Limit theory of sparse random geometric graphs in high dimensions

Gilles Bonnet

joint work with Daniel Willhalm, Christian Hirsch & Daniel Rosen

GPSD Essen

March 7, 2023

Model

- ▷ $\mathcal{P}_d :=$ Poisson process in $\mathbb{R}^d$ with intensity $\lambda_d^d > 0$
- ▷ $\mathsf{GG}_d(t) = l_\infty$-Gilbert graph at parameter $t^{1/d}$ with $0 \leqslant t \leqslant 1$
- ▷ $W_d =$ cubical observation window

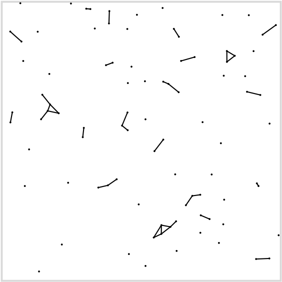
- $\mathcal{P}_d :=$ Poisson process in $\mathbb{R}^d$ with intensity $\lambda_d^d > 0$
- $\mathsf{GG}_d(t) = l_\infty$-Gilbert graph at parameter $t^{1/d}$ with $0 \leqslant t \leqslant 1$
- $W_d =$ cubical observation window

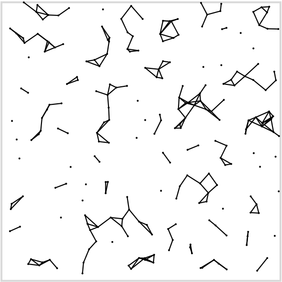
- $\mathsf{GG}_d^*(t) := \bigcup_{G \in \mathsf{Comp}_d(t)} G$
  - $\mathsf{Comp}_d(t) := \{G \subseteq \mathsf{GG}_d(t) \colon G$ ***is a component centered in*** $W_d\}$
  - ***Center of*** $G =$ smallest vertex in lexicographic order

- $\mathcal{P}_d :=$ Poisson process in $\mathbb{R}^d$ with intensity $\lambda_d^d > 0$
- $\mathsf{GG}_d(t) = l_\infty$-Gilbert graph at parameter $t^{1/d}$ with $0 \leqslant t \leqslant 1$
- $W_d =$ cubical observation window

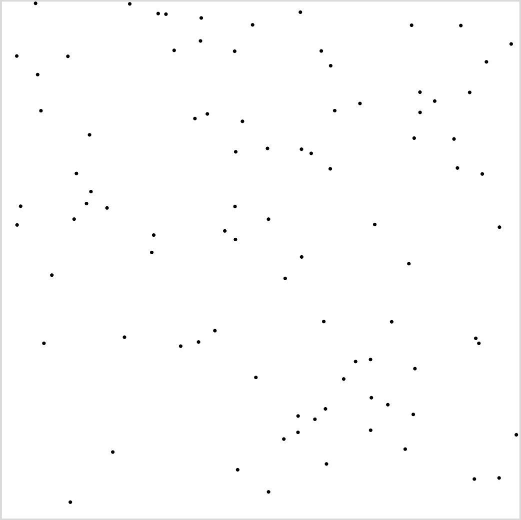
- $\mathsf{GG}_d^*(t) := \bigcup_{G \in \mathsf{Comp}_d(t)} G$
  - $\mathsf{Comp}_d(t) := \{G \subseteq \mathsf{GG}_d(t) \colon G$ ***is a component centered in*** $W_d\}$
  - ***Center of*** $G =$ smallest vertex in lexicographic order

Goal. Describe topological and geometric functionals as $\lambda_d \xrightarrow{d\to\infty} 0$ ***and*** $|W_d| \xrightarrow{d\to\infty} \infty$



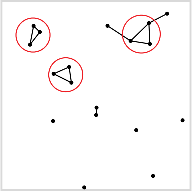


Animation of $(\mathrm{GG}^*_d(t))_{t \leqslant 1}$



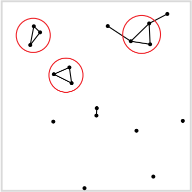
t = 0

Subgraph count



- G_0 = Fixed graph
- $\#\{G' \subseteq \mathsf{GG}_d^*(t) \colon G' \cong G_0\}$
- $G_0 = \Delta \rightsquigarrow$ Triangle count

Subgraph count



- ▷ G_0 = Fixed graph
- ▷ $\#\{G' \subseteq \mathsf{GG}^*_d(t) \colon G' \cong G_0\}$
- ▷ $G_0 = \Delta \rightsquigarrow$ Triangle count

Betti numbers

- ▷ Geometric graph gives rise to abstract simplicial complexes, e.g. the Rips complex.
- ▷ Functional as ith Betti number of the clique complex on the graph G
- ▷ β_0 = number of connected components
- ▷ β_1 = number of loops
- ▷ β_2 = number of cavities

1 Additive functionals

2 Multi-additive functionals

3 Proof of CLT

4 Proof of Poisson Approximation

5 Outlook

Additive functionals and moment asymptotics

Goal. Describe $d \uparrow \infty$-asymptotics of $A_{d,t} := a(\mathsf{GG}^*_d(t))$, where

- $a(\cdot) =$ Nonnegative functional on abstract graphs
- Assume ***additivity*** $a(G \sqcup G') = a(G) + a(G')$

Goal. Describe $d \uparrow \infty$-asymptotics of $A_{d,t} := a(\mathsf{GG}^*_d(t))$, where

- $a(\cdot) =$ Nonnegative functional on abstract graphs
- Assume ***additivity*** $a(G \sqcup G') = a(G) + a(G')$

Scaling for $\mathbb{E}[A_{d,t}]$. Set

$$\rho_d := |W_d|\lambda_d^{(k_0+1)d} \max_{G \in \mathcal{A}_{k_0}} v(G)^d$$

- $\mathcal{A}_k :=$ Connected graphs on $\{0, \dots, k\}$ with $a(G) > 0$
- $k_0 := \min\{k \geqslant 0 \colon \mathcal{A}_k \neq \varnothing\}$
- $v(G)^d :=$ ***"integral over indicator that $k_0 + 1$ points form graph at least G"***
- $\mathcal{A}'_{k_0} = \{G \colon v(G) = \max_{G' \in \mathcal{A}_{k_0}} v(G')\}$

Goal. Describe $d \uparrow \infty$-asymptotics of $A_{d,t} := a(\mathsf{GG}^*_d(t))$, where

- $a(\cdot) =$ Nonnegative functional on abstract graphs
- Assume ***additivity*** $a(G \sqcup G') = a(G) + a(G')$

Scaling for $\mathbb{E}[A_{d,t}]$***.*** Set

$$\rho_d := |W_d|\lambda_d^{(k_0+1)d} \max_{G \in \mathcal{A}_{k_0}} v(G)^d$$

- $\mathcal{A}_k :=$ Connected graphs on $\{0, \dots, k\}$ with $a(G) > 0$
- $k_0 := \min\{k \geqslant 0 \colon \mathcal{A}_k \neq \varnothing\}$
- $v(G)^d :=$ ***"integral over indicator that*** $k_0 + 1$ ***points form graph at least*** G***"***
- $\mathcal{A}'_{k_0} = \{G \colon v(G) = \max_{G' \in \mathcal{A}_{k_0}} v(G')\}$

Theorem (Moment asymptotics)

Assume that a *is additive and that* $a(G) \in e^{O(|G|)}$*. Then, for* $0 \leqslant t \leqslant t' \leqslant 1$*,*

- $\mathbb{E}[A_{d,t}] \sim \frac{\rho_d t^{k_0}}{(k_0+1)!} \sum_{G \in \mathcal{A}'_{k_0}} a(G)$, *as* $d \to \infty$
- $\mathrm{Cov}(A_{d,t'}, A_{d,t}) \sim \frac{\rho_d t^{k_0}}{((k_0+1)!)^2} \sum_{G \in \mathcal{A}'_{k_0}} a(G)^2$, *as* $d \to \infty$

- $r(t) := \frac{t^{k_0}}{((k_0+1)!)^2} \sum_{G \in \mathcal{A}'_{k_0}} a(G)^2$
- $(B_t)_{t \leqslant 1} =$ Standard Brownian motion
- $(N_t^{(G)})_{t \leqslant 1} =$ Poisson process with intensity $K k_0((k_0+1)!)^{-1} t^{k_0-1} \mathrm{d}t$
  - Independent for different $G \in \mathcal{A}'_{k_0}$
  - $K := \lim_{d \to \infty} \rho_d$ if it exists

- $r(t) := \frac{t^{k_0}}{((k_0+1)!)^2} \sum_{G\in\mathcal{A}'_{k_0}} a(G)^2$
- $(B_t)_{t\leqslant 1} =$ Standard Brownian motion
- $(N_t^{(G)})_{t\leqslant 1} =$ Poisson process with intensity $Kk_0((k_0+1)!)^{-1}t^{k_0-1}\mathrm{d}t$
  - Independent for different $G \in \mathcal{A}'_{k_0}$
  - $K := \lim_{d\to\infty} \rho_d$ if it exists

Theorem (CLTs and Poisson approximation)

Assume that a is additive and that $a(G) \in e^{O(|G|)}$.

1. *If $\rho_d^{1/d} \uparrow \infty$:* $(\rho_d^{-1/2}(A_{d,t} - \mathbb{E}[A_{d,t}]))_{t\leqslant 1} \overset{\text{fidi}}{\Rightarrow} (B_{r(t)})_{t\leqslant 1}$
2. *If $\rho_d^{1/d} \uparrow \infty$ & $\lambda_d\rho_d^{2/(3d)} \to 0$:* $(\rho_d^{-1/2}(A_{d,t} - \mathbb{E}[A_{d,t}]))_t \overset{Skorokhod}{\Longrightarrow} (B_{r(t)})_t$
3. *If $\rho_d \to K$:* $(A_{d,t})_{t\leqslant 1} \overset{Skorokhod}{\Longrightarrow} \left(\sum_{G\in\mathcal{A}'_{k_0}} N_t^{(G)} a(G)\right)_{t\leqslant 1}$

Remark: All convergences as dimension d tends to ∞.

Goal. Describe $d \uparrow \infty$-asymptotics of $A_{d,\boldsymbol{t}} := a(\mathsf{GG}^*_d(\boldsymbol{t}))$, where

- ▷ $\boldsymbol{t} := (t_1, \dots, t_m)$ for $0 \le t_1 \le \cdots \le t_m \le 1$
- ▷ $\mathsf{GG}^*_d(\boldsymbol{t}) := \big(\mathsf{GG}_d(t_1) \cap \mathsf{GG}^*_d(t_m), \dots, \mathsf{GG}_d(t_{m-1}) \cap \mathsf{GG}^*_d(t_m), \mathsf{GG}^*_d(t_m)\big)$
- ▷ $a(\cdot) = m$-variate nonnegative functional on abstract graphs

Goal. Describe $d \uparrow \infty$-asymptotics of $A_{d,\boldsymbol{t}} := a(\mathsf{GG}^*_d(\boldsymbol{t}))$, where

- ▷ $\boldsymbol{t} := (t_1, \dots, t_m)$ for $0 \le t_1 \le \cdots \le t_m \le 1$
- ▷ $\mathsf{GG}^*_d(\boldsymbol{t}) := \big(\mathsf{GG}_d(t_1) \cap \mathsf{GG}^*_d(t_m), \dots, \mathsf{GG}_d(t_{m-1}) \cap \mathsf{GG}^*_d(t_m), \mathsf{GG}^*_d(t_m)\big)$
- ▷ $a(\cdot) = m$-variate nonnegative functional on abstract graphs

Assume ***multi-additivity.***

$$a(\boldsymbol{R} \cup \boldsymbol{R}') = a(\boldsymbol{R}) + a(\boldsymbol{R}')$$

- ▷ where $\boldsymbol{R} = (G_1, \dots, G_m)$ with $G_1 \subseteq \cdots \subseteq G_m$ and $\boldsymbol{R}' = (G'_1, \dots, G'_m)$ with $G'_1 \subseteq \cdots \subseteq G'_m$ and $G_m \cap G'_m = \varnothing$.

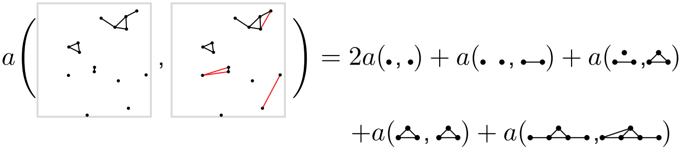
Goal. Describe $d \uparrow \infty$-asymptotics of $A_{d,\boldsymbol{t}} := a(\mathsf{GG}^*_d(\boldsymbol{t}))$, where

- $\boldsymbol{t} := (t_1, \dots, t_m)$ for $0 \le t_1 \le \dots \le t_m \le 1$
- $\mathsf{GG}^*_d(\boldsymbol{t}) := \big(\mathsf{GG}_d(t_1) \cap \mathsf{GG}^*_d(t_m), \dots, \mathsf{GG}_d(t_{m-1}) \cap \mathsf{GG}^*_d(t_m), \mathsf{GG}^*_d(t_m)\big)$
- $a(\cdot) = m$-variate nonnegative functional on abstract graphs

Assume ***multi-additivity.***

$$a(\boldsymbol{R} \cup \boldsymbol{R}') = a(\boldsymbol{R}) + a(\boldsymbol{R}')$$

- where $\boldsymbol{R} = (G_1, \dots, G_m)$ with $G_1 \subseteq \dots \subseteq G_m$ and $\boldsymbol{R}' = (G'_1, \dots, G'_m)$ with $G'_1 \subseteq \dots \subseteq G'_m$ and $G_m \cap G'_m = \varnothing$.



Linear combinations of univariate functionals

$a(G_1, \dots, G_m) := \alpha_1 a'(G_1) + \cdots + \alpha_m a'(G_m)$, where $\alpha_1, \dots, \alpha_m \geqslant 0$ and where a' is a nonnegative additive functional.

Linear combinations of univariate functionals

$a(G_1, \ldots, G_m) := \alpha_1 a'(G_1) + \cdots + \alpha_m a'(G_m)$, where $\alpha_1, \ldots, \alpha_m \geqslant 0$ and where a' is a nonnegative additive functional.

Dynamic subgraph count

$(G_{0,1}, \ldots, G_{0,m})$ fixed sequence of graphs and
$a(G_1, \ldots, G_m) := \#\{G'_1 \subseteq \cdots \subseteq G'_m \colon G'_i \subseteq G_i$ and $G'_i \cong G_{0,i}$ for all $i \leqslant m\}$

Linear combinations of univariate functionals

$a(G_1, \dots, G_m) := \alpha_1 a'(G_1) + \cdots + \alpha_m a'(G_m)$, where $\alpha_1, \dots, \alpha_m \geqslant 0$ and where a' is a nonnegative additive functional.

Dynamic subgraph count

$(G_{0,1}, \dots, G_{0,m})$ fixed sequence of graphs and
$a(G_1, \dots, G_m) := \#\{G'_1 \subseteq \cdots \subseteq G'_m : G'_i \subseteq G_i$ and $G'_i \cong G_{0,i}$ for all $i \leqslant m\}$

Persistent Betti numbers

$a_i(G, G')$ as ith persistent Betti number associated with the clique complexes of the graphs $G \subseteq G'$.

Linear combinations of univariate functionals

$a(G_1, \ldots, G_m) := \alpha_1 a'(G_1) + \cdots + \alpha_m a'(G_m)$, where $\alpha_1, \ldots, \alpha_m \geqslant 0$ and where a' is a nonnegative additive functional.

Dynamic subgraph count

$(G_{0,1}, \ldots, G_{0,m})$ fixed sequence of graphs and
$a(G_1, \ldots, G_m) := \#\{G'_1 \subseteq \cdots \subseteq G'_m : G'_i \subseteq G_i$ and $G'_i \cong G_{0,i}$ for all $i \leqslant m\}$

Persistent Betti numbers

$a_i(G, G')$ as ith persistent Betti number associated with the clique complexes of the graphs $G \subseteq G'$.

Theorem (CLT for multi-additive functionals)

Assume that a is multi-additive, dominated and that $a(G, \cdots, G) \in e^{O(|G|)}$. Let $\boldsymbol{t} = (t_1, \ldots, t_m)$ with $0 \leqslant t_1 \leqslant \cdots \leqslant t_m \leqslant 1$.

1. *If $\rho_d^{1/d} \uparrow \infty$: $\mathbb{E}[A_{d,\boldsymbol{t}}] \asymp \rho_d$ and $\mathsf{Var}[A_{d,\boldsymbol{t}}] \asymp \rho_d$.*
2. *If $\rho_d^{1/d} \uparrow \infty$: $(\mathsf{Var}[A_{d,\boldsymbol{t}}])^{-1/2}(A_{d,\boldsymbol{t}} - \mathbb{E}[A_{d,\boldsymbol{t}}]) \Rightarrow \mathcal{N}(0,1)$*

Proof of multivariate CLT

- ***Cramér-Wold theorem*** $\rightsquigarrow$ consider $\sum_{i \leqslant m} c_i A_{d,t_i}$ as $d \uparrow \infty$.

Proof of multivariate CLT

- ***Cramér-Wold theorem*** $\rightsquigarrow$ consider $\sum_{i\leqslant m} c_i A_{d,t_i}$ as $d \uparrow \infty$.
- Using the moment (variance) asymptotics, it follows that we can ignore components with more than $k_0 + 1$ vertices.

- ***Cramér-Wold theorem*** $\rightsquigarrow$ consider $\sum_{i\leqslant m} c_i A_{d,t_i}$ as $d \uparrow \infty$.
- Using the moment (variance) asymptotics, it follows that we can ignore components with more than $k_0 + 1$ vertices.
- After normalizing the centered process, we need to show that such an expression converges to a standard normal distribution. The variance asymptotics are handled by part 1 of the theorem.

Proof of multivariate CLT

- ***Cramér-Wold theorem*** $\rightsquigarrow$ consider $\sum_{i\leqslant m} c_i A_{d,t_i}$ as $d\uparrow\infty$.
- Using the moment (variance) asymptotics, it follows that we can ignore components with more than k_0+1 vertices.
- After normalizing the centered process, we need to show that such an expression converges to a standard normal distribution. The variance asymptotics are handled by part 1 of the theorem.
- Divide W_d in a grid consisting of boxes of side length about $2k_0$ and represent the functional of interest as a sum of random variables restricted to the boxes of the grid.

- ***Cramér-Wold theorem*** $\rightsquigarrow$ consider $\sum_{i \leqslant m} c_i A_{d,t_i}$ as $d \uparrow \infty$.
- Using the moment (variance) asymptotics, it follows that we can ignore components with more than $k_0 + 1$ vertices.
- After normalizing the centered process, we need to show that such an expression converges to a standard normal distribution. The variance asymptotics are handled by part 1 of the theorem.
- Divide W_d in a grid consisting of boxes of side length about $2k_0$ and represent the functional of interest as a sum of random variables restricted to the boxes of the grid.
- Apply [Penrose, 2003, Theorem 2.4] ***(normal approximation for a sum of weakly dependent variables by Stein's method)***

Proof of functional CLT for additive functionals

By [Billingsley, 1999, Theorem 13.1] ***finite-dimensional convergence in distribution*** and ***tightness*** is sufficient for the ***convergence in distribution on a process level***.

By [Billingsley, 1999, Theorem 13.1] ***finite-dimensional convergence in distribution*** and ***tightness*** is sufficient for the ***convergence in distribution on a process level***.

To show tightness, we can show a ***condition for the 4th moment***. More, precisely that there exist $c, \varepsilon > 0$ such that

$$\rho_d^{-2}\mathbb{E}[\bar{A}_d(E)^4] \leqslant c|E|^{1+\varepsilon}$$

for all $d \geqslant 1$ and all intervals $E = [t_-, t_+] \subseteq [0, 1]$, where $\bar{A}_d(E)$ denotes the centered increment.

By [Billingsley, 1999, Theorem 13.1] ***finite-dimensional convergence in distribution*** and ***tightness*** is sufficient for the ***convergence in distribution on a process level***.

To show tightness, we can show a ***condition for the 4th moment***. More, precisely that there exist $c, \varepsilon > 0$ such that

$$\rho_d^{-2}\mathbb{E}[\bar{A}_d(E)^4] \leqslant c|E|^{1+\varepsilon}$$

for all $d \geqslant 1$ and all intervals $E = [t_-, t_+] \subseteq [0,1]$, where $\bar{A}_d(E)$ denotes the centered increment.

Strategy:

- Reduce the moment condition to interval that are not too small.

By [Billingsley, 1999, Theorem 13.1] ***finite-dimensional convergence in distribution*** and ***tightness*** is sufficient for the ***convergence in distribution on a process level***.

To show tightness, we can show a ***condition for the 4th moment***. More, precisely that there exist $c, \varepsilon > 0$ such that

$$\rho_d^{-2}\mathbb{E}[\bar{A}_d(E)^4] \leqslant c|E|^{1+\varepsilon}$$

for all $d \geqslant 1$ and all intervals $E = [t_-, t_+] \subseteq [0, 1]$, where $\bar{A}_d(E)$ denotes the centered increment.

Strategy:

- Reduce the moment condition to interval that are not too small.
- Use the ***cumulant identity***: Let Z be a random variable with finite fourth moment. Then, Z satisfies $\mathbb{E}[(Z - \mathbb{E}[Z])^4] = 3\mathrm{Var}(Z)^2 + c_4(Z)$. The fourth order cumulant is the 4th derivative of the cumulant-generating function $s \mapsto \log \mathbb{E}[e^{sX}]$ evaluated at 0.

By [Billingsley, 1999, Theorem 13.1] ***finite-dimensional convergence in distribution*** and ***tightness*** is sufficient for the ***convergence in distribution on a process level***.

To show tightness, we can show a ***condition for the 4th moment***. More, precisely that there exist $c, \varepsilon > 0$ such that

$$\rho_d^{-2}\mathbb{E}[\bar{A}_d(E)^4] \leqslant c|E|^{1+\varepsilon}$$

for all $d \geqslant 1$ and all intervals $E = [t_-, t_+] \subseteq [0, 1]$, where $\bar{A}_d(E)$ denotes the centered increment.

Strategy:

- Reduce the moment condition to interval that are not too small.
- Use the ***cumulant identity***: Let Z be a random variable with finite fourth moment. Then, Z satisfies $\mathbb{E}[(Z - \mathbb{E}[Z])^4] = 3\text{Var}(Z)^2 + c_4(Z)$. The fourth order cumulant is the 4th derivative of the cumulant-generating function $s \mapsto \log \mathbb{E}[e^{sX}]$ evaluated at 0.
- Derive bounds on the variance and the fourth-order cumulant c_4.

General approach similar to CLT: For high dimension d only components in $\mathcal{A}'_{k_0}$ significantly contribute to the functional.

General approach similar to CLT: For high dimension d only components in $\mathcal{A}'_{k_0}$ significantly contribute to the functional.

- Restrict $A_{d,t}$ to contributions from ***subgraphs in $\mathcal{A}'_{k_0}$***, and that, once formed, ***do not change over time*** anymore. Call it A'_{d,t,k_0}. Then, A'_{d,t,k_0} approximates $A_{d,t}$.

General approach similar to CLT: For high dimension d only components in $\mathcal{A}'_{k_0}$ significantly contribute to the functional.

- Restrict $A_{d,t}$ to contributions from ***subgraphs in*** $\mathcal{A}'_{k_0}$, and that, once formed, ***do not change over time*** anymore. Call it A'_{d,t,k_0}. Then, A'_{d,t,k_0} approximates $A_{d,t}$.
- With a general ***Poisson approximation result*** [Decreusefond et al., 2016, Theorem 3.1], we show that the sum in A'_{d,t,k_0} over $(k_0 + 1)$-tuples of points can be approximated by a sum over a Poisson process. $\Rightarrow$ Approximate A'_{d,t,k_0} by a Poisson functional.

General approach similar to CLT: For high dimension d only components in $\mathcal{A}'_{k_0}$ significantly contribute to the functional.

- Restrict $A_{d,t}$ to contributions from ***subgraphs in $\mathcal{A}'_{k_0}$***, and that, once formed, ***do not change over time*** anymore. Call it A'_{d,t,k_0}. Then, A'_{d,t,k_0} approximates $A_{d,t}$.
- With a general ***Poisson approximation result*** [Decreusefond et al., 2016, Theorem 3.1], we show that the sum in A'_{d,t,k_0} over (k_0+1)-tuples of points can be approximated by a sum over a Poisson process. $\Rightarrow$ Approximate A'_{d,t,k_0} by a Poisson functional.
- Categorizing the Poisson process into the graphs in $\mathcal{A}'_{k_0}$ that are formed yields ***independent*** Poisson Processes.

Outlook

- CLT & Poisson approximation for Gilbert graph in high dimensions ✓
- Large deviations **!**
- $\lambda_d < 1/2$ but not dependent on dimension d **!**

Billingsley, P. (1999).
Convergence of Probability Measures.
John Wiley & Sons, New York, second edition.

Decreusefond, L., Schulte, M., and Thäle, C. (2016).
Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry.
Ann. Probab., 44(3):2147–2197.

Kallenberg, O. (2017).
Random Measures, Theory and Applications.
Springer, Cham.

Penrose, M. (2003).
Random geometric graphs, volume 5 of *Oxford Studies in Probability*.
Oxford University Press, Oxford.