GEOMETRY AND TOPOLOGY OF THE BOOLEAN MODEL ON A STATIONARY POINT PROCESSES : A BRIEF SURVEY

D. YOGESHWARAN

ABSTRACT. We shall review some of the author's recent results concerning geometric and topological features of the boolean model on a stationary point processes. While study of geometric features of the Poisson boolean model span a very rich literature, the literature for topological features of the Poisson boolean model is very nascent and that for a general stationary point process is very little. In particular, the focus will be on asymptotics of geometric or topological statistics of the boolean model on a stationary point process. We shall mainly give details about topological phase transitions and central limit theorem for geometric and topological statistics.

1. INTRODUCTION

Consider a locally finite collection of random points $\mathcal{P} = \{X_i\}_{i \geq 1}$ (called as *point processes* from now onwards) on a euclidean space $\mathbb{R}^d, d \geq 2$. We shall assume that $\mathcal{P}$ is stationary (i.e., $\mathcal{P} + x$ has same distribution as $\mathcal{P}$) and $\mathcal{P}$ is simple (i.e., $X_i \neq X_j$ for $i \neq j$). For example, consider the following three examples of stationary point point processes restricted to a finite window in Figure 1. Such point processes arise in diverse applications. For example, point processes might denote node locations in a network ([Baccelli and Blaszczyszyn, 2009, Baccelli and Blaszczyszyn, 2010, Haenggi, 2012, Yukich, 2006]), configuration of sites in disordered or amorphous materials ([Hiraoka et al., 2016, Agarwala and Shenoy, 2017]), data points in data analysis ([Carlsson, 2014, Bobrowski and Kahle, 2014, Adler, 2015] [Penrose and Yukich, 2013]) and so on. A common theme in many of these applications is to build suitable geometric or topological structures to either understand such structures or use them as a tool to understand the point process itself. We shall consider a very simple object constructed from point processes called as *boolean model*. The boolean model is obtained by taking unions of balls of radius r centred at the points of $\mathcal{P}$. More formally, it is defined as $C(\mathcal{P}, r) := \cup_{X_i \in \mathcal{P}} B_r(X_i)$ where $B_r(x)$ denotes the euclidean ball of radius r centred at $x \in \mathbb{R}^d$. While it is true that the afore-cited articles sometimes consider more complicated structures on $\mathcal{P}$ or point processes on more general spaces, we shall see that

(DY) STATISTICS AND MATHEMATICS UNIT, INDIAN STATISTICAL INSTITUTE, BENGALURU, INDIA.

E-mail address: `d.yogesh@isibang.ac.in`.
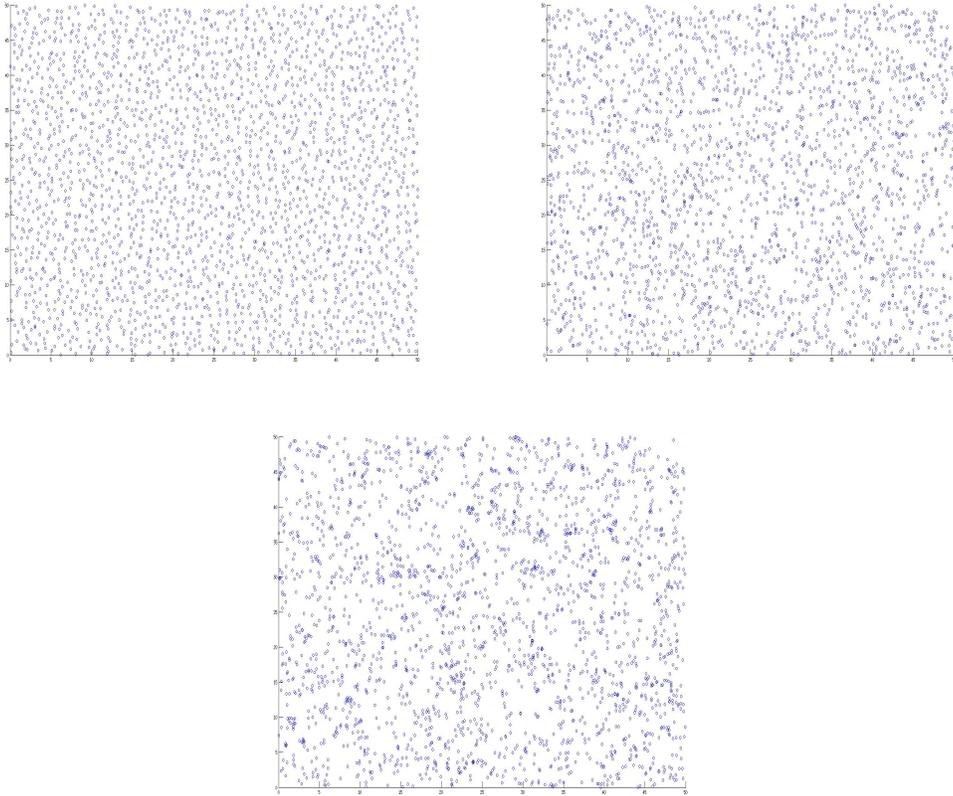
FIGURE 1. Hypergeometric perturbed lattice, Poisson point process and Negative binomial perturbed lattice

there are still many interesting unanswered questions about the boolean model on the euclidean space itself. For many questions, it is more meaningful to consider the asymptotics of $C(\mathcal{P}_n, r)$ as $n \to \infty$ where $\mathcal{P}_n = \mathcal{P} \cap W_n$ with $W_n = [-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$ being a window of volume n. For example, $V_d(C(\mathcal{P}, r)) = \infty$ (V_d denotes the volume) for many 'nice' $\mathcal{P}$ whereas asking the growth of $C(\mathcal{P}_n, r)$ still makes sense. In this short survey, we shall endeavour to describe some recent studies of such asymptotics by the author as well as some related work. The results presented here are more for illustrative purposes and hence the reader is referred to the papers [Błaszczyszyn and Yogeshwaran, 2015, Yogeshwaran and Adler, 2015, Yogeshwaran et al., 2017, Blaszczyszyn et al., 2016] for complete details and the most general results. After introducing the necessary notions and notations in sections 1.1 and 1.2, we shall present a sample of our results in Section 2 as follows :

- Thresholds for topological phase transitions on stationary point processes (Section 2.1) : As we vary r, the topology of $C(\mathcal{P}, r)$ undergoes two phase transitions - from trivial to non-trivial and then back to trivial. Though qualitatively the phase transitions remain the same for many point processes, there do exist quantitative differences in terms of how the points are spatially distributed in a point process. To summarise

2

the differences in words, the more regular the spatial distributions of the point processes are, the more narrower is the window between the two thresholds. See Figures 1 and 2 for an illustration.

- Limit theory for geometric statistics of 'asymptotically independent' point processes (Section 2.2) : Now, we turn our attention to the thermodynamic regime for point processes i.e., $r \in (0, \infty)$. For many geometric statistics (i.e.,those obtained as a sum of 'local information' around the points) and point processes satisfying a weak notion of 'asymptotic independence', we will state central limit theorems assuming variance lower bounds. We will mention examples of point processes satifying our conditions.

- Limit theory for topological statistics of Poisson point process (Section 2.3) : We shall consider topological invariants of the boolean model but over Poisson point processes. Due to the long-range dependence of topological invariants, we are more restrictive on our choice of point processes. Again, we shall state our central limit theorem and mention some related results.

Finally, we shall conclude with some omitted, on-going and future directions. We would also like to warn the reader that some of the material has already been covered in the survey [Bobrowski and Kahle, 2014] on random geometric complexes.

1.1. **Geometric and Topological Statistics :** Having described the boolean model, let us now spill some details on what we mean by geometric or topological statistics. By geometric, we refer to functionals of the boolean model that can be obtained by collecting "local information" about the boolean model around each point $X \in \mathcal{P}_n$. To be more precise, geometric statistics (denoted as G have the form)

$$(1.1) \qquad G_n := G^\xi(\mathcal{P}_n) := \sum_{X \in \mathcal{P}_n} \xi(X, \mathcal{P}_n),$$

where $\xi : \mathbb{R}^d \times \mathcal{N} \to \mathbb{R}$ ($\mathcal{N}$ denotes the space of all point processes i.e., locally-finite counting measures) is a measurable function satisfying the following conditions : (1) $\xi(x, \mathcal{P}) = 0$ if $x \in \mathcal{P}$, (2) ξ is translation invariant i.e., $\xi(x, \mathcal{P} + x) = \xi(O, \mathcal{P})$ for all $x \in \mathbb{R}^d$ with O denoting the origin and (3) there exists $r > 0$ such that $\xi(O, \mathcal{P}) = \xi(O, (\mathcal{P} \cap B_O(r)) \cup \mathcal{A})$ for any locally finite $\mathcal{A} \subset B_O(r)^c$. We call r as *the radius of stabilization* and such G_n's are called as *local geometric statistics*. We also allow r to be a random variable with a (stretched) exponentially decaying tail and we refer to such G_n's as *quasi-local* statistics. Examples of local statistics in the context of boolean model include $V_j(C(\mathcal{P}, r)), j = 0, \ldots, d$ where V_j is the jth intrinsic volume with V_d being the volume or Lebesgue measure, V_{d-1} being the surface measure and V_0 is the famed Euler-Poincaré characteristic. These are very important statistics which have applications in stereology and image analysis ([Göring et al., 2013, Klette and Rosenfeld, 2004, Kong and Rosenfeld, 1989, Svane, 2017]) and have been of interest in stochastic and integral geometry ([Schneider and Weil, 2008])

since the earliest days of the subject. Examples of quasi-local statistics include nearest-neighbour distances (i.e., $\xi_{NN}(x, \mathcal{P})$ = distance from x to its nearest-neighbour in $\mathcal{P}$) and we denote the corresponding geometric statistics as L_n. This arises in computational geometry and combinatorial optimization ([Yukich, 2006]) and is also of interest due to its relation with connectivity of the boolean model (see [Penrose, 1997]).

We use the term 'topological statistics' loosely to denote a collection of statistics that capture the topology of the boolean model or the underlying point process and what makes them interesting as well as difficult is that they cannot be expressed like in geometric statistics in (1.1). In this article, the main example of topological statistics are Betti numbers $\beta_k(C(\mathcal{P}_n, r)), k = 0, \ldots, d-1$ and persistent Betti numbers $\beta_k(C(\mathcal{P}_n, r, s)), k = 0, \ldots, d-1$. Though a precise definition is beyond the scope of the article, we try to give a heuristic description now and refer to [Carlsson, 2014, Edelsbrunner and Harer, 2010, Munkres, 1996] for details. $\beta_0(C(\mathcal{P}_n, s))$ denotes the number of connected components in $C(\mathcal{P}_n, s)$, $\beta_k(C(\mathcal{P}_n, r))$ denote the number of $(k+1)$-dimensional 'holes' or equivalently the number of independent k-dimensional non-trivial cycles. The persistent Betti numbers are rather meant to count the 'holes' or non-trivial cycles that persist from r to s.

More generally, as we track the evolution of $C(\mathcal{P}_n, r)$ as r varies from 0 to ∞, we can associate a birth-time and death-time to every non-trivial cycle or hole. Since $C(\mathcal{P}_n, r)$ are increasing in r, every non-trivial cycle or hole appears for the first time for some r called *the birth-time of the cycle* and vanishes (i.e., the cycle becomes trivial or the hole is filled) for some s ($\geq r$) called *the death-time of the cycle*. The collection of all birth and death times corresponding to the kth Betti number is the kth persistent diagrams : $Per_k := \{(b_i, d_i)\}$. Given Per_k, we can infer both the Betti and persistent Betti numbers as follows :

$$\beta_k(C(\mathcal{P}_n, r, s)) = \sum_i \mathbf{1}[b_i \leq r \leq s \leq d_i], \ \ \beta_k(C(\mathcal{P}_n, r)) = \beta_k(C(\mathcal{P}_n, r, r)).$$

Further, instead of representing Per_k as points in $\mathbb{R}^2$, we can also represent Per_k as barcodes by representing each (b_i, d_i) as a bar from b_i to d_i. We give the barcodes of Per_1 for a related model (the Vietoris-Rips complex) in Figure 2 corresponding to the three point processes in Figure 1. For an illustration of Per_1 of the boolean model of Poisson point process and two other point processes (Ginibre and Poisson cluster), we refer the reader to [Duy et al., 2016, Figure 1]. One notable feature of this diagrams is that the support of the barcode is narrower if the spatial distribution of points is more regular i.e., the support of the barcodes is shorter for hypergeometric perturbed lattice compared to that of the Poisson point process which is shorter compared to that of the negative Binomial perturbed lattice.

Persistent diagrams are the key tool in the new and thriving area of topological data analysis or applied topology ([Carlsson, 2014, Bobrowski and Kahle, 2014, Adler, 2015, Ghrist, 2014]). In this subject, point process represent data points and persistent diagrams capture the topology of the points as well as the underlying space on which the points are distributed. In many applications, it is reasonable to assume that one is unaware of the space on which the data points are distributed and one is interested in gleaning some information
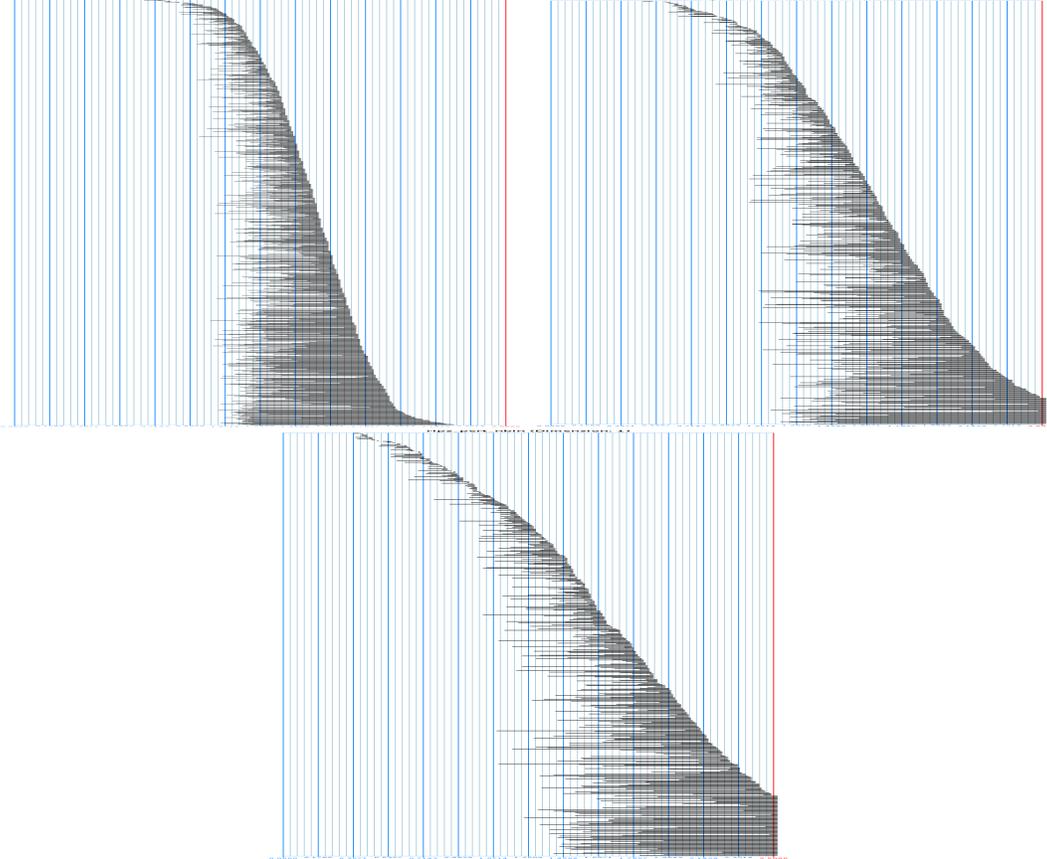
FIGURE 2. Barcode representation of Per_1 of Hypergeometric perturbed lattice, Poisson point process and Negative binomial perturbed lattice

about the space via the points. Of course, our space is euclidean, the simplest possible underlying space and has trivial topological features. However, as we shall see, a complete mathematical understanding of point process on even this simple space eludes us. There is some progress about understanding the topology on more general spaces in [Bobrowski and Mukherjee, 2015, Bobrowski and Weinberger, 2017] but describing these results is beyond the scope of this article.

1.2. **Point processes :** We shall now introduce some notions and examples related to stationary point processes and refer the reader to [Ben-Hough et al., 2009, Blaszczyszyn et al., 2016, Yogeshwaran and Adler, 2015, Błaszczyszyn and Yogeshwaran, 2015, Schneider and Weil, 2008, Last and Penrose, 2017] for details.

By stationarity we have that $\mathbb{E}(\mathcal{P}(B)) = \lambda V_d(B)$ for any Borel set B. As a standing assumption, we set $\lambda \in (0, \infty)$. Note that a point process can be viewed as a random set $\{X_i\}$ or as a random counting measure $\mathcal{P} = \sum_i \delta_{X_i}(.)$. We set $\alpha^{(k)}(.)$ to be the k th order factorial moment measure on $(\mathbb{R}^d)^k$ and assume that it has a Radon-Nikodyn density $\rho^{(k)}(.)$ with respect to the Lebesgue measure, also called k-point correlation functions. If $\alpha^{(k)}$ is locally finite, there exists a family of probability distributions called the Palm measures $P_{x_1,...,x_k}$

on $(\mathcal{N}, \mathcal{B})$, defined and unique up to an $\alpha^{(k)}$-null set of $(\mathbb{R}^d)^k$ and satisfy the disintegration formula (also known as the refined Campbell theorem)

$$(1.2) \quad \mathbb{E}\Big(\sum_{(x_1,\ldots,x_k)\in\mathcal{P}^{(k)}} f(x_1,\ldots,x_k;\mathcal{P})\Big) = \int_{\mathbb{R}^{kd}} \int_{\mathcal{N}} f(x_1,\ldots,x_k;\mu)\, P_{x_1,\ldots,x_k}(\mathrm{d}\mu)\alpha^{(k)}(\mathrm{d}x_1,\ldots\mathrm{d}x_k)$$

for any (say non-negative) measurable functions f on $(\mathbb{R}^d)^k \times \mathcal{N}$. We denote the corresponding expectation $\mathbb{E}_{x_1,\ldots,x_k}$. Intuitively, $P_{x_1,\ldots,x_k}$ is the distribution of $\mathcal{P}$ conditioned on $\{x_1,\ldots,x_k\} \subset \mathcal{P}$. The factorial moment measures or correlation functions characterize point processes under some reasonable assumptions (see [Ben-Hough et al., 2009, Remark 1.2.4]) and so do the void probabilities defined as $\nu(B) = \mathbb{P}(\mathcal{P}(B) = \emptyset)$ for Borel subsets B. Here are a few examples of stationary point processes.

(1) Poisson point process - $\mathcal{P}_\lambda$: The most ubiquituous model and characterized either by $\rho^{(k)}(.) \equiv \lambda^k$ or $\nu(B) = e^{-\lambda V_d(B)}$.

(2) Determinantal point process - $\mathcal{P}_{det}$: A point process $\mathcal{P}$ is said to a determinantal point process with kernel K if $\rho^{(k)}(x_1,\ldots,x_k) = det\big(K(x_i,x_j)\big)_{1\leq i,j\leq k}$ for all $x_1,\ldots,x_k \in \mathbb{R}^d$. For a kernel K to generate a correlation function as mentioned, K must be a locally trace-class Hermitian kernel. The most popular euclidean determinantal point process is the Ginibre point process $\mathcal{P}_{Gin}$ on $\mathbb{C}$ (identified with $\mathbb{R}^2$) and defined by the kernel $K(z,w) = \exp-\frac{|z|^2+|w|^2-2z\bar{w}}{2}, z, w \in \mathbb{C}$.

(3) Zeros of Gaussian entire function - $\mathcal{P}_{GEF}$: This point process is $\mathcal{P}_{GEF} = F^{-1}(0)$ where F is a Gaussian analytic function defined by $F(z) := \sum_{j\geq 0} X_j \frac{z^j}{\sqrt{j!}}$ with X_j's being i.i.d. standard complex normal random variables.

For more details on the last two examples, we refer to [Ben-Hough et al., 2009]. We refer the reader to [Blaszczyszyn et al., 2016, Section 2.2] and [Yogeshwaran and Adler, 2015, Section 2.2] for more examples of point process that fall within our framework.

2. Preview of results

2.1. **Some topological phase transitions.** We hinted at the regulairty of spatial distribution of points playing a crucial role in determining their topological phase transitions. A mathematical way of characterizing the same is as follows : We say that $\mathcal{P}$ is *weakly sub-Poisson* if for $\lambda = \alpha([0,1]^d)$ we have that

$$(2.1) \qquad\qquad \rho^{(k)}(x_1,\ldots,x_k) \leq \lambda^k \ ; \ \mathbb{P}(\mathcal{P}(B) = \emptyset) \leq e^{-\lambda V_d(B)}$$

i.e., the correlation functions and void probabilities are lesser than that of the corresponding Poisson point process. $\mathcal{P}$ is said to be *weakly super-Poisson* if the above inequalities are reversed. These definitions and some other notions of comparison of point processes were introduced in the author's thesis and we refer to the survey [Błaszczyszyn and Yogeshwaran, 2015] for details. Here weakly sub-Poisson means the points are distributed more regularly than a Poisson point process and weakly super-Poisson indicates that the points are more clustered than a Poisson point process. For example, the hypergeometric perturbed lattice and determinantal point processes are weakly sub-Poisson while the negative Binomial perturbed

lattice and the permanental point process are weakly super-Poisson. Our first theorem establishes bounds for the two thresholds for topological phase transitions for weakly sub-Poisson point processes.

Theorem 2.1. *([Yogeshwaran and Adler, 2015, Theorem 4.4]) Let $\mathcal{P}$ be a stationary weakly sub-Poisson point process. Then if r_n is a sequence of radii such that either $r_n^{d(k+1)} = o(n^{-1})$ or $r_n^d = \omega(\log n)$ [1] then as $n \to \infty$,*

$$\mathbb{P}(\beta_k(C(\mathcal{P}_n, r_n)) = 0) \to 1, k = 1, \dots, d - 1,$$

and further if $r_n^d = \omega(\log n)$, we have that as $n \to \infty$,

$$\mathbb{P}(\beta_0(C(\mathcal{P}_n, r_n)) = 1) \to 1.$$

Such a result was proven for Poisson point processes alone in [Kahle, 2011] and along with the results of [Bobrowski and Weinberger, 2017], one knows that the thresholds are sharp for the Poisson point process. Hence, our result implies that for weakly sub-Poisson point processes, "non-trivial topology appears later than that of Poisson but vanishes earlier than Poisson". It is feasible at times to quantify 'how later/earlier than Poisson'. Such quantifications either involve the stronger property of negative association or are very much dependent on the behaviour of the k-correlation functions near the diagonal and the void probabilities for large balls. Thus, a complete statement is beyond the scope of this paper but as a sample, we shall mention the result in the special case of the Ginibre point process.

Proposition 2.2. *([Yogeshwaran and Adler, 2015, Theorems 4.4 and 4.6]) Let $\mathcal{P}$ be the Ginibre point process $\mathcal{P}_{Gin}$ on $\mathbb{R}^2$. If r_n is a sequence of radii such that either $r_n^{(k+1)(k+4)} = o(n^{-1})$ or $r_n^d = \omega((\log n)^{1/4})$, we have that as $n \to \infty$,*

$$\mathbb{P}(\beta_k(C(\mathcal{P}_n, r_n)) = 0) \to 1, k = 1, \dots, d - 1$$

and if $r_n^{(k+1)(k+4)} = \omega(n^{-1})$ and $r_n = O(1)$, then as $n \to \infty$,

$$\mathbb{P}(\beta_k(C(\mathcal{P}_n, r_n)) \neq 0) \to 1, k = 1, \dots, d - 1.$$

Further if $r_n^d = \omega((\log n)^{1/4})$, then $\mathbb{P}(\beta_0(C(\mathcal{P}_n, r_n)) = 1) \to 1$ as $n \to \infty$.

As a comparison, for the Poisson point process on $\mathbb{R}^2$, the above proposition holds with $(k + 4)(k + 1)$ replaced by $2(k + 1)$ and $1/4$ by $1/2$. Thus, this substantiates our heuristic argument that "more regularly the points are distributed spatially, the support of the persistence barcodes are narrower". The key to the proofs of the above result is first using the refined Campbell theorem ((1.2)) to compute first and second moments for suitable geometric statistics and using them as an approximation for topological statistics.

---

[1]We are using the Bachmann-Landau big O-small o notation for asymptotics.

2.2. Geometry of the boolean model : Thermodynamic regime.

Having broadly etched the thresholds for topological phase transitions, we wish to focus on distributional results within the different radii regimes. In this survey, we shall focus on thermodynamic regime (i.e., $r \in (0, \infty)$) as it has a rich set of results. We call this the thermodynamic regime as the asymptotic density of balls intersecting a window W_n (i.e., $n^{-1}\sharp\{i : B_r(X_i) \cap W_n \neq \emptyset\}$) converges to a constant. Recall the definition of geometric statistics (1.1) and score functions defined therein. Let us consider the following simple score function $\xi_k, k \geq 0$ of the boolean model : Setting $X_0 = x$, we define

$$\xi_k(x, \mathcal{P}) := \frac{1}{k!} \sum_{\{X_1,\ldots,X_k\} \subset \mathcal{P}}^{\neq} \mathbf{1}[\cap_{i=0}^{k} B_r(X_i) \neq \emptyset].$$

Denote the corresponding geometric statistic as $G_{n,k} := G^{\xi_k}(\mathcal{P}_n)$. Essentially $G_{n,k}$ counts the number of k-wise intersections in the boolean model $C(\mathcal{P}_n, r)$. In combinatorial topology, $G_{n,k}$ represents the number of k-simplices in the Čech complex on $\mathcal{P}_n$ formed by balls of radius r (see [Edelsbrunner and Harer, 2010] for a precise definition). The advantage of teh Čech complex is its combinatorial nature and at the same time, it is homotopy equivalent to the boolean model because of the nerve theorem ([Björner, 1995, Theorem 10.7]). Simplices are the building block of any complex and hence one of the first quantities of interest in a random complex. Thus, we can add $G_{n,k}$ also to the list of geometric statistics we have already encountered. We shall now illustrate our results for such geometric statistics by stating a few special cases here. Firstly, for a score function ξ, define

$$(2.2) \qquad \sigma^2(\xi) := \mathbb{E}_0 \xi^2(O, \mathcal{P}) \rho^{(1)}(O) + \int_{\mathbb{R}^d} (m_{(2)}(O, x) - m_{(1)}(O)^2) \, \mathrm{d}x.$$

where we define $m_{(p)}(x_1,\ldots,x_p) := \mathbb{E}_{x_1,\ldots,x_p}(\xi(x_1, \mathcal{P})\ldots\xi(x_p, \mathcal{P}))\rho^{(p)}(x_1,\ldots,x_p)$ for $p \geq 1, x_1,\ldots,x_p \in \mathbb{R}^d$ with $\mathbb{E}_{x_1,\ldots,x_p}$ and $\rho^{(p)}(x_1,\ldots,x_p)$ being the Palm expectation and correlation functions respectively as defined in Section 1.2.

Theorem 2.3. *Let $k \in \mathbb{N}$ and $\mathcal{P} = \mathcal{P}_{GEF}$ or $\mathcal{P}$ is a determinantal point process with kernel $K(x, y)$ such that $|K(x, y)| \leq \phi(|x - y|)$ where ϕ is exponentially decaying (i.e., $\liminf_{t\to\infty} t^b \log \phi(t) < 0$ for some $b > 0$). Then, we have that*

$$\lim_{n\to\infty} n^{-1}\mathbb{V}arG_{n,k} = \sigma^2(\xi_k).$$

Additionally, if $\mathbb{V}arG_{n,k} = \Omega(n^\nu)$ for some $\nu \in (0, \infty)$, then as $n \to \infty$

$$\frac{G_{n,k} - \mathbb{E}G_{n,k}}{\sqrt{\mathbb{V}arG_{n,k}}} \xrightarrow{\mathcal{D}} N,$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution and N denotes standard normal random variable.

Theorem 2.4. *Let $\mathcal{P}$ be a stationary determinantal point process on $\mathbb{R}^d$ as in Theorem 2.3. We have that for all $j = 0,\ldots,d$,*

$$\lim_{n\to\infty} n^{-1}\mathbb{V}arV_j(C(\mathcal{P}_n, r)) = \sigma^2(\xi_j) \; ; \; \lim_{n\to\infty} n^{-1}\mathbb{V}arL_n = \sigma^2(\xi_{NN}),$$

where $\xi_j, j = 0, \ldots, d$ are the score functions generating the corresponding intrinsic volumes V_j's as in (1.1). Moreover, if $\mathbb{V}arV_j(C(\mathcal{P}_n, r)) = \Omega(n^\nu)$ for some $\nu \in (0, \infty)$ then as $n \to \infty$

$$\frac{V_j(C(\mathcal{P}_n, r)) - \mathbb{E}V_j(C(\mathcal{P}_n, r))}{\sqrt{\mathbb{V}arV_j(C(\mathcal{P}_n, r))}} \xrightarrow{\mathcal{D}} N.$$

Similarly, if $\mathbb{V}arL_n) = \Omega(n^\nu)$ for some $\nu \in (0, \infty)$ then as $n \to \infty$

$$\frac{L_n - \mathbb{E}L_n}{\sqrt{\mathbb{V}arL_n}} \xrightarrow{\mathcal{D}} N.$$

Remark 2.5 (Remarks on Theorems 2.3 and 2.4)**.**

(1) *Firstly, we again re-emphasize that the above theorems are illustrative and results as above hold true for more general point processes satisfying the following condition : For all $x_1, \ldots, x_{p+q}$ with $s := \min_{1 \le i \le p, 1 \le j \le q} |x_i - x_{p+j}|$, we have that*

(2.3) $\qquad |\rho^{(p+q)}(x_1, \ldots, x_{p+q}) - \rho^{(p)}(x_1, \ldots, x_p)\rho^{(q)}(x_{p+1}, \ldots, x_{p+q})| \le C_{p+q}\phi(c_{p+q}s),$

for a fast decreasing ϕ and for all $k \ge 1$, $C_k < \infty, c_k > 0$. Theorem 2.3 holds for this class of point processes provided they satisfy some moment conditions. Apart from the two examples mentioned above, many α-permanental point processes also satisfy (2.3).

(2) *For applying Theorem 2.4 to general point processes, apart from satisfying (2.3) and some moment conditions, it is necessary for $\inf_k c_k > 0$ and $C_k = O(k^{ak})$ for some $a < 1$. Such examples as expected are rarer. Apart from determinantal point processes mentioned in the aboe theorems, certain Gibbs point processes, finite-range dependent point processes and some Cox point processes are the well-known examples. We refer the reader to [Blaszczyszyn et al., 2016, Section 1.4] for the general and precise statements and [Blaszczyszyn et al., 2016, Section 2] for more examples and applications.*

(3) *We refer to [Blaszczyszyn et al., 2016, Section 1] for weak laws and for more detailed variance behaviour. For example, it is shown that if variance is not of volume order, it is at most of surface order. However, it is a challenge to prove suitable variance bounds except in specific cases.*

(4) *We refer to remarks in [Blaszczyszyn et al., 2016, Section 1.4] for more details on the previous literature and comparisons. However, we mention that similar results for Poisson point process and for certain Gibbs point processes was proven in [Baryshnikov and Yukich, 2005] and [Schreiber and Yukich, 2008] respectively. Refer to [Yukich, 2013] for a detailed survey of these results.*

(5) *In full generality, the above theorems are stated for linear functionals of the random measure $\mu_n^\xi := \sum_{X \in \mathcal{P}_n} \xi(x, \mathcal{P}_n)\delta_{n^{-1/d}x}(.)$.*

(6) *In a recent pre-print [Ram Reddy et al., 2017], these results have been extended to random fields on finitely generated Cayley graphs.*

(7) *The key tool to the proofs is to derive a series expansion for the mixed moments $m_{(p)}(.)$ using the factorial moment expansion for functionals of point processes (see*

[Błaszczyszyn, 1995, Błaszczyszyn et al., 1997]). *Using the above mentioned expansion, we show clustering as in (2.3) for mixed moments and from there, prove the central limit theorem via the classical cumulant method.*

2.3. **Topology of the Poisson boolean model : Thermodynamic regime.** In this section, we shall state the strong law for certain topological statistics of the boolean model - the Betti numbers and the persistent Betti number- on ergodic point processes and the central limit theorem under the further assumption that $\mathcal{P}$ is a stationary Poisson point process . For a boolean model, the non-trivial Betti numbers are $\beta_i, 0 \leq i \leq d-1$.

Theorem 2.6. *Let $\mathcal{P}$ be an ergodic point process such that $\mathbb{E}(\mathcal{P}(B)^m) < \infty$ for all $m \in \mathbb{N}$ and for all bounded Borel subsets B. Let $0 < r \leq s < \infty$. Then, for $0 \leq k \leq d-1$, there exist constants $\beta_k^\infty(r,s) \in [0,\infty)$ such that*

$$n^{-1}\beta_k(C(\mathcal{P}_n, r, s)) \overset{a.s}{\to} \beta_k^\infty(r,s).$$

Theorem 2.7. *Let $\mathcal{P}$ be the stationary Poisson process and and $0 < r \leq s < \infty$. Then, for $k \in \{1, \ldots, d-1\}$, there exists a constant $\sigma_k^2(r,s) \in [0,\infty)$ such that, as $n \to \infty$,*

$$n^{-1}\mathbb{V}ar\beta_k(C(\mathcal{P}_n, r, s) \to \sigma_k^2(r,s),$$

and

$$n^{-1/2}\left(\beta_k(C(\mathcal{P}_n, r, s)) - \mathbb{E}\beta_k(C(P_n, r, s))\right) \Rightarrow N(0, \sigma_k^2(r,s)).$$

Remark 2.8 (Remarks on Theorems 2.6 and 2.7). (1) *For the case $k = 0$, the above results are well-known from [Penrose, 2003]. The above results for the case $r = s$ was proven in [Yogeshwaran et al., 2017, Theorems 3.5 and 4.7] and later extended to the case $r \leq s$ by [Duy et al., 2016, Theorems 1.11 and 1.12].*
(2) *It is not necessary that $\beta_k^\infty(r,s) > 0$ but it holds for many well-known point processes (see [Yogeshwaran and Adler, 2015, Theorem 3.3] and [Duy et al., 2016, Theorem 1.9]. In fact, the strong law has been shown to hold for inhomogeneous Poisson point processes as well in [Duy, 2017].*
(3) *As for variance asymptotics, while it is shown in [Yogeshwaran et al., 2017, Theorem 4.7] that $\sigma_k^2(r,r) > 0$ no such assertion can be made about $\sigma_k^2(r,s)$ for $r < s$. Further, variance lower bounds for inhomogeneous Poisson point processes for β_0, β_{d-1} have been shown in [Yogeshwaran et al., 2017, Lemma 4.3] and we wish to mention that using the construction from [Bobrowski et al., 2016, Section 7], it might be possible to show lower bounds for other Betti numbers.*
(4) *The proof of the strong law rests on exploiting near-additivity of (persistent) Betti numbers along with ergodicity of the point process. The central limit theorem proceeds via a general martingale-difference based central limit theorem for Poisson functional derived in [Penrose and Yukich, 2001, Theorem 2.1]. The key step in the proof is to verify stabilization of the first-order difference operator or the add-one cost of $\beta_k(C(\mathcal{P}_n, r, s))$. For the case $r = s$, this is done via the Mayer-Vietoris exact sequence but for the case $r \leq s$, this is done via a more direct analysis of the add-one cost.*

3. Omitted, on-going and future directions :

As is to be expected in such a short survey, we have omitted numerous related results of which we shall refer to the latest here. The three notable omissions are the work on discrete random complexes (see the survey [Kahle, 2014]), thresholds for vanishing of homology in random geometric complexes (see [Bobrowski and Weinberger, 2017, Bobrowski and Oliveira, 2017] for the most recent results) and effect of the tail of the density of the inhomogeneous Poisson point process on the topology (see [Owada, 2016]). We refer the reader to the surveys [Kahle, 2014, Bobrowski and Kahle, 2014] for more on both discrete and geometric random complexes.

In an on-going project, we are extending the limit theorems for Betti numbers and persistent Betti numbers to lifetime sum of barcodes i.e., $\int_0^\infty \beta_k(C(\mathcal{P}_n, r))$. In the case of $k = 0$, this corresponds to the length of the minimal spanning tree on the complete graph on $\mathcal{P}_n$ with edge-weights being the euclidean distances. In a seminal work, [Kesten and Lee, 1996] proved the CLT for length of minimal spanning tree using martingale-difference argument and this served as the motivation for the general CLT of [Penrose and Yukich, 2001, Theorem 2.1] that we referenced earlier. Given this connection, it is natural to wonder whether for higher k, there is a correspondance to higher-dimensional versions of trees. Such a correspondance is estabilished to 'minimal spanning acycles' in [Hiraoka and Shirai, 2015] and [Skraba et al., 2017] via different methods. Hence, this on-going work shall provide limit theorems for statistics of euclidean minimal spanning acycles.

For a reader interested in exploring more about this subject, we mention a few open problems. A basic question would be to derive strong laws and variance lower bounds for persistent Betti numbers of the boolean model on an inhomogeneous Poisson point process. A more non-trivial question would be to estabilish CLT as well as rates of convergence for topological statistics of the boolean model on an inhomogeneous Poisson point process. A starting point for such studies might be the recent work on CLT rates for minimal spanning tree in [Chatterjee and Sen, 2017] and a more general result for CLT rates of Poisson functionals via the Malliavin-Stein method in [Last et al., 2016, Peccati and Reitzner, 2016]. Now, if we move onto stationary point processes, there are no CLTs for topological statistics and rates of CLT are unavailable even for geometric statistics. Possible approaches to such a question of rates might be via deriving more careful bounds on cumulants and then using [Grote and Thäle, 2016, Lemma 4.2] or [Heinrich and Schmidt, 1985, Theorem 1]. Such methods might also be useful in proving other limit theorems such as moderate deviations.

References

[Adler, 2015] Adler, R. J. (2014-2015). *TOPOS*. http://bulletin.imstat.org/category/robert-adler/.

[Agarwala and Shenoy, 2017] Agarwala, A. and Shenoy, V. B. (2017). Topological insulators in amorphous systems. *Phys. Rev. Lett.*, 118:236402.

[Baccelli and Blaszczyszyn, 2009] Baccelli, F. and Blaszczyszyn, B. (2009). *Stochastic Geometry and Wireless Networks: Volume I Theory*, volume 3. Now Publishers, Inc.

[Baccelli and Blaszczyszyn, 2010] Baccelli, F. and Blaszczyszyn, B. (2010). *Stochastic geometry and wireless networks: Volume II Applications*, volume 4. Now Publishers, Inc.

[Baryshnikov and Yukich, 2005] Baryshnikov, Y. and Yukich, J. E. (2005). Gaussian limits for random measures in geometric probability. *Ann. Appl. Prob.*, 15(1):213–253.

[Ben-Hough et al., 2009] Ben-Hough, J., Krishnapur, M., Peres, Y., and Virág, B. (2009). *Zeros of Gaussian analytic functions and determinantal point processes*, volume 51. American Mathematical Society Providence, RI.

[Björner, 1995] Björner, A. (1995). *Topological methods. In Handbook of Combinatorics*. Elsevier, Amsterdam.

[Błaszczyszyn, 1995] Błaszczyszyn, B. (1995). Factorial moment expansion for stochastic systems. *Stoch. Proc. Appl.*, 56(2):321–335.

[Błaszczyszyn et al., 1997] Błaszczyszyn, B., Merzbach, E., and Schmidt, V. (1997). A note on expansion for functionals of spatial marked point processes. *Stat. Probab. Letters*, 36(3):299–306.

[Błaszczyszyn and Yogeshwaran, 2015] Błaszczyszyn, B. and Yogeshwaran, D. (2015). *Clustering Comparison of Point Processes, with Applications to Random Geometric Models*, pages 31–71. Springer International Publishing, Cham.

[Blaszczyszyn et al., 2016] Blaszczyszyn, B., Yogeshwaran, D., and Yukich, J. E. (2016). Limit theory for geometric statistics of clustering point processes. arXiv:1606.03988.

[Bobrowski and Kahle, 2014] Bobrowski, O. and Kahle, M. (2014). Topology of random geometric complexes: A survey. arXiv:1409.4734.

[Bobrowski et al., 2016] Bobrowski, O., Kahle, M., and Skraba, P. (2016). Maximally Persistent Cycles in Random Geometric Complexes. arXiv:1509.04347 , to appear in Ann. Appl. Prob.

[Bobrowski and Mukherjee, 2015] Bobrowski, O. and Mukherjee, S. (2015). The topology of probability distributions on manifolds. *Prob. Th. Rel. Fields*, 161:651–686.

[Bobrowski and Oliveira, 2017] Bobrowski, O. and Oliveira, G. (2017). Random \v{C}ech Complexes on Riemannian Manifolds. arXiv:1704.07204.

[Bobrowski and Weinberger, 2017] Bobrowski, O. and Weinberger, S. (2017). On the vanishing of homology in random Čech complexes. *Rand. Struct. Alg.*, 51(1):14–51.

[Carlsson, 2014] Carlsson, G. (2014). Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289–368.

[Chatterjee and Sen, 2017] Chatterjee, S. and Sen, S. (2017). Minimal spanning trees and steins method. *Ann. Appl. Probab.*, 27(3):1588–1645.

[Duy, 2017] Duy, T. K. (2017). A remark on the convergence of betti numbers in the thermodynamic regime. *Pac. J. Math. for Industry*, 9(1).

[Duy et al., 2016] Duy, T. K., Hiraoka, Y., and Shirai, T. (2016). Limit theorems for persistence diagrams. arXiv:1612.08371.

[Edelsbrunner and Harer, 2010] Edelsbrunner, H. and Harer, J. (2010). *Computational Topology, An Introduction*. American Mathematical Society, Providence, RI.

[Ghrist, 2014] Ghrist, R. (2014). *Elementary Applied Topology*. Createspace.

[Göring et al., 2013] Göring, D., Klatt, M., Stegmann, C., and Mecke, K. (2013). Morphometric analysis in gamma-ray astronomy using Minkowski functionals-source detection via structure quantification. *Astronomy & Astrophysics*, 555:A38.

[Grote and Thäle, 2016] Grote, J. and Thäle, C. (2016). Gaussian polytopes: a cumulant-based approach. arXiv:1602.06148.

[Haenggi, 2012] Haenggi, M. (2012). *Stochastic geometry for wireless networks*. Cambridge University Press.

[Heinrich and Schmidt, 1985] Heinrich, L. and Schmidt, V. (1985). Normal convergence of multidimensional shot noise and rates of this convergence. *Adv. Appl. Prob.*, 17(4):709–730.

[Hiraoka et al., 2016] Hiraoka, Y., Nakamura, T., Hirata, A., Escolar, E. G., Matsue, K., and Nishiura, Y. (2016). Hierarchical structures of amorphous solids characterized by persistent homology. *Proceedings of the National Academy of Sciences*, 113(26):7035–7040.

[Hiraoka and Shirai, 2015] Hiraoka, Y. and Shirai, T. (2015). Minimum spanning acycle and lifetime of persistent homology in the Linial-Meshulam process. arXiv:1503.05669.

[Kahle, 2011] Kahle, M. (2011). Random geometric complexes. *Discrete Comput. Geom.*, 45(3):553–573.

[Kahle, 2014] Kahle, M. (2014). Topology of random simplicial complexes: a survey. In *Algebraic topology: applications and new directions*, volume 620 of *Contemp. Math.*, pages 201–221. Amer. Math. Soc., Providence, RI.

[Kesten and Lee, 1996] Kesten, H. and Lee, S. (1996). The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.*, 6(2):495–527.

[Klette and Rosenfeld, 2004] Klette, R. and Rosenfeld, A. (2004). *Digital geometry: geometric methods for digital picture analysis*. Elsevier.

[Kong and Rosenfeld, 1989] Kong, T. Y. and Rosenfeld, A. (1989). Digital topology: introduction and survey. *Computer Vision, Graphics, and Image Processing*, 48(3):357–393.

[Last et al., 2016] Last, G., Peccati, G., and Schulte, M. (2016). Normal approximation on poisson spaces: Mehler's formula, second order poincaré inequalities and stabilization. *Prob. Th. Rel. Fields*, 165(3):667–723.

[Last and Penrose, 2017] Last, G. and Penrose, M. D. (2017+). *Lectures on the Poisson Process.* to be published as IMS Textbook by Cambridge University Press. `http://www.math.kit.edu/stoch/~last/page/lectures_on_the_poisson_process/en`.

[Munkres, 1996] Munkres, J. R. (1996). *Elements Of Algebraic Topology,* 1st edition. Westview Press.

[Owada, 2016] Owada, T. (2016). Limit Theorems for the Sum of Persistence Barcodes. arXiv:1604.04058.

[Peccati and Reitzner, 2016] Peccati, G. and Reitzner, M., editors (2016). *Stochastic analysis for Poisson point processes. Malliavin calculus, Wiener-Itô chaos expansions and stochastic geometry*, volume 7 of *Bocconi & Springer Series*. Bocconi University Press, Springer, [Cham].

[Penrose, 2003] Penrose, M. (2003). *Random Geometric Graphs.* Oxford University Press, New York.

[Penrose and Yukich, 2001] Penrose, M. and Yukich, J. (2001). Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.*, 11(4):1005–1041.

[Penrose, 1997] Penrose, M. D. (1997). The longest edge of the random minimal spanning tree. *Ann. Appl. Prob.*, pages 340–361.

[Penrose and Yukich, 2013] Penrose, M. D. and Yukich, J. E. (2013). Limit theory for point processes in manifolds. *Ann. Appl. Probab.*, 23(6):2161–2211.

[Ram Reddy et al., 2017] Ram Reddy, T., Vadlamani, S., and Yogeshwaran, D. (2017). Central limit theorem for quasi-local statistics of spin models on Cayley graphs. arXiv:1709.10424.

[Schneider and Weil, 2008] Schneider, R. and Weil, W. (2008). *Stochastic and Integral Geometry.* Probability and its Applications (New York). Springer-Verlag, Berlin.

[Schreiber and Yukich, 2008] Schreiber, T. and Yukich, J. (2008). Variance asymptotics and central limit theorems for generalized growth processes with applications to convex hulls and maximal points. *Ann. Probab.*, 36(1):363–396.

[Skraba et al., 2017] Skraba, P., Thoppe, G., and Yogeshwaran, D. (2017). Randomly Weighted $d-$complexes: Minimal Spanning Acycles and Persistence Diagrams. arXiv:1701.00239v1.

[Svane, 2017] Svane, A. (2017). *Valuations in image analysis*, pages 435–454. Springer International Publishing, Cham.

[Yogeshwaran and Adler, 2015] Yogeshwaran, D. and Adler, R. J. (2015). On the topology of random complexes built over stationary point processes. *Ann. Appl. Probab.*, 25(6):3338–3380.

[Yogeshwaran et al., 2017] Yogeshwaran, D., Subag, E., and Adler, R. J. (2017). Random geometric complexes in the thermodynamic regime. *Prob. Th. Rel. Fields*, 167(1-2):107–142.

[Yukich, 2006] Yukich, J. E. (2006). *Probability theory of classical Euclidean optimization problems.* Springer.

[Yukich, 2013] Yukich, J. E. (2013). Limit theorems in discrete stochastic geometry. In *Stochastic geometry, spatial statistics and random fields*, volume 2068 of *Lecture Notes in Math.*, pages 239–275. Springer, Heidelberg.