Component counts of dense random geometric graphs

Mathew Penrose*
(University of Bath, UK)

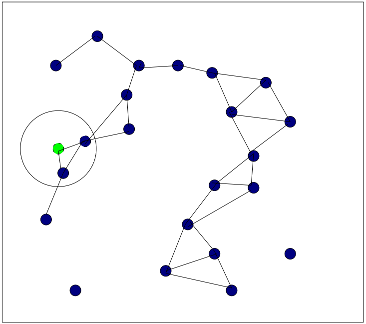
Workshop ‘Stochastic Geometry in Action’
Bath, September 2024

* Joint work with Xiaochuan Yang (Banque Internationale à Luxembourg.)
Supported by EPSRC grant EP/T028653/1

- Penrose, M.D. and Yang, X. (2023) On k-clusters of high-intensity random geometric graphs (v3). Arxiv
- Penrose, M.D. and Yang, X. (2024+) On the components of random geometric graphs in the dense limit. In preparation.

Random geometric graphs

Let $d \in \mathbb{N}$, $r > 0$. Given locally finite $F \subset \mathbb{R}^d$, define the geometric graph $G(F, r) = (V, E)$ with $V = F$ and $\{x, y\} \in E \Leftrightarrow \|x - y\| \leq r$.



Let $A \subset \mathbb{R}^d$ be compact with ∂A smooth and $\mathrm{vol}(A) = 1$.
Let $X_1, X_2, \ldots$ be independent uniform random points in A.
Let $F_n := \{X_1, \ldots, X_n\}$. Given $(r_n)_{n \geq 1}$, consider $G(F_n, r_n)$ as $n \to \infty$.

Note if $r_n \to 0$, $\mathbb{E}[\mathrm{Deg}(X_1)] \sim \theta n r_n^d$ as $n \to \infty$, where $\theta := \theta_d := \mathrm{vol}(B_1)$.

The k-component count

Fix $k \in \mathbb{N}$; let $S_{n,k} :=$ number of components of order k in $G(F_n, r_n)$;

Case 1: Suppose $nr_n^d \to \lambda$ as $n \to \infty$. In this case (see Penrose 03) $n^{-1}\mathbb{E}[S_{n,k}] \to k^{-1}p_k(\lambda)$ as $n \to \infty$, and for some $\sigma > 0$,

$$n^{-1/2}(S_{n,k} - \mathbb{E}[S_{n,k}]) \xrightarrow{\mathcal{D}} N(0, \sigma).$$

Here $p_k(\lambda) = p_k(\lambda, d) \in (0, \infty)$ (we'll discuss later).
Note $p_1(\lambda) = \lim_{n\to\infty} \mathbb{P}[\mathrm{Deg}(X_1) = 0] = e^{-\lambda\theta}$.

Case 2: Suppose $d \geq 2$ and $\liminf(n\theta r_n^d/(\log n)) > 2 - 2/d$. Then as $n \to \infty$, $\mathbb{P}[\sum_k S_{n,k} = 1] \to 1$ (Penrose 03), so $\mathbb{P}[S_{n,k} = 0] \to 1$.

Case 0: Suppose $k \geq 2$ and $n(nr_n^d)^{k-1} \to c \in (0, \infty)$ (so $nr_n^d \to 0$). Then $S_{n,k}$ is asymptoticlaly Poisson (see P. 03).

All limiting regimes are of interest (e.g. in Topological Data Analysis).

Case 1.5 (in between Cases 1 and 2)

THEOREM 1 (P. and Yang 2023) Suppose that $d \geq 2$ and that

$$\lim_{n\to\infty} (nr_n^d) = \infty \quad \text{and} \quad \limsup_{n\to\infty} (n\theta r_n^d / \log n) < 2/d. \tag{1}$$

Let $I_{n,k} := \mathbb{E}[S_{n,k}]$, the mean number of k-components in $G(F_n, r_n)$. Then there exists $\alpha_k \in (0,\infty)$ such that as $n \to \infty$ with $k \in \mathbb{N}$ fixed:

(i) $kI_{n,k} \sim \alpha_k n (nr_n^d)^{(1-k)(d-1)} \exp(-\theta n r_n^d)$; **(Note:** $\alpha_1 = 1$)

(ii) $\mathbb{V}\mathrm{ar}(S_{n,k}) \sim I_{n,k}$;

(iii) $\frac{S_{n,k} - \mathbb{E}[S_{n,k}]}{\sqrt{I_{n,k}}} \xrightarrow{\mathcal{D}} N(0,1)$.

These results also hold for $S'_{n,k}$, defined like $S_{n,k}$ but with a Poisson(n) number of points.

P+Y also give bounds on the rates of convergence (not presented here). Note: $2/d \leq 2 - 2/d$ which was the threshold for Case 2 (connectivity).

Singletons in Case 1.8 (between Cases 1.5 and 2)

THEOREM 2 (P. and Yang 2024+) Suppose that $d \geq 3$ and $nr_n^d \to \infty$ and

$$\liminf_{n\to\infty}(nr_n^d/\log n) > 2/d; \quad (2)$$

$$\limsup_{n\to\infty}(nr_n^d/\log n) < 2 - 2/d. \quad (3)$$

Let $S_n := S_{n,1}$, the number of isolated vertices (singletons) in $G(F_n, r_n)$. Let $I_n := \mathbb{E}[S_n]$, $I'_n := \mathbb{E}[S'_n]$ where $S'_n := S'_{n,1}$. Then as $n \to \infty$:

(i) $I_n \sim I'_n \sim \theta_{d-1}^{-1}|\partial A| r_n^{1-d} \exp(-\theta n r_n^d/2)$;

(ii) $\mathbb{V}\mathrm{ar}(S_n) \sim I_n$;

(iii) $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{I_n}} \xrightarrow{\mathcal{D}} N(0,1)$.

Results (ii) and (iii) also hold for S'_n. Also we can relax condition (3) to

$$n\theta r_n^d - (2 - 2/d)\log n - \log\log n \to -\infty$$

which is equivalent to $I_n \to \infty$.

Number of components and giant component

Let $K_n := \sum_{k=1}^{n} S_{n,k}$ and $L_n := \max\{k : S_{n,k} > 0\}$ and $R_n := n - L_n$.

If $nr_n^d \to \lambda$, LLN and CLT for K_n and L_n are given in Penrose (2003).
Now let $nr_n^d \to \infty$ but $I_n \to \infty$ (eg $\limsup_{n\to\infty}(n\theta r_n^d/\log n) < 2 - 2/d$)
Ganesan '13 (for $A = [0,1]^2$): $\exists c > 0$ with $\mathbb{P}[R_n < ne^{-cnr_n^2}] \to 1$.

THM 3 (P. and Y. 2024+) If $d \geq 2$ and ξ_n denotes any of R_n, K_n, R'_n, K'_n,

$$\mathbb{E}[\xi_n] \sim I_n \sim ne^{-n\theta r_n^d} + e^{-n\theta r_n^d/2} r_n^{1-d} |\partial A| \quad \text{as} \quad n \to \infty$$

If also $d \geq 3$ or ξ_n is R'_n or K'_n, then $\mathbb{V}\mathrm{ar}[\xi_n] \sim I_n$ and

$$\frac{\xi_n - \mathbb{E}[\xi_n]}{I_n^{1/2}} \xrightarrow{\mathcal{D}} N(0,1).$$

If $\limsup(n\theta r_n^d/\log n) < \max(1/2, 1 - 2/d)$ then $\frac{\xi_n}{I_n} \to 1$ almost surely.

Idea of proof of Theorem 3 [$\xi_n = K_n, K'_n, R_n$ or R'_n]

For $0 < a < b < \infty$ let $\xi_{n,a,b}$ denote the contribution to ξ_n from components of Euclidean diameter in the range $(ar_n, br_n]$. Given $\rho > \varepsilon > 0$,

$$\xi_n - S_n = \xi_{n,0,\varepsilon} + \xi_{\varepsilon,\rho} + \xi_{\rho,\infty}.$$

Can choose ε small and ρ large such that all of

$$\mathbb{E}[\xi_{n,0,\varepsilon}], \quad \mathbb{E}[\xi_{n,\varepsilon,\rho}] \quad \mathbb{E}[\xi_{n,\rho,\infty}]$$

are $o(I_n)$. Likewise for

$$\mathbb{V}\mathrm{ar}[\xi_{n,0,\varepsilon}], \quad \mathbb{V}\mathrm{ar}[\xi_{n,\varepsilon,\rho}] \quad \mathbb{V}\mathrm{ar}[\xi_{n,\rho,\infty}]$$

except when $d = 2$ and $\xi_n = K_n$ or R_n.
[We can use spatial independence for $\mathbb{V}\mathrm{ar}[K'_{n,0,\varepsilon}]$ or $\mathbb{V}\mathrm{ar}[R'_{n,0,\varepsilon}]$ but have to use the Efron-Stein inequality for $\mathbb{V}\mathrm{ar}[K_{n,0,\varepsilon}]$ or $\mathbb{V}\mathrm{ar}[R_{n,0,\varepsilon}]$.]

Characterization of p_k and α_k

[Recall Case 1: if $nr_n^d \to \lambda$ then $\mathbb{E}[S_{n,k}/n] \to k^{-1}p_k(\lambda)$.
Theorem 1(i): if $nr_n^d \to \infty$ and $\limsup_{n\to\infty}(n\theta r_n^d/\log n) < 2/d$ then $k\mathbb{E}[S_{n,k}] \sim \alpha_k n(nr_n^d)^{(1-k)(d-1)}\exp(-\theta nr_n^d)$.]

Given $\lambda > 0$, let $\mathcal{H}_\lambda$ be a homogeneous Poisson point process in $\mathbb{R}^d$ with intensity λ. Let o be the origin in $\mathbb{R}^d$ and $\mathcal{H}_\lambda^o := \mathcal{H}_\lambda \cup \{o\}$. Then

$$p_k(\lambda) := \mathbb{P}[|\mathcal{C}_1(o, \mathcal{H}_\lambda^o)| = k],$$

where for $x \in F \subset \mathbb{R}^d$, $\mathcal{C}_r(x, F) := \{y \in F : x \leftrightarrow y \text{ in } G(F, r)\} \cup \{x\}$.

THEOREM A (Penrose and Yang 2023). Let $d, k \in \mathbb{N}$. As $\lambda \to \infty$,

$$p_{k+1}(\lambda) \sim \alpha_k e^{-\theta\lambda}\lambda^{-k(d-1)} \tag{4}$$

where for a certain 'energy' function $g(z_1, \ldots, z_k)$ to be defined later,

$$\alpha_k := \frac{1}{k!}\int_{(\mathbb{R}^d)^k} \exp(-g(z_1, \ldots, z_k))d(z_1, \ldots, z_k).$$

[Previously (Alexander 1993) $p_{k+1}(\lambda) = \Theta(e^{-\theta\lambda}\lambda^{-k(d-1)})$.]

Towards proving Theorem A: a formula for $p_k(\lambda)$

For $F \subset \mathbb{R}^d$, $r > 0$, let $F^r := \cup_{x \in F} B_r(x)$. Also set $h_r(F) := \mathbf{1}\{G(F,r)) \in \mathcal{K}\}$, where $\mathcal{K}$ is the class of connected graphs. Then $\mathbf{1}\{|\mathcal{C}_1(o, \mathcal{H}_\lambda^0)| = k+1\}$ equals $1/k!$ times

$$\sum_{x_1,\ldots,x_k \in \mathcal{H}_\lambda}^{\neq} h_1(\{o, x_1, \ldots, x_k\}) \mathbf{1}\{(\mathcal{H}_\lambda \setminus \{x_1, \ldots, x_k\}) \cap \{o, x_1, \ldots, x_k\}^1 = \emptyset\},$$

where $\sum^{\neq}$ means the the sum is over ordered k-tuples of distinct points in $\mathcal{H}_\lambda$. Thus by the multivariate Mecke formula (eg Last and Penrose 2018),

$$p_{k+1}(\lambda) = \frac{\lambda^k}{k!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} h_1(\{o, x_1, \ldots, x_k\}) \exp(-\lambda \mathrm{vol}(\{o, x_1, \ldots, x_k\}^1)) \, d(x_1, \ldots, x_k).$$

For short $p_{k+1}(\lambda) = \frac{\lambda^k}{k!} \int_{(\mathbb{R}^d)^k} h_1(o, \mathbf{x}) \exp(-\lambda V(o, \mathbf{x})) d\mathbf{x}$.

Idea for proof of Theorem A

For $\mathbf{x} = (x_1, \ldots, x_k) \in (\mathbb{R}^d)^k$ set $V(o, \mathbf{x}) = \mathrm{vol}(\{o, x_1, \ldots, x_k\}^1)$ and $V'(\mathbf{x}) := V(o, \mathbf{x}) - \theta$. Taking $\mathbf{z} = \lambda \mathbf{x}$,

$$
\begin{aligned}
p_{k+1}(\lambda) &= \frac{\lambda^k}{k!} \int_{(\mathbb{R}^d)^k} h_1(o, \mathbf{x}) \exp(-\lambda V(o, \mathbf{x})) d\mathbf{x} \\
&= \frac{\lambda^{k-kd} e^{-\theta\lambda}}{k!} \int_{(\mathbb{R}^d)^k} h_1(o, \lambda^{-1}\mathbf{z}) \exp(-\lambda V'(\lambda^{-1}\mathbf{z})) d\mathbf{z}.
\end{aligned}
$$

LEMMA: Let $\mathbf{z} \in (\mathbb{R}^d)^k$. Then $r^{-1} V'(r\mathbf{z}) \to g(\mathbf{z})$ (defined later) as $r \downarrow 0$.

Using the lemma (with $r = \lambda^{-1}$) and fact that the first factor tends to 1 as $\lambda \to \infty$ for all $\mathbf{z}$, and dominated convergence, gives

$$
k! e^{\lambda\theta} \lambda^{kd-k} p_{k+1}(\lambda) \to \int_{(\mathbb{R}^d)^k} \exp(-g(\mathbf{z})) d\mathbf{z}
$$

as $\lambda \to \infty$, which is a weak version of Theorem A.

Idea of proof of Lemma: $r^{-1}V'(r\mathbf{z}) \to g(\mathbf{z})$.

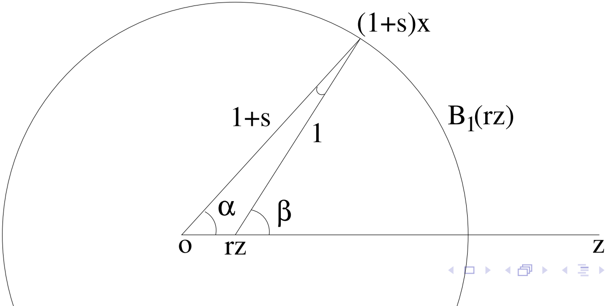
Recall $V'(\mathbf{x}) = \mathrm{vol}(\{o, x_1, \ldots, x_k\}^1) - \theta$. For $k = 1$,

$$r^{-1}V'(rz) \to \theta_{d-1} \quad \text{as } r \downarrow 0.$$

By some Euclidean geometry, given $x \in \partial B_1(o)$, with $s = s(x, rz)$ as shown and $\alpha(x, z) = \angle\, xoz$,

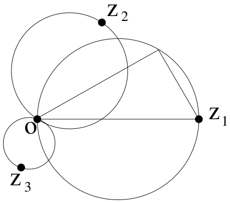
$$s(x, rz) \sim r\|z\|(\cos(\alpha(x, z)))^+ \quad \text{as } r \downarrow 0.$$

Using polar coordinates, with σ for Lebesgue surface measure on $\partial B_1(o)$,
$r^{-1}V(r\mathbf{z}) \to \int_{\partial B_1(o)} \max_{i \le k}(\|z_i\|(\cos\alpha(x, z_i))^+)\sigma(dx) =: g(\mathbf{z})$



Geometrical interpretation of $g(\mathbf{z})$, $\mathbf{z} \in (\mathbb{R}^d)^k$

From the last slide, $g(\mathbf{z}) := \int_{\partial B_1(o)} \max_{i \le k}(\|z_i\|(\cos\alpha(x, z_i))^+)\sigma(dx)$.



For $\mathbf{z} = (z_1, \ldots, z_k) \in (\mathbb{R}^d)^k$ set

$$D(\mathbf{z}) := \cup_{i=1}^k B_{\|z\|/2}((1/2)z_i).$$

Then (using Thales' Theorem)

$$g(\mathbf{z}) = \int_{D(\mathbf{z})} \|x\|^{1-d} d\mathbf{x}.$$

Interpretation for $d = 2$: the *gravitational energy* of $D(\mathbf{z})$ with respect to a large point mass at the origin.

THEOREM B (P. and Yang 2023)

Let $d, k \in \mathbb{N}$. If $Y_1, \ldots, Y_k$ denote the points of $\mathcal{C}_1(o, \mathcal{H}_\lambda^0) \setminus \{o\}$ taken from left to right, then

$$\mathbb{P}[(\lambda Y_1, \ldots, \lambda Y_k) \in d\mathbf{z} | (|\mathcal{C}_1(o, \mathcal{H}_\lambda^0)|) = k + 1] \Longrightarrow \mathbb{P}[(Z_{(1)}, \ldots, Z_{(k)}) \in d\mathbf{z}]$$

where $Z_{(1)}, \ldots Z_{(k)}$ are the points $Z_1, \ldots, Z_k$ taken from left to right, and

$$\mathbb{P}[(Z_1, \ldots, Z_k) \in d\mathbf{z}] = (k!\alpha_k)^{-1} \exp(-g(\mathbf{z}))d\mathbf{z}.$$

Re-statement of THEOREM 1 (Case 1.5)

THEOREM 1 (P. and Yang 2023) Suppose that $d \geq 2$ and that

$$\lim_{n\to\infty}(nr_n^d) = \infty \quad \text{and} \quad \limsup_{n\to\infty}(n\theta r_n^d/\log n) < 2/d. \tag{5}$$

Let $I_{n,k} := \mathbb{E}[S_{n,k}]$, the mean number of k-components in $G(F_n, r_n)$. Then there exists $\alpha_k \in (0,\infty)$ such that as $n \to \infty$ with $k \in \mathbb{N}$ fixed:

(i) $kI_{n,k} \sim \alpha_k n(nr_n^d)^{(1-k)(d-1)} \exp(-\theta nr_n^d)$; **(Note:** $\alpha_1 = 1$)

(ii) $\mathbb{V}\mathrm{ar}(S_{n,k}) \sim I_{n,k}$;

(iii) $\frac{S_{n,k}-\mathbb{E}[S_{n,k}]}{\sqrt{I_{n,k}}} \xrightarrow{\mathcal{D}} N(0,1)$.

These results also hold for $S'_{n,k}$, defined like $S_{n,k}$ but with a Poisson(n) number of points.

Ideas of proof of Theorem 1

(i) By scaling and binomial-Poisson approximation,

$$k\mathbb{E}[S_{n,k}] = n \int_A \mathbb{P}[|\mathcal{C}_{r_n}(x, F_{n-1} \cup \{x\})| = k]dx$$
$$\sim n\mathbb{P}[|\mathcal{C}_1(o, \mathcal{H}_{nr_n^d} \cup \{o\})| = k] = np_k(nr_n^d)$$

and (i) then follows by applying Theorem A.
(ii) Similar (but more involved) second moment computation.
(iii) Let $Y_i := \mathbf{1}\{|\mathcal{C}_{r_n}(X_i, F_n)| = k, X_i \prec X_j \ \forall X_j \in \mathcal{C}_{r_n}(X_i, F_n)\}$.
Poisson approximation using Stein's method/coupling (BHJ92)

$$d_{TV}(S_{n,k}, \mathrm{Po}(\mathbb{E}[S_{n,k}])) \leq \mathbb{E}[|S_{n,k} - V|]$$

for any coupled V satisfying $\mathcal{L}(V) = \mathcal{L}(S_{n,k} - 1|Y_1 = 1)$.
For V, resample $X_1, \ldots X_k$ with conditional law given $\{Y_1 = 1\} \cap \{C_{r_n}(X_1, F_n) = \{X_1, \ldots, X_k\}\}$. Then resample those $X_j, j > k$ which lie in $\{X_1, \ldots, X_k\}^1$. This works.

References

- Alexander, K.S. (1993) Finite clusters in continuous percolation: compression and sphericality. *Probab. Theory Related Fields* **97**, 35–63.
- Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.
- Last, G. and Penrose, M. (2018) *Lectures on the Poisson process.* Cambridge University Press.
- Penrose, M.D. (2003). *Random Geometric Graphs.* Oxford University Press.
- Penrose, M.D. and Yang, X. (2023) On k-clusters of high-intensity random geometric graphs (v3). Arxiv