibre point process, we can improve on the threshold results of the last section. The radius regime for contractibility of Čech complexes over the Ginibre point process and zeros of GEF can be made more precise, as more is known about void probabilities in these cases. Once we have the contractibility or connectivity results, the upper bounds on the thresholds for vanishing of Betti numbers in this special case can be improved.

THEOREM 4.6 (Contractibility of Čech complexes). *Let* Φ *be the Ginibre point process or zeros of GEF. Then there exists a* $C_d > 0$ (*depending on the point process*) *such that for* $r_n \geq C_d(\log n)^{1/4}$, *w.h.p.* $C(\Phi_n, r_n)$ *is contractible. Hence,* $\beta_0(C(\Phi_n, r_n)) = 1$, $\beta_k(C(\Phi_n, r_n)) = 0$, $k \geq 1$ *and* $\chi(C(\Phi_n, r_n)) = 1$ *w.h.p. for* $r_n^2 = \omega(\sqrt{logn})$.

PROOF. The proof follows along similar lines as the proof of Theorem 4.3 except that in this case, the void probabilities are of strictly lower order and so, the radius for contractibility as well. More precisely, we know from [20], Proposition 7.2.1 and Theorem 7.2.3, that for the Ginibre point process and zeros of GEF, $-\log(\mathbb{P}\{\Phi(B_O(r)) = 0\}) = \Theta(r^4)$ as $r \to \infty$. All that remains is to substitute these bounds into the proof of Theorem 4.3 to derive the corresponding results for the Ginibre point process and zeros of GEF. □

For Vietoris–Rips complexes, we do not have a contractibility result for the Ginibre point proceses, but as a consequence of the upper bounds for the Palm void probabilities, we can obtain upper bounds on the threshold for the vanishing of the Betti numbers as well.

THEOREM 4.7 (Disappearence of homology groups for Vietoris–Rips complexes). *Let* Φ *be the Ginibre point process. Then there exists a* $C_{d,k} > 0$ *such that for* $r_n \geq C_{d,k}(\log n)^{1/4}$, *we have that w.h.p.* $\beta_k(R(\Phi_n, r_n)) = 0$, $k \geq 1$.

The proof uses the discrete Morse theoretic approach (see [16]) similar to that of [22], Theorem 5.1, and the reader is referred to that proof and the Appendix in [22] for missing details. As in [22], Theorem 5.1, our proof actually shows topological k-connectivity, though we do not state it here explicitly to avoid defining further topological notions.

PROOF OF THEOREM 4.7. As the point process is simple and stationary, index the points in Φ as $X_1, X_2, \dots$ such that $\|X_1\| < \|X_2\| < \|X_3\| < \cdots$. Define V to be the collection of pairs of simplices (V_1, V_2), $V_1 \subset V_2$ with

$$V_1 = [X_{i_1}, \dots, X_{i_k}] \quad \text{and} \quad V_2 = [X_{i_0}, X_{i_1}, \dots, X_{i_k}],$$

where $i_0 < i_1 < \cdots < i_k$. In words, we pair a simplex with another simplex of codimension 1 in the original simplex only if the additional point is closer to the origin than the rest. A simplex that is not in V is said to be a critical simplex. Let C_k be the number of critical k-simplices of V. From discrete Morse theory, we know that $\beta_k(R(\Phi_n, r_n)) \le C_k$. Thus, we only need to show that $\mathbb{E}\{C_k\} \to 0$ for all $k \ge 1$, for an appropriate choice of radii.

A k-simplex $\mathbf{X} = [X_{i_0}, \dots, X_{i_k}]$ where $i_0 < i_1 < \cdots < i_k$ is critical only if

$$\Phi_n\left(\bigcap_{j=0}^{k} B_{X_{i_j}}(r) \cap B_O(\|X_{i_0}\|)\right) = \{X_{i_0}\}.$$

Hence, using Campbell–Mecke formula for the first inequality, then using [22], Lemma 5.3—that is, for a critical k-simplex as above, there exists an $\epsilon_d > 0$ and $x \in \mathbb{R}^d$ such that

$$B_x(\epsilon_d) \subset \bigcap_{j=0}^{k} B_{X_{i_j}}(r) \cap B_O(\|X_{i_0}\|)$$

—and Lemma A.4 for the second inequality and finally $\rho^{(k)} \le 1$ for the last inequality, we find that

$$\begin{aligned}
\mathbb{E}\{C_k\} &\le \int_{W_n^{k+1}} 1[\mathbf{x} \text{ is a simplex}] 1\big[\|x_{i0}\| < \|x_{i_1}\| < \cdots < \|x_{i_k}\|\big] \\
&\qquad \times \mathbb{P}^!_{\mathbf{x}}\left\{\Phi_n\left(\bigcap_{j=0}^{k} B_{x_{i_j}}(r) \cap B_O(\|x_{i_0}\|)\right) = 0\right\} \rho^{(k)}(\mathbf{x})\, d\mathbf{x} \\
&\le \exp\left\{(k+1)(\epsilon_d r)^2 - (\epsilon_d r)^4\left(\frac{1}{4} + o(1)\right)\right\} \int_{W_n \times B_{x_{i0}}(r)^k} \rho^{(k)}(\mathbf{x})\, d\mathbf{x} \\
&\le n r^{2k} \exp\left\{k(\epsilon_d r)^2 - (\epsilon_d r)^4\left(\frac{1}{4} + o(1)\right)\right\}.
\end{aligned}$$

It is easy to see that there exists a constant $C_{d,k} > 0$ such that $\mathbb{E}\{C_k\} \to 0$ for $r_n \ge C_{d,k}(\log n)^{1/4}$, and so we are done. $\square$

5. Morse theory for random geometric complexes. Our aim in this section is to present a collection of results concerning random geometric complexes, but from the viewpoint of Morse theory.

In fact, we have already used discrete Morse theory to derive some of the connectivity thresholds for Vietoris–Rips complexes in Theorem 4.7. However, in addition to this essentially combinatorial Morse theory, there is a different and more geometric version of Morse theory for nonsmooth functions on "nice" manifolds [17]. While discrete Morse theory can be applied to study simplicial complexes without requiring any information on an ambient space in which the complex is embedded, in a geometric setting such as ours one can exploit knowledge of the ambient (Euclidean, in our case) space to apply the so-called "min-type" Morse theory.

This theory has also been exploited in the past to study of random geometric complexes on Poisson and i.i.d. point processes in [6], where it was shown that this Morse theoretic approach can give an intrinsically richer set of results than that obtained by attacking homology directly. Further, these Morse theoretic results have, as usual, implications about Betti numbers. We wish to point out that each of thse quite distinct versions of Morse theory have proved to be useful tools in the study of random complexes.

We do not intend to give full proofs here, but rather to set things up in such a way that parallels between the structures that have appeared in previous sections and those that are natural to the Morse theoretic approach become clear, and it becomes "obvious" what the Morse theoretic results will be. Full proofs would require considerable more space, but would add little in terms of insight. We note, however, that this does not make the proofs of [6] in any way redundant. On the one hand, the results there go beyond what we have here (albeit only for the Poisson and i.i.d. cases), and it is their existence that allows us to be certain that the parallels work properly.

We start with some definitions and a quick description of the Morse theoretic setting.

5.1. *Morse theory.* Morse theory for geometric complexes is based on the distance function, $d_\Phi : \mathbb{R}^d \to \mathbb{R}_+$, defined by

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

Note that while classical Morse theory deals with smooth functions, the distance function is piecewise linear, but nondifferentiable along subspaces. The extension to the distance function of classical Morse theory is discussed in detail in [6], based on the definitions and results in [17], and we shall adopt the same approach. The main difference between smooth Morse theory and that based on the distance function lies in the definition of the indices of critical points.

Critical points of index 0 of the distance function are the points where $d_\Phi = 0$, which are local and global minima, and are the points of Φ. For higher indices, define the critical points as follows: A point $c \in \mathbb{R}^d$ is said to be a *critical point* with index $1 \le k \le d$ if there exists a collection of points $\mathbf{X} = \{X_1, \dots, X_{k+1}\} \subset \Phi^{(k+1)}$ such that the following conditions hold:

(1) $d_\Phi(c) = \|c - X_i\|$ for all $1 \le i \le k+1$ and $d_\Phi(c) < \|c - Y\|$ for all $Y \in \Phi \setminus \mathbf{X}$.

(2) The points X_i, $1 \le i \le k+1$ lie in general position; namely, they do not lie in a $(k-1)$-dimensional affine space.

(3) $c \in \mathrm{conv}^o(\mathbf{X})$, where $\mathrm{conv}^o(\mathbf{X})$ denotes the interior of the convex hull formed by the points of $\mathbf{X}$.

Let $C(\mathbf{X})$ denote the center of the unique $(k-1)$-dimensional sphere (if it exists) containing the points of $\mathbf{X} \in \Phi^{(k+1)}$ and $R(\mathbf{X})$ be the radius of the ball. The conditions in the definition of critical points can be reduced to the following more workable conditions; see [6], Lemma 2.2. A set of points $\mathbf{X} \in \Phi^{(k+1)}$ in general position generates an index k critical point if and only if

$$C(\mathbf{X}) \in \mathrm{conv}^o(\mathbf{X}) \quad \text{and} \quad \Phi\big(B_{C(\mathbf{X})}\big(R(\mathbf{X})\big)\big) = 0.$$

Our interest lies in critical points which are at most at a distance r from Φ, namely, those for which $d_\Phi(c) \le r$, or, equivalently $R(\mathbf{X}) \le r$. The reason for this lies in the simple fact that

$$d_\Phi^{-1}\big([0, r]\big) = \bigcup_{x \in \Phi} B_x(r),$$

and, as we already noted earlier, by the nerve theorem this is homotopy equivalent to the Čech complex $C(\Phi, r)$.

The following indicator functions will be required to draw the analogy between counting critical points and counting components of random geometric graphs. For $\mathbf{X} \in \Phi^{(k+1)}$, define

$$h(\mathbf{X}) := \mathbb{1}\big[C(\mathbf{X}) \in \mathrm{conv}^o(\mathbf{X})\big],$$
$$h_r(\mathbf{X}) := \mathbb{1}\big[C(\mathbf{X}) \in \mathrm{conv}^o(\mathbf{X})\big]\mathbb{1}\big[R(\mathbf{X}) \le r\big].$$

Note that these functions are translation and scale invariant, as were the h_Γ functions defined for the subgraph and component counts in Section 3; namely, for all $x \in \mathbb{R}^d$ and $\mathbf{y} = (0, y_1, \dots, y_k) \in \mathbb{R}^{d(k+1)}$,

$$h_r(x, x + ry_1, \dots, x + ry_k) = h_1(\mathbf{y}).$$

This was the key property of h_Γ used to derive asymptotics for component counts. Thus, once we manage to represent the numbers of critical points as counting statistics of h_r, the analogy with component counts is made. To this end, let

$N_k(\Phi, r)$ be the number of critical points of index k for the distance function d_Φ that are at most at a distance r from Φ. Then

$$N_k(\Phi, r) = \sum_{\mathbf{X} \in \Phi^{(k+1)}} h_r(\mathbf{X}) \mathbb{1}\big[\Phi\big(B_{C(\mathbf{X})}\big(R(\mathbf{X})\big)\big) = 0\big]. \tag{5.1}$$

The similarity between the expression for $N_k(\Phi_n, r_n)$ and J_n [cf. (3.4)] should convince the reader that the method of proof used for component counts will also suffice for a derivation of the asymptotics of Morse critical points. Although the void indicator term is slightly different, we can use the fact that $R(\mathbf{X}) \leq r$ for $h_r(\mathbf{X}) = 1$ to apply the techniques of Section 3 with only minor changes.

5.2. *Limit theorems for expected numbers of critical points.* As in previous sections, we shall give results for the sparse and thermodynamic regimes separately. In the Betti number results, in the sparse regime ($r_n \to 0$) the scaling factor of n for J_n (see Theorem 3.1) arose from the translation invariance of h_Γ and Φ. The factor of $r_n^{d(k-1)}$ was due to the scale invariance of h_Γ, and the factor of $f^k(r_n)$ came from the scaling of the joint intensities $\rho^{(k)}$. Since h_r is also translation and scale invariant, we work under the same assumptions on Φ as in Theorems 3.1 and 3.5 with corresponding conditions h_r in order to obtain asymptotics for expected number of critical points of the distance function. Also, $\mathbb{E}\{N_0(\Phi_n, r)\} = \mathbb{E}\{\Phi(W_n)\} = n$ for all $r \geq 0$ and so we shall focus only on N_k, $1 \leq k \leq d$. The corresponding result is as follows:

THEOREM 5.1 (Sparse regime). *Let Φ be a stationary point process in $\mathbb{R}^d$ satisfying the assumptions of Theorem 3.1 for all $1 \leq k \leq (d+1)$. Let $r_n \to 0$ and $\mathbf{y} = (0, y_1, y_2, \dots, y_k)$. Then, for $1 \leq k \leq d$,*

$$\begin{aligned} \lim_{n\to\infty} \frac{\mathbb{E}\{N_k(\Phi_n, r_n)\}}{n r_n^{dk} f^{k+1}(r_n)} &= \nu_k(\Phi, 0) \\ &:= \frac{1}{(k+1)!} \int_{\mathbb{R}^{dk}} h_1(\mathbf{y}) g_\rho^{k+1}(\mathbf{y})\, d\mathbf{y}. \end{aligned}$$

Further, $\mathsf{Var}(N_k(\Phi_n, r_n)) = O(\mathbb{E}\{N_k(\Phi_n, r_n)\})$ *for negatively associated point processes.*

One point that is deserving of additional comment for the proof is that, as in Theorem 3.1, we can omit the void probability term in the limit by the following reasoning: since $R(\mathbf{y}) \leq r$ if $h_r(\mathbf{y}) = 1$, $\mathbf{y} = (0, y_1, \dots, y_k)$, we have that whenever $h_{r_n}(\mathbf{y}) = 1$,

$$\big\{\Phi\big(B_{C(r_n^{1/d}\mathbf{y})}\big(r_n^{1/d}\big)\big) = 0\big\} \subset \big\{\Phi\big(B_{C(r_n^{1/d}\mathbf{y})}\big(r_n^{1/d} R(\mathbf{y})\big)\big) = 0\big\},$$

and the probability of the left event here (and hence the right as well) tends to 1. This follows from similar arguments to those in the proof of Theorem 3.1.

Turning now to the thermodynamic regime, we saw in Theorem 3.4 that the sole scaling factor of n for component counts is due to the translation invariance of h_Γ and Φ. The same remains true for mean numbers of critical points.

THEOREM 5.2 (Thermodynamic regime: $r_n^d \to \beta$). *Let Φ be a stationary point process in $\mathbb{R}^d$ satisfying the assumptions of Theorem 3.4 for all $1 \le k \le (d+1)$. Let $r_n^d \to \beta \in (0,\infty)$ and $\mathbf{y} = (0, y_1, y_2, \dots, y_k)$. Then, for $1 \le k \le d$,*

$$\begin{aligned}\lim_{n\to\infty} & \frac{\mathbb{E}\{N_k(\Phi_n, r_n)\}}{n} \\ &= \nu_k(\Phi, \beta) \\ &:= \frac{\beta^k}{(k+1)!}\int_{\mathbb{R}^{dk}} h_1(\mathbf{y})\mathbb{P}^!_{\beta^{1/d}\mathbf{y}}\big(\Phi\big(B_{C(\beta^{1/d}\mathbf{y})}(\beta^{1/d}R(\mathbf{y}))\big) = 0\big)\rho^{(k)}(\beta^{1/d}\mathbf{y})\,d\mathbf{y}.\end{aligned}$$

Further, assume that Φ is also a negatively associated point process such that

$$\mathbb{P}\{\Phi(B_{C(\mathbf{x})}(\beta^{1/d})) = 0\} > 0$$

for a.e. $\mathbf{x} = (0, x_1, \dots, x_k) \in B_0(3\beta^{1/d})^{k+1}$, and for all $1 \le k \le d$. Then $\nu_k(\Phi, \beta) > 0$ for all $1 \le k \le d$.

Also, $\mathsf{Var}(N_k(\Phi_n, r_n)) = O(\mathbb{E}\{N_k(\Phi_n, r_n)\})$ for negatively associated point processes.

As previously, the void probability needs some attention. In this case, to show its positivity, we again use the fact that $R(\mathbf{y}) \le 1$ if $h_1(\mathbf{y}) = 1$, and hence, whenever $h_1(\mathbf{y}) = 1$,

$$\{\Phi(B_{C(\beta^{1/d}\mathbf{y})}(\beta^{1/d})) = 0\} \subset \{\Phi(B_{C(\beta^{1/d}\mathbf{y})}(\beta^{1/d}R(\mathbf{y}))) = 0\}.$$

The positivity of the first event under Palm probability is guaranteed by our assumption via Lemma 2.1.

Finally, we turn to a result about Euler characteristics that is not accessible from the non-Morse theory. We already defined the Euler characteristic in terms of Betti numbers at (4.3), and showed in Theorem 4.3 that, in the connectivity regime, it is 1 with high probability. However, taking an alternative, but equivalent, definition via numbers of Morse critical points, we can deduce its L_1 asymptotics in the sparse and thermodynamic regimes as a corollary of the previous results in this section. The alternative definition, which is more amenable to computations due to the bounded number of terms in the following sum, is

$$\chi(C(\Phi, r)) := \sum_{k=0}^{d} (-1)^k N_k(\Phi, r).$$

THEOREM 5.3. *Let Φ be a stationary point process in $\mathbb{R}^d$ satisfying the assumptions of Propositions 3.1 and 3.4 for all $1 \le k \le (d+1)$:*

(i) *If* $r_n \to 0$, *then*

$$n^{-1}\mathbb{E}\{\chi(C(\Phi_n, r_n))\} \to 1.$$

(ii) *If* $r_n^d \to \beta \in (0,\infty)$, *then*

$$n^{-1}\mathbb{E}\{\chi(C(\Phi_n, r_n))\} \to 1 + \sum_{k=1}^{d}(-1)^k \nu_k(\Phi, \beta).$$

(iii) *If* Φ *is also a negatively associated point process*, *then the above convergences also hold in the* L_2-*norm*.

To prove the part (iii) of the theorem, we need variance bounds, which is why we require the additional assumption of negative association. For example, in the sparse regime, we have the following bound via the Cauchy–Schwarz inequality:

$$\begin{aligned}
&\mathbb{E}\left\{\left\|\frac{\chi(C(\Phi_n, r_n))}{n} - 1\right\|^2\right\}\\
&\quad = \mathbb{E}\left\{\left\|\left(\frac{\Phi(W_n)}{n} - 1\right) + \sum_{k=1}^{d}(-1)^k \frac{N_k(\Phi, r)}{n}\right\|^2\right\}\\
&\quad \le d\left(\frac{\mathsf{Var}(\Phi(W_n))}{n^2} + \sum_{k=1}^{d}\frac{\mathbb{E}\{N_k(\Phi_n, r_n)^2\}}{n^2}\right)\\
&\quad = d\left(\frac{\mathsf{Var}(\Phi(W_n))}{n^2} + \sum_{k=1}^{d}\frac{\mathsf{Var}(N_k(\Phi_n, r_n))}{n^2} + \sum_{k=1}^{d}\frac{(\mathbb{E}\{N_k(\Phi_n, r_n)\})^2}{n^2}\right).
\end{aligned}$$

The L_2 convergence follows once it is noted that all the terms on right-hand side converge to 0 due to the variance bounds proven for negatively associated point processes. A slight modification of this argument handles the thermodynamic regime as well.

APPENDIX

In this section, we prove the result about Palm void probabilities of Ginibre point process that is used in the proof of Theorem 4.7. The proof is due to Manjunath Krishnapur.

LEMMA A.4. *Let* $D = B_0(r) \subset \mathbb{R}^2$ *for some* $r > 0$ *and* Φ *be the Ginibre point process*. *Then for* $k \ge 1$ *and* $\mathbf{x} \in \mathbb{R}^{2k}$,

$$\mathbb{P}^!_{\mathbf{x}}\{\Phi(D) = 0\} \le \exp\{kr^2\}\mathbb{P}\{\Phi(D) = 0\} = \exp\{kr^2 - r^4(\tfrac{1}{4} + o(1))\}.$$

PROOF. We shall prove the result for $k = 1$. The proof for the general case then follows by a recursive application of the same argument.

Let $\mathcal{K}_D$ be the restriction to D of the integral operator $\mathcal{K}$ corresponding to Ginibre point process. Since the Palm process of the Ginibre point process is also a determinantal point process, let $\mathcal{L}_D$ be the integral operator corresponding to the Palm point process restricted to D. Let $\lambda_i, i = 1, 2, \dots$ and $\mu_i, i = 1, 2, \dots$ be the eigenvalues of $\mathcal{K}_D$ and $\mathcal{L}_D$, respectively. From [37], Theorem 6.5, we know that $\mathcal{K}_D - \mathcal{L}_D$ has rank 1, and hence, by a generalization of Cauchy's interlacement theorem, the respective eigenvalues are interlaced with $\lambda_i \geq \mu_i \geq \lambda_{i+1}$ for $i = 1, 2, \dots$.

Now, consider the case $k = 1$ and we have the following inequality:

$$\mathbb{P}^!_{\mathbf{x}}\{\Phi(D) = 0\} = \prod_{i \geq 1}(1 - \mu_i) \leq \prod_{i \geq 2}(1 - \lambda_i) = (1 - \lambda_1)^{-1}\mathbb{P}\{\Phi(D) = 0\},$$

where the two equalities are due to [20] Theorem 4.5.3, and the inequality is due to the generalization of Cauchy's interlacement theorem described above. Now using [20], Proposition 7.2.1, to bound $\mathbb{P}\{\Phi(D) = 0\}$ and from the fact that $1 - \lambda_1 = \mathbb{P}\{\mathrm{EXP}(1) > r^2\} = \exp\{-r^2\}$ (see [20], proof of Theorem 4.7.1), where EXP(1) is the exponential random variable with mean 1, we have the desired inequality for the case $k = 1$. $\square$

Acknowledgments. The authors wish to thank Manjunath Krishnapur for pointing out an error in Remark 3.2(9) and other useful discussions regarding determinantal point processes. The authors also thank an anonymous referee for many helpful comments on the first draft. Most of this work was done when D. Yogeshwaran was a postdoctoral researcher at Technion, Israel and D. Yogeshwaran thanks Technion for its support.

REFERENCES

[1] BABSON, E., HOFFMAN, C. and KAHLE, M. (2011). The fundamental group of random 2-complexes. *J. Amer. Math. Soc.* **24** 1–28. MR2726597

[2] BJÖRNER, A. (1995). Topological methods. In *Handbook of Combinatorics*, *Vol.* 1, 2 1819–1872. Elsevier, Amsterdam. MR1373690

[3] BŁASZCZYSZYN, B. and YOGESHWARAN, D. (2015). Clustering, percolation and comparison of point processes. In *Stochastic Geometry*, *Spatial Statistics and Random Fields*: *Models and Algorithms* (V. Schmidt, ed.). *Lecture Notes in Mathematics* **2120**. Springer, Cham. To appear.

[4] BŁASZCZYSZYN, B. and YOGESHWARAN, D. (2013). Clustering and percolation of point processes. *Electron. J. Probab.* **18** no. 72, 20. MR3091718

[5] BŁASZCZYSZYN, B. and YOGESHWARAN, D. (2014). On comparison of clustering properties of point processes. *Adv. in Appl. Probab.* **46** 1–20. MR3189045

[6] BOBROWSKI, O. and ADLER, R. J. (2014). Distance functions, critical points, and topology for some random complexes. *Homology*, *Homotopy Appl.* **16** 311–344.

[7] BOBROWSKI, O. and MUKHERJEE, S. (2015). The topology of probability distributions on manifolds. *Probab. Theory Related Fields* **161** 651–686. MR3334278

[8] BOBROWSKI, O., MUKHERJEE, S. and TAYLOR, J. (2014). Topological consistency via kernel estimation. Preprint. Available at arXiv:1407.5272.

[9] BUBENIK, P., CARLSSON, G., KIM, P. T. and LUO, Z.-M. (2010). Statistical topology via Morse theory persistence and nonparametric estimation. In *Algebraic Methods in Statistics and Probability II*. *Contemp*. *Math*. **516** 75–92. Amer. Math. Soc., Providence, RI. MR2730741

[10] BURTON, R. and WAYMIRE, E. (1985). Scaling limits for associated random measures. *Ann*. *Probab.* **13** 1267–1278. MR0806223

[11] CARLSSON, G. (2009). Topology and data. *Bull*. *Amer. Math*. *Soc*. (*N.S.*) **46** 255–308. MR2476414

[12] CARLSSON, G. (2014). Topological pattern recognition for point cloud data. *Acta Numer.* **23** 289–368. MR3202240

[13] CHAZAL, F., COHEN-STEINER, D. and MÉRIGOT, Q. (2011). Geometric inference for probability measures. *Found*. *Comput*. *Math*. **11** 733–751. MR2859954

[14] COHEN, D., COSTA, A., FARBER, M. and KAPPELER, T. (2011). Topology of random 2-complexes. *Discrete Comput*. *Geom*. **47** No. 1.

[15] EDELSBRUNNER, H. and HARER, J. L. (2010). *Computational Topology*: *An Introduction*. Amer. Math. Soc., Providence, RI. MR2572029

[16] FORMAN, R. (2002). A user's guide to discrete Morse theory. *Sém*. *Lothar. Combin*. **48** Art. B48c, 35.

[17] GERSHKOVICH, V. and RUBINSTEIN, H. (1997). Morse theory for Min-type functions. *Asian J. Math*. **1** 696–715. MR1621571

[18] GHOSH, S. (2012). Determinantal processes and completeness of random exponentials: The critical case. Preprint. Available at arXiv:1211.2435.

[19] GHRIST, R. (2008). Barcodes: The persistent topology of data. *Bull*. *Amer. Math*. *Soc*. (*N.S.*) **45** 61–75. MR2358377

[20] HOUGH, J. B., KRISHNAPUR, M., PERES, Y. and VIRÁG, B. (2009). *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. *University Lecture Series* **51**. Amer. Math. Soc., Providence, RI. MR2552864

[21] KAHLE, M. (2009). Topology of random clique complexes. *Discrete Math*. **309** 1658–1671. MR2510573

[22] KAHLE, M. (2011). Random geometric complexes. *Discrete Comput*. *Geom*. **45** 553–573. MR2770552

[23] KAHLE, M. (2014). Topology of random simplicial complexes: A survey. In *Algebraic Topology*: *Applications and New Directions*. *Contemp*. *Math*. **620** (U. Tillmann, S. Galatius and D. Sinha, eds.). Amer. Math. Soc., Providence, RI. MR3290093

[24] KAHLE, M. and MECKES, E. (2013). Limit theorems for Betti numbers of random simplicial complexes. *Homology*, *Homotopy Appl*. **15** 343–374. MR3079211

[25] KALLENBERG, O. (1983). *Random Measures*, 3rd ed. Akademie-Verlag, Berlin. MR0818219

[26] LAVANCIER, F., MØLLER, J. and RUBAK, E. (2012). Determinantal point process models and statistical inference. Preprint. Available at arXiv:1205.4818.

[27] LINIAL, N. and MESHULAM, R. (2006). Homological connectivity of random 2-complexes. *Combinatorica* **26** 475–487. MR2260850

[28] MEESTER, R. and ROY, R. (1996). *Continuum Percolation*. *Cambridge Tracts in Mathematics* **119**. Cambridge Univ. Press, Cambridge. MR1409145

[29] MESHULAM, R. and WALLACH, N. (2009). Homological connectivity of random k-dimensional complexes. *Random Structures Algorithms* **34** 408–417. MR2504405

[30] MILEYKO, Y., MUKHERJEE, S. and HARER, J. (2011). Probability measures on the space of persistence diagrams. *Inverse Problems* **27** 124007, 22. MR2854323

[31] MIYOSHI, N. and SHIRAI, T. (2013). Cellular networks with α-Ginibre configurated base stations. Available at https://www.researchgate.net/publication/260094649.

[32] NAZAROV, F. and SODIN, M. (2012). Correlation functions for random complex zeroes: Strong clustering and local universality. *Comm. Math. Phys.* **310** 75–98. MR2885614

[33] PEMANTLE, R. (2000). Towards a theory of negative dependence. *J. Math. Phys.* **41** 1371–1390. MR1757964

[34] PEMANTLE, R. and PERES, Y. (2014). Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *Combin. Probab. Comput.* **23** 140–160. MR3197973

[35] PENROSE, M. (2003). *Random Geometric Graphs. Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198

[36] SCHNEIDER, R. and WEIL, W. (2008). *Stochastic and Integral Geometry*. Springer, Berlin. MR2455326

[37] SHIRAI, T. and TAKAHASHI, Y. (2003). Random point fields associated with certain Fredholm determinants. I. Fermion, Poisson and boson point processes. *J. Funct. Anal.* **205** 414–463. MR2018415

[38] STOYAN, D., KENDALL, W. and MECKE, J. (1995). *Stochastic Geometry and Its Applications*. Wiley, Chichester.

[39] TAUSZ, A. and VEJDEMO-JOHANSSON, M. (2011). JavaPlex: A research software package for persistent (co) homology. Available at http://code.google.com/p/javaplex/.

[40] TURNER, K., MILEYKO, Y., MUKHERJEE, S. and HARER, J. (2014). Fréchet means for distributions of persistence diagrams. *Discrete Comput. Geom.* **52** 44–70. MR3231030

[41] YOGESHWARAN, D., SUBAG, E. and ADLER, R. J. (2014). Random geometric complexes in the thermodynamic regime. Preprint. Available at arXiv:1403.1164.

[42] ZOMORODIAN, A. J. (2009). *Topology for Computing. Cambridge Monographs on Applied and Computational Mathematics* **16**. Cambridge Univ. Press, Cambridge. MR2549932

STATISTICS AND MATHEMATICS UNIT
INDIAN STATISTICAL INSTITUTE
BANGALORE—560059
INDIA
E-MAIL: d.yogesh@isibang.ac.in

FACULTY OF ELECTRICAL ENGINEERING
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA 32000
ISRAEL
E-MAIL: robert@ee.technion.ac.il