LMS-EPSRC Short Course
Random Geometric Graphs (Draft notes)

Mathew Penrose
University of Bath
http://www.maths.bath.ac.uk∼masmdp

August 16, 2013

1 Introduction

Given finite $\mathcal{X} \subset \mathbb{R}^d$, and $r > 0$, the *geometric graph* $G(\mathcal{X}, r)$ has vertex set $\mathcal{X}$ and edge set $\{\{x, y\} : \|x - y\| \leq r\}$, where $\| \cdot \|$ is the Euclidean norm.

Motivation: radio stations/communications. Trees/disease. Stars/constellations.

A *random* geometric graph (RGG) is obtained by taking $\mathcal{X}$ to be a random set of points.

Let $\xi_1, \xi_2, \ldots$ be independent random d-vectors, uniformly distributed over $[0, 1]^d$ (typically $d = 2$). Set

$$\mathcal{X}_n := \{\xi_1, \ldots, \xi_n\}.$$

and $\mathcal{P}_n := \{\xi_1, \ldots, \xi_{N_n}\}$ with N_n Poisson distributed with parameter n, independent of $(\xi_1, \xi_2, \ldots)$. Then $\mathcal{P}_n$ is a *Poisson point process* with n times Lebesgue measure as its mean measure, i.e.

$$\mathcal{P}_n(A) \sim \mathrm{Po}[n|A|]; \tag{1.1}$$

$$\mathcal{P}_n(A_1), \ldots, \mathcal{P}_n(A_k) \text{ are independent for } A_1, \ldots, A_k \text{ disjoint} \tag{1.2}$$

where $\mathcal{X}(A)$ means the number of points of $\mathcal{X}$ in A.

Exercise 1.1. *(exPP)* Prove this.

The RGGs we consider are $G(\mathcal{X}_n, r_n)$ and $G(\mathcal{P}_n, r_n)$, with $(r_n)_{n\geq 1}$ a specified sequence of distance parameters.

One reason to study RGGs is to explore 'typical' properties of geometric graphs. Another reason is to assess statistical tests based on the graph $G(\mathcal{X}_n, r_n)$, for example tests for uniformity. It is of interest to compare this random graph model with others, such as the Erdős-Rényi random graph $G(n, p)$.

Notation. Many of the results described in this course are asymptotic results as $n \to \infty$. Unless stated otherwise, any limiting statement in the sequel is as $n \to \infty$. Also, for positive real-valued sequences a_n and b_n we use the following asymptotic notational conventions:

- $a_n = O(b_n)$ means $\limsup(a_n/b_n) < \infty$.
- $a_n = \Theta(b_n)$ means that both $a_n = O(b_n)$ and $b_n = O(a_n)$.
- $a_n = o(b_n)$ means that $a_n/b_n \to 0$. This may also be written as $a_n \ll b_n$ or as $b_n \gg a_n$, or as $b_n = \omega(a_n)$.
- $a_n \sim b_n$ means $a_n/b_n \to 1$.

We let θ denote the volume of the unit ball in $\mathbb{R}^d$.

Given two points $x, y \in \mathbb{R}^d$, we shall say x lies *to the left* of y if x precedes y in the lexicographic ordering on $\mathbb{R}^d$.

Exercise 1.2. *(Ex1)* Prove that if $r_n \to 0$, then $\mathbb{E}[\text{Degree}(\xi_1)] \sim \theta n r_n^d$.

If $r_n \to 0$, the expected number of edges incident to a 'typical vertex' of $\mathcal{X}_n$ or $\mathcal{P}_n$ goes like $\theta n r_n^d$ (in the case of $\mathcal{X}_n$, Exercise 1.2 makes this statement precise).

As this suggests, we often get different limiting behaviour depending on the limit behaviour of nr_n^d. We refer to cases with $nr_n^d \to 0$ as the *sparse limit*, and $nr_n^d \to \infty$ as the *dense limit*, and $nr_n^d = \Theta(1)$ as the *thermodynamic limit*.

2 Edge counts

Let $\mathcal{E}_n$ be the number of edges of $G(\mathcal{X}_n, r_n)$, and let $\mathcal{E}'_n$ be the number of edges of $G(\mathcal{P}_n, r_n)$. We consider the limiting behaviour of the probability distribution of $\mathcal{E}_n$ and $\mathcal{E}'_n$. More generally one could (by similar methods) also consider the limiting distribution of the number of subgraphs isomorphic to some specified connected finite graph; see [3] for details.

Note that if $r_n \to 0$ then

$$\mathbb{E}\mathcal{E}_n \sim \theta(n^2 r_n^d)/2. \qquad \underline{(0802c)} \qquad (2.1)$$

Exercise 2.1. *(exExp)* Prove this.

First we consider the sparse limit ($nr_n^d \to 0$). In this case we can show $\mathcal{E}_n$ is well approximated by a Poisson distributed variable (though if also $n^2 r_n^d$ is large this is itself well approximated by a normal random variable).

We prove this using the technique of *dependency graphs*. Suppose $(V, \sim)$ is a finite graph without loops (i.e. for $\alpha, \beta \in V$ we write $\alpha \sim \beta$ if α, β are adjacent.)

This is a *dependency graph* for a set of random variables $(W_\alpha, \alpha \in V)$ if whenever $A \subset V, B \subset V$ with $A \cap B = \emptyset$ and no edges connecting A to B,

$$(W_\alpha, \alpha \in A) \text{ is independent of } (W_\beta, \beta \in B).$$

Lemma 2.2. (Poisson Approximation Lemma) [1], or [3, Theorem 2.1]. *Suppose $(V, \sim)$ is a finite graph and $(W_\alpha)_{\alpha \in V}$ is a family of $0-1$ valued random variables, having $(V, \sim)$ as a dependency graph. For $\alpha, \beta \in V$ set $p_\alpha = \mathbb{P}[W_\alpha = 1]$ and $p_{\alpha\beta} = \mathbb{P}[W_\alpha = 1, W_\beta = 1]$. Then if we set $W = \sum_{\alpha \in V} W_\alpha$ and $\lambda = \sum_{\alpha \in V} p_\alpha$, we have*

$$\sum_{k=0}^{\infty} |\mathbb{P}[W = k] - e^{-\lambda}\lambda^k/k!| \leq \min(2/\lambda, 6) \times \left(\sum_{\alpha \in V} p_\alpha^2 + \sum_{\alpha} \sum_{\beta \sim \alpha} (p_{\alpha\beta} + p_\alpha p_\beta) \right). \tag{2.2}$$

Proof. Omitted.

Theorem 2.3. *(PoLimEd) Suppose $nr_n^d \to 0$. Let $\lambda_n = \mathbb{E}\mathcal{E}_n$. Then*

$$\lim_{n \to \infty} \left(\sum_{k=0}^{\infty} |\mathbb{P}[\mathcal{E}_n = k] - e^{-\lambda_n}\lambda_n^k/k!| \right) = 0.$$

Proof. Let $V = \{\alpha = \{i, j\} : 1 \leq i \leq j \leq n\}$ with $\alpha \sim \beta$ if $\alpha \cap \beta \neq \emptyset$ and $\alpha \neq \beta$. Set

$$W_{\{i,j\}} = \mathbf{1}\{\|\xi_i - \xi_j\| \leq r_n\}.$$

Then $\mathcal{E}_n = \sum_{\alpha \in V} W_\alpha$ and $(V, \sim)$ is a dependency graph for $\{W_\alpha\}$. Now p_α depends on n but not α and

$$p_\alpha \sim \theta r_n^d$$

and similarly for $\alpha \sim \beta$ we have

$$p_{\alpha\beta} \sim (\theta r_n^d)^2$$

so that

$$\lambda_n = \sum_{\alpha \in V} p_\alpha \sim \binom{n}{2} \theta r_n^d \sim \frac{n^2 \theta r_n^d}{2}$$

and

$$\sum_{\alpha \in V} p_\alpha^2 \sim \binom{n}{2} (\theta r_n^d)^2 \sim \lambda_n (\theta r_n^d)$$

while

$$\begin{aligned} \sum_{\alpha \in V} \sum_{\beta \sim \alpha} (p_{\alpha\beta} + p_\alpha p_\beta) &\sim \binom{n}{2} \times 2(n-1) \times \theta^2 r_n^{2d} \\ &= O(nr_n^d \lambda_n) \end{aligned}$$

and since we assume $nr_n^d \to 0$, Lemma 2.2 gives us the result. $\square$

The following lemma will be useful to us. It was called 'Palm theory for the Poisson process' in [3] but here we call it the 'Mecke formula'. In the proof we use notation

$$(n)_{(k)} := n(n-1) \cdots (n-k+1) \text{ for } n, k \in \mathbb{N}$$

(the so-called 'descending factorial'). Also, in the following formula (and elsewhere) the region of integration, when not specified otherwise, is to be taken to be $[0, 1]^d$.

Lemma 2.4. (Mecke formula.) *Let $k \in \mathbb{N}$. For any measurable real-valued function f, defined on the product of $(\mathbb{R}^d)^k$ and the space of finite subsets of $[0,1]^d$, for which the following expecation exists,*

$$\mathbb{E} \sum_{X_1,\ldots,X_k \in \mathcal{P}_n}^{\neq} f(X_1, X_2, \ldots, X_k, \mathcal{P}_n \setminus \{X_1, \ldots, X_k\}) = n^k \int dx_1 \cdots \int dx_k \mathbb{E} f(x_1, \ldots, x_k, \mathcal{P}_n)$$

where $\sum^{\neq}$ means the sum is over ordered k-tuples of distinct points of $\mathcal{P}_n$.

Proof. We condition on the number of points of $\mathcal{P}_n$. Then

$$\begin{aligned}
&\mathbb{E} \sum_{X_1,\ldots,X_k \in \mathcal{P}_n}^{\neq} f(X_1, X_2, \ldots, X_k, \mathcal{P}_n \setminus \{X_1, \ldots, X_k\}) \\
&= \sum_{m=k}^{\infty} \left(e^{-n} \frac{n^m}{m!} \right) (m)_k \int dx_1 \cdots \int dx_m f(x_1, \ldots, x_k, \{x_{k+1}, \ldots, x_m\}) \\
&= n^k \int dx_1 \cdots \int dx_k \sum_{m=k}^{\infty} \left(\frac{e^{-n} n^{m-k}}{(m-k)!} \right) \int dy_1 \cdots \int dy_{m-k} f(x_1, \ldots, x_k, \{y_1, \ldots, y_{m-k}\}) \\
&= n^k \int dx_1 \cdots \int dx_k \sum_{r=0}^{\infty} \left(\frac{e^{-n} n^r}{r!} \right) \int dy_1 \cdots \int dy_r f(x_1, \ldots, x_k, \{y_1, \ldots, x_r\}) \\
&= n^k \int dx_1 \cdots \int dx_k \mathbb{E} f(x_1, \ldots, x_k, \mathcal{P}_n)
\end{aligned}$$

where in the third line we made the substitution $y_j = x_{k+j}$ for $k < j \leq m$, and in the fourth line we set $r = m - k$. $\square$

3 Edge counts: Normal approximation

Given $x \in \mathbb{R}^d$ and $r > 0$, let $B(x; r)$ be the Euclidean ball of radius r centred at x.

We now give limit behaviour of the variance of $\mathcal{E}'_n$ in thermodynamic or dense limit. For a more general result (considering number of induced subgraphs isomorphic to a specified graph, rather than just number of edges), see [3, Proposition 3.7].

Proposition 3.1. *(propvar) If $\liminf(nr_n^d) > 0$ and $r_n \to 0$ then*

$$\mathrm{Var}(\mathcal{E}'_n) \sim n[(\theta n r_n^d)^2 + (1/2)\theta n r_n^d].$$

Note that in the dense limit this simplifies to $\mathrm{Var}(\mathcal{E}'_n) \sim n(\theta n r_n^d)^2$.

Exercise 3.2. *(exfac)* Prove that if X is a Poisson variable with parameter λ, and $k \in \mathbb{N}$, then $\mathbb{E}[(X)_k] = \lambda^k$.

Proof of Proposition 3.1 Let $g_n(x, y) = \mathbf{1}\{0 < \|x - y\| \leq r_n\}$ for $x, y \in \mathbb{R}^d$. Then by the Mecke formula

$$2\mathbb{E}\mathcal{E}'_n = \mathbb{E} \sum_{X,Y \in \mathcal{P}_n}^{\neq} g_n(X, Y) = n^2 \int \int g_n(x, y) dx dy \qquad \underline{(0814a)} \tag{3.1}$$

with all integrals over $[0,1]^d$ in this proof. Thus,

$$\mathbb{E}\mathcal{E}'_n \sim n^2\theta r_n/2. \qquad \underline{(0813a)} \tag{3.2}$$

Now consider $(\mathcal{E}'_n)^2$. This is the number of ordered pairs of edges in $G(\mathcal{P}_n, r_n)$. We may decompose this as

$$(\mathcal{E}'_n)^2 = S_{n,0} + S_{n,1} + S_{n,2}$$

where for $i = 0,1,2$ we let $S_{n,i}$ be the number of ordered pairs of edges with i endpoints in common. Then $S_{n,0} = \frac{1}{4}\sum^{\neq}_{U,V,X,Y\in\mathcal{P}_n} g_n(U,V)g_n(X,Y)$ where $\sum^{\neq}$ means the sum is over ordered 4-tuples of distinct points in $\mathcal{P}_n$. By the Mecke formula, followed by (3.1),

$$\mathbb{E}S_{n,0} = \frac{n^4}{4}\int\int\int\int g_n(u,v)g_n(x,y)dudvdxdy = (\mathbb{E}\mathcal{E}'_n)^2.$$

Also, $S_{n,2} = \mathcal{E}'_n$ and therefore

$$\mathrm{Var}(\mathcal{E}'_n) = \mathbb{E}[(\mathcal{E}'_n)^2] - (\mathbb{E}\mathcal{E}'_n)^2 = \mathbb{E}S_{n,1} + \mathbb{E}\mathcal{E}'_n \tag{3.3}$$

Next consider $S_{n,1}$. We have

$$S_{n,1} = \sum_{X\in\mathcal{P}_n}\sum^{\neq}_{Y,Z\in\mathcal{P}_n\setminus\{X\}} g_n(X,Y)g_n(X,Z) = \sum_{X\in\mathcal{P}_n} h_n(X,\mathcal{P}_n)$$

where $h_n(x;\mathcal{P}_n) := (\mathcal{P}_n(B(x;r_n)\setminus\{x\}))_{(2)}$ and $(n)_{(2)} := n(n-1)$ is the descending factorial. Using Exercise 3.2 , and the Mecke formula, we have

$$\mathbb{E}S_{n,1} = \mathbb{E}\sum_{X\in\mathcal{P}_n} h_n(X;\mathcal{P}_n) = n\int \mathbb{E}h_n(x;\mathcal{P}_n)dx \sim n(\theta r_n^d)^2. \tag{3.4}$$

By combining (3.3), (3.4) and (3.2) we get the result. $\square$

We shall prove a central limit theorem for $\mathbb{E}'_n$ in the thermodynamic or dense limit, using the following result of Chen and Shao [2] (alternatively we could use Theorem 2.4 of [3] which would give a slower rate of convergence in the CLT). We let Φ be the standard normal cumulative distribution fuction, i.e. $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^x e^{-t^2/2}dt$ for $x\in\mathbb{R}$.

Lemma 3.3. (Normal Approximation Lemma; see [2, Theorem 2.7].) *Let W_i, $i \in V$, be random variables indexed by the vertices of a dependency graph with $|V|$ vertices, all of degree at most D. Let $W = \sum_{i\in V} W_i$. Assume that $\mathbb{E}[W^2] = 1, \mathbb{E}[W_i] = 0$, and for some $\beta > 0$, that $\mathbb{E}|W_i|^3 \le \beta$ for all $i \in V$. Then*

$$\sup_t |P[W \le t] - \Phi(t)| \le 75D^{10}|V|\beta.$$

Proof. Omitted. Lemmas 3.3 and 2.2 are both proved by versions of *Stein's method*, which is an important topic in its own right, but beyond the scope of this short course.

Now we can give our central limit theorem.

Theorem 3.4. *Suppose $r_n \to 0$ but $\liminf(nr_n^d) > 0$. Then for all $x \in \mathbb{R}^d$,*

$$\mathbb{P}\left[\frac{\mathcal{E}'_n - \mathbb{E}\mathcal{E}'_n}{\sqrt{\mathrm{Var}(\mathcal{E}'_n)}} \le x\right] \to \Phi(x) \qquad \textit{as } n\to\infty.$$

Proof. Given n, partition $\mathbb{R}^d$ into cubes of side r_n, and let those cubes in the partition that have non-empty intersection with $[0,1]^d$ be denoted $C_1, \ldots, C_{k_n}$. Then $k_n \sim r_n^{-d}$. Let $M_i := M_i(n)$ denote the number of edges of $G(\mathcal{P}_n, r_n)$ with left-endpoint in C_i. Set

$$W_i := W_i(n) := \frac{M_i - \mathbb{E}M_i}{\sqrt{\mathrm{Var}(\mathcal{E}'_n)}}.$$

Then

$$\frac{\mathcal{E}'_n - \mathbb{E}\mathcal{E}'_n}{\sqrt{\mathrm{Var}(\mathcal{E}'_n)}} = \sum_{i=1}^{k_n} W_i.$$

Set $V = \{1, 2, \ldots, k_n\}$ and for $i, j \in V$ put $i \sim j$ if C_i and C_j are neighbouring cubes (i.e. they touch, so allowing diagonal neighbours) or they have a common neighbour. Then $(V, \sim)$ is a dependency graph for $(W_i, i \in V)$ and the maximal degree of this graph is at most $5^d - 1$, independent of n.

By Proposition 3.1 we have that $\mathrm{Var}(\mathcal{E}'_n) = \Theta(n(nr_n^d)^2)$. Therefore by Lemma 3.3 it suffices to prove that

$$\underline{(0813b)} k_n \max_{1 \le i \le k_n} \frac{\mathbb{E}[|M_i - \mathbb{E}M_i|^3]}{(n^{1/2} n r_n^d)^3} \to 0. \tag{3.5}$$

We shall estimate $\mathbb{E}[|M_i - \mathbb{E}M_i|^3]$ by first estimating $\mathbb{E}[|M_i - \mathbb{E}M_i|^4]$ (which is the same as $\mathbb{E}[(M_i - \mathbb{E}M_i)^4]$), and then using Hölder's inequality. By the Binomial Theorem,

$$\mathbb{E}[(M_i - \mathbb{E}M_i)^4] = \sum_{j=0}^{4} \binom{4}{j} (-\mathbb{E}M_i)^{4-j} \mathbb{E}[M_i^j]. \qquad \underline{(0813c)} \tag{3.6}$$

Note that

$$\mathbb{E}[M_i^4] = \mathbb{E} \sum_{e} \sum_{e'} \sum_{e''} \sum_{e'''} 1$$

where each sum runs through all edges of the $G(\mathcal{P}_n, r_n)$ having left-endpoint in the cube C_i. The leading order term in this expectation comes from when all of e, e', e'', e''' have distinct endpoints and using the Mecke formula, we have that the this leading order term is

$$\mathbb{E} \sum_{X_1, Y_1, \ldots, X_4, Y_4 \in \mathcal{P}_n}^{\neq} g_{i,n}(X_1, Y_1) \cdots g_{i,n}(X_4, Y_4)$$

where $g_{i,n}(x, y)$ is the indicator of the event that $\|x - y\| \le r_n$ and also $x \in C_i$ and also x lies to the left of y. Therefore by the Mecke formula, the leading term equals

$$\begin{aligned} n^8 \int \cdots \int dx_1 \cdots dx_4 dy_1 \cdots dy_4 \prod_{i=1}^{4} g_{i,n}(x_i, y_i) \\ = \left(n^2 \int \int dx dy g_{i,n}(x, y) \right)^4 \\ = (\mathbb{E}[M_i])^4 \end{aligned}$$

(recall that all integrals are over $[0,1]^d$ unless specified otherwise).

Similarly, the leading-order term in $\mathbb{E}M_i^3$ (coming from triples of edges with no endpoints in common) is equal to $(\mathbb{E}[M_i])^3$, and the leading-order term in $\mathbb{E}M_i^2$ (coming from pairs of edges with no endpoints in common) is equal to $(\mathbb{E}[M_i])^2$. Therefore if we collect together all the leading-order terms in (3.6) we get

$$\sum_{j=1}^{4} \binom{4}{j} (-\mathbb{E}M_i)^{4-j} (\mathbb{E}M_i)^j$$

which is equal to $(\mathbb{E}M_i - \mathbb{E}M_i)^4$ and therefore comes to zero.

Next we consider second order terms. Let R_i be the number of ordered pairs of edges having a common endpoint and both with left endpoint in C_i. Then

$$\begin{aligned}
\mathbb{E}R_i &= \mathbb{E} \sum_{X,Y,Z \in \mathcal{P}_n}^{\neq} (g_{i,n}(X,Y) + g_{i,n}(Y,X))(g_{i,n}(X,Z) + g_{i,n}(Z,X)) \\
&= \int\int\int (g_{i,n}(x,y) + g_{i,n}(y,x))(g_{i,n}(x,z) + g_{i,n}(z,x)) dxdydz
\end{aligned}$$

where the last line comes from the Mecke formula.

The second order term in $\mathbb{E}M_i^4$ comes from quadruples of edges e, e', e'', e''' having seven distinct endpoints between them (so having precisely one common endpoint among the four edges). There are 6 ways to choose which two of the edges e, e', e'', e''' share an endpoint, and therefore the second-order term in $\mathbb{E}M_i^4$ comes to

$$6 \sum_{X,Y,Z,X_1,X_2,Y_1,Y_2 \in \mathcal{P}_n}^{\neq} (g_{i,n}(X,Y) + g_{i,n}(Y,X))(g_{i,n}(X,Z) + g_{i,n}(Z,X)) g_{i,n}(X_1,Y_1) g_{i,n}(X_2,Y_2)$$

and by the Mecke formula this comes to

$$\begin{aligned}
6n^7 \int \cdots \int dxdydzdx_1dy_1dx_2dy_2 (g_{i,n}(x,y) + g_{i,n}(y,x))(g_{i,n}(x,z) + g_{i,n}(z,x)) \\
\times g_{i,n}(x_1,y_1) g_{i,n}(x_2,y_2) \\
= 6(\mathbb{E}M_i)^2 \mathbb{E}R_i
\end{aligned}$$

Similarly, the second order term in $\mathbb{E}M_i^3$ comes to $3\mathbb{E}M_i\mathbb{E}R_i$, and the second order term in $\mathbb{E}M_i^2$ comes to $\mathbb{E}R_i$. There is no second order term in $\mathbb{E}M_i$.

Therefore, combining all second-order terms in (3.6) we get

$$(\mathbb{E}M_i)^2 \mathbb{E}R_i \left(\binom{4}{4} \times 6 - \binom{4}{3} \times 3 + \binom{4}{2} \times 1 \right) = 0.$$

Therefore the leading order non-zero term in (3.6) comes from the third-order terms. The third-order term in $\mathbb{E}M_i^4$ comes from 4-tuples of edges e, e', e'', e''' having two shared endpoints between them (so with a total of six distinct endpoints). This is bounded by some combinatorial constant, times

$$\sum_{X_1,\ldots,X_6 \in \mathcal{P}_n}^{\neq} h_{i,n}^*(X_1, \ldots, X_n)$$

where $h^*_{i,n}(x_1,\ldots,x_6)$ is the indicator of the event that $x_1,\ldots,x_6$ all lie in C_i or in one of the neighbouring cubes. Therefore by the Mecke formula, the third-order term in $\mathbb{E}M_i^4$ is bounded by a constant times

$$n^6 \int \cdots \int h^*_{i,n}(x_1,\ldots,x_6) dx_1 \cdots dx_6$$

and thus is $O((nr_n^d)^6)$. Similarly the higher order terms are $O((nr_n^d)^5)$. Combining all this, we find from (3.6) that

$$\mathbb{E}[(M_i - \mathbb{E}M_i)^4] = O((nr_n^d)^6)$$

and also by Hölder's inequality for any random variable X we have $\mathbb{E}[|X|^3] \leq (\mathbb{E}X^4)^{3/4}$, so that

$$\mathbb{E}[|M_i - \mathbb{E}M_i|^3] = O((nr_n^d)^{9/2}),$$

uniformly over $i \leq k_n$. Therefore

$$k_n \max_{1\leq i\leq k_n} \frac{\mathbb{E}[|M_i - \mathbb{E}M_i|^3]}{(n^{1/2}nr_n^d)^3} = O\left(r_n^{-d}(nr_n^d)^{3/2}n^{-3/2}\right) = O(r_n^{d/2})$$

which tends to zero so we have (3.5) as required. □

4 The maximum degree

Let Δ_n be the maximum degree of vertices in $G(\mathcal{X}_n, r_n)$, and let Δ'_n be the maximum degree in $G(\mathcal{P}_n, r_n)$. Given $k \in \mathbb{N}$ let $N_{\geq k}(n)$ (respectively $N'_{\geq k}(n)$) be the number of vertices of $G(\mathcal{X}_n, r_n)$ (respectively $G(\mathcal{P}_n, r_n)$) of degree at least k. Also set $N_k(n) := N_{\geq k}(n) - N_{\geq k+1}(n)$ and $N'_k(n) := N'_{\geq k}(n) - N'_{\geq k+1}(n)$ (the number of vertices of degree exactly k).

First consider the sparse limit with $nr_n^d \to 0$. As long as this convergence to zero is not very slow, we can show that the maximum degree remains bounded in probability.

Indeed, suppose for some $k \in \mathbb{N}$ that in fact $nr_n^d = o(n^{-1/k})$.

Then $n^{(k+1)/k}r_n^d \to 0$, and $n^{k+1}(r_n^d)^k \to 0$. For $\lambda > 0$, let $\mathrm{Po}(\lambda)$ denote a Poisson distributed random variable with parameter λ. Note that $\mathbb{P}[\mathrm{Po}(\lambda) \geq k] \sim \lambda^k/k!$ as $\lambda \downarrow 0$. Therefore using the Mecke formula we have

$$\mathbb{E}N'_{\geq k}(n) \sim n(\theta nr_n^d)^k/k!$$

which tends to zero, so by Markov's inequality $\mathbb{P}[N'_{\geq k}(n) > 0] \to 0$ and hence $\mathbb{P}[\Delta'_n \geq k] \to 0$.

Similarly, if $nr_n^d = \omega(n^{-1/k})$ then $\mathbb{E}N'_{\geq k}(n) \to \infty$, and in fact it can also be shown in this case that $\mathbb{P}[\Delta'_n \geq k] \to 1$. Therefore if $n^{-1/k} \ll nr_n^d \ll n^{-1/(k+1)}$ then $\mathbb{P}[\Delta'_n = k] \to 1$.

If $nr_n^d \sim \alpha n^{-1/k}$ for some positive finite constant α then $\lim_{n\to\infty} \mathbb{P}[\Delta_n = k]$ exists and lies in $(0,1)$, and

$$\lim_{n\to\infty} \mathbb{P}[\Delta_n \in \{k-1, k\}] = 1.$$

This last fact is the so-called *two-point concentration* (or *focusing*) property of the distribution of Δ_n.

We prove the above assertions only for $k = 1$. In this case if $nr_n^d \sim \alpha n^{-1}$, then $n^2 r_n^d \to \alpha$, and $\mathbb{E}\mathcal{E}_n \to \theta\alpha/2$ by (2.1). Then by Theorem 2.3 we have

$$\mathbb{P}[\Delta_n = 0] = \mathbb{P}[\mathcal{E}_n = 0] \to \exp(-\theta\alpha/2)$$

but also since $nr_n^d = o(n^{-1/2})$ we have $\mathbb{P}[\Delta'_n \geq 2] \to 0$.

Exercise 4.1. *Using the preceding statement and monotonicity, prove that if $n^2 r_n^d \to \infty$ then $\mathbb{P}[\Delta_n = 0] \to 0$.*

Next we briefly consider the thermodynamic limit with $nr_n^d \to \alpha$ for some $\alpha \in (0, \infty)$. For any k,

$$\mathbb{E}N'_k \sim n \exp(-\theta\alpha)(\theta\alpha)^k / k!$$

which tends to ∞ as $k \to \infty$.

Exercise 4.2. *Prove that $\mathbb{P}[N_k \geq 1] \to 1$, i.e. $\mathbb{P}[\Delta'_n \geq k] \to 1$, in the case where* $\liminf nr_n^d > 0$.

Thus in this case, and in the previous case, $\Delta'_n \gg nr_n^d$ in probability.

Now we consider the case with

$$\frac{nr_n^d}{\log n} \to \alpha, \qquad \alpha \in (0, \infty) \qquad \underline{(connlim)} \tag{4.1}$$

In this case we shall find it is *not* the case that $\Delta_n \gg nr_n^d$ in probability. We shall give a strong law showing $\Delta_n/(nr_n^d)$ tends to a positive constant almost surely.

To state the result we shall need more notation. Define the function $H : (0, \infty) \to \mathbb{R}$ by

$$H(a) = 1 - a + a \log a.$$

Some simple calculus shows that $H(1) = 0$ is the unique minimum value of $H(\cdot)$ with $H(\cdot)$ increasing on $(0, \infty)$ and decreasing on $(0, 1)$; also $\lim_{a \downarrow 0} H(a) = 1$ and $\lim_{a \to \infty} H(a) = +\infty$.

For $x > 0$ let $h_+^{-1}(x)$ be the $a \in (1, \infty)$ with $H(a) = x$, and if $0 < x < 1$ let let $h_+^{-1}(x)$ be the $a \in (0, 1)$ with $H(a) = x$.

Theorem 4.3. *<u>(Thmax)</u> Suppose (4.1) holds. Then $\Delta_n/(n\theta r_n^d) \to H_+^{-1}(1/\alpha)$ almost surely.*

Remark. It can be shown that for the random geometric graph in the *torus*, if (4.1) holds with $\alpha > 1$ and δ_n denotes the *minimum* degree of $G(\mathcal{X}_n, r_n)$ then $\delta_n/(n\theta r_n^d) \to H_-^{-1}(1/\alpha)$ almost surely.

The function H arises from the following large-deviations results concerning the Poisson distribution.

Lemma 4.4. *<u>(LDlem)</u> It is the case that if $np = \mu$ then*

$$\mathbb{P}[\mathrm{Po}(\lambda) \geq k] \leq \exp(-\lambda H(k/\lambda)), \quad k \geq \lambda \qquad \underline{(LD3)} \tag{4.2}$$

$$\mathbb{P}[\mathrm{Po}(\lambda) \leq k] \leq \exp(-\lambda H(k/\lambda)), \quad k \leq \lambda \qquad \underline{(LD4)} \tag{4.3}$$

and for any $a > 1$ and $\varepsilon > 0$, there exists $\lambda_0 \in (0, \infty)$ such

$$\mathbb{P}[\mathrm{Po}(\lambda) \geq a\lambda] \geq \exp(-(1 + \varepsilon)\lambda H(a)), \quad \lambda \geq \lambda_0. \qquad \underline{(LD5)} \tag{4.4}$$

Proof. Set $X = \mathrm{Po}(\lambda)$. For $z \geq 1$, by Markov's inequality applied to the random variable z^X,

$$\mathbb{P}[\mathcal{P}(\lambda) \geq k] \leq z^{-k}\mathbb{E}[z^X] = z^{-k}e^{\lambda(z-1)}. \qquad \underline{(0814b)} \tag{4.5}$$

and similarly for $z \leq 1$,

$$\mathbb{P}[\mathcal{P}(\lambda) \leq k] \leq z^{-k}\mathbb{E}[z^X] = z^{-k}e^{\lambda(z-1)}. \qquad \underline{(0815a)} \tag{4.6}$$

Put $z = k/\lambda$. If $k \geq \lambda$ then this choice of z satisfies $z \geq 1$ so by (4.5) we obtain

$$\mathbb{P}[\mathcal{P}(\lambda) \geq k] \leq \left(\frac{\lambda}{k}\right)^k e^{k-\lambda} = \exp(-\lambda H(k/\lambda))$$

which proves (4.2). If $k \leq \lambda$ then the same choice of z satisfies $z \leq 1$ so we obtain (4.3) from (4.6).

Finally to prove (4.4), we use the following inequality (a weak form of Stirling's formula):

$$\log k! = \sum_{i=1}^{k} \log i \leq \int_1^{k+1} \log x = (k+1)\log(k+1) - k$$

so $k! \leq (k+1)^{k+1}e^{-k}$. Thus if we fix $a > 1$ and put $k = \lceil a\lambda \rceil$ we obtain

$$\begin{aligned}\mathbb{P}[\mathrm{Po}(\lambda) \geq a\lambda] \geq \mathbb{P}[\mathrm{Po}(\lambda) = k] &= e^{-\lambda}\lambda^k/k! \\ &\geq \frac{e^{-\lambda}e^k\lambda^k}{(k+1)^{k+1}}\end{aligned}$$

and hence

$$\begin{aligned}\lambda^{-1}\log \mathbb{P}[\mathrm{Po}(\lambda) \geq a\lambda] &\geq -1 + (k/\lambda) - (k/\lambda)\log((k+1)/\lambda) - \lambda^{-1}\log(k+1) \\ &\geq -H(a)(1+\varepsilon), \qquad \lambda \text{ large.}\end{aligned}$$

which proves (4.4) $\square$

Proof of Theorem 4.3. Assume $n\theta r_n^d/\log n \to \alpha \in (0,\infty)$.

Let $\beta < H_+^{-1}(1/\alpha)$, so that $H(\beta) > 1/\alpha$. Let $\delta > 0$ and let $\varepsilon > 0$ (to be chosen later).

Cover $[0,1]^d$ by balls of radius εr_n. The number of balls required, denoted k_n, is $O(r_n^{-d})$ and therefore is $O(n/\log n)$.

Let the centres of these balls be denoted $x_1, \ldots, x_{k_n}$.

If $\Delta_n \geq n\theta r_n^d\beta$, then there is a point X of $\mathcal{P}_n$ with degree at least $n\theta r_n^d\beta$ in $G(\mathcal{P}_n, r_n)$, and this must lie in one of the balls $B(x_i, \varepsilon r_n)$, $1 \leq i \leq k_n$, say for $i = I$. Then by the triangle inequality, we must have

$$\mathcal{P}_n(B(x_I,(1+\varepsilon)r_n)) \geq n\theta r_n^d\beta.$$

Therefore by the union bound,

$$\mathbb{P}[\Delta_n \geq n\theta r_n^d\beta] \leq \sum_{i=1}^{k_n} \mathbb{P}[\mathcal{P}_n(B(x_i,(1+\varepsilon)r_n)) \geq n\theta r_n^d\beta]. \qquad \underline{(0815b)} \tag{4.7}$$

Now for each $i \le k_n$, the random variable $W_i := \mathcal{P}_n(B(x_i, (1+\varepsilon)r_n))$ is Poisson distributed with mean satisfying

$$\mathbb{E}W_i \le n\theta(1+\varepsilon)^d r_n^d \le (1+\varepsilon)^{d+1}\alpha \log n$$

so that by Lemma 4.4,

$$\begin{aligned}\mathbb{P}[W_i > n\theta r_n^d \beta] &\le \exp\left[-(1+\varepsilon)^{d+1}\alpha(\log n)H\left(\frac{n\theta r_n^d\beta}{n\theta(1+\varepsilon)^d r_n^d}\right)\right] \\ &= \exp\left[(1+\varepsilon)^{d+1}\alpha(\log n)H\left(\frac{\beta}{(1+\varepsilon)^d}\right)\right]\end{aligned}$$

and if we choose ε small enough so that

$$(1+\varepsilon)^{d+1}\alpha H(\beta/(1+\varepsilon)^d) > 1+\delta$$

then we have

$$\mathbb{P}[W_i > n\theta r_n^d\beta] \le n^{-(1+\delta)}.$$

Therefore

$$\sum_{i=1}^{k_n} \mathbb{P}\left[W_i \ge n\theta(1+\varepsilon)^d r_n^d\right] = O((n/\log n)n^{-(1+\delta)})$$

which is summable in n, and therefore by (4.7) we have for any $\beta > H_+^{-1}(1/\alpha)$.

$$\sum_{n=1}^{\infty} \mathbb{P}[\Delta_n \ge n\theta r_n^d\beta] < \infty,$$

and therefore by the first Borel-Cantelli lemma

$$\mathbb{P}[\limsup_{n\to\infty} \Delta_n/(n\theta r_n^d) > \beta] = 0, \qquad \beta > H_+^{-1}(1/\alpha). \qquad \underline{(0815c)} \tag{4.8}$$

Now suppose $\beta < H_+^{-1}(1/\alpha)$, so that $H(\beta) < 1/\alpha$. Let $\varepsilon > 0$ and choose $\delta > 0$, to be chosen below. Choose a maximal collection of points $x_1, \ldots, x_{j_n}$ such that the balls $B(x_i, (1-\delta)r_n), 1 \le i \le j_n$ are disjoint and all contained in $[0,1]^d$. Then $j_n = \Theta(r_n^{-d}) = \theta(n/\log n)$. For $1 \le i \le j_n$, define the event

$$A_i = \{\mathcal{P}_n(B(x_i, \delta r_n)) \ge 1\} \cap \{\mathcal{P}_n(B(x_i,(1-\delta)r_n) \setminus B(x_i, \delta r_n)) \ge n\theta r_n^d\beta\}.$$

Then we have the event inclusion

$$\cup_{i=1}^{j_n} A_i \subset \{\Delta_n' \ge n\theta r_n^d \beta\} \qquad \underline{(0815d)} \tag{4.9}$$

Now, $V_i := \mathcal{P}_n(B(x_i,(1-\delta)r_n) \setminus B(x_i,\delta r_n))$ is Poisson with mean

$$\mathbb{E}V_i = n\theta((1-\delta)^d - \delta^d)r_n^d > (1-\delta)^{d+1}\alpha\log n$$

where the last inequality holds for all large enough n. So

$$\mathbb{P}[V_i \ge n\theta r_n^d\beta] \ge \exp\left[-\alpha(\log n)H\left(\frac{\beta}{(1-\delta)^d}\right)\right]$$

and if δ is chosen so small that $\alpha H\left(\frac{\beta}{(1-\delta)^d}\right) > (1-\varepsilon)$, we obtain that

$$\mathbb{P}[V_i \geq n\theta r_n^d \beta] \geq \exp(-(1-\delta)\log n) = n^{\delta-1}.$$

Since the event $\{\mathcal{P}_n(B(x_i, \delta r_n)) \geq 1\}$ has probability greater than $1/2$ (for large enough n) and is independent of the event $\{V_i \geq n\theta r_n^d \beta\}$, we have $\mathbb{P}[A_i] \geq (1/2)n^{\delta-1}$. Therefore, since the events $A_1, \ldots, A_{j_n}$ are independent we have for some $c > 0$ that

$$\mathbb{P}[\cap_{i=1}^{j_n} A_i^c] \leq (1 - \frac{1}{2}n^{1-\delta})^{j_n} \leq \exp(-cn^{\delta-1} \times (n/\log n))$$

which tends to zero. Combined with (4.9), this shows that

$$\sum_{n=1}^{\infty} \mathbb{P}[\Delta_n' < \beta n \theta r_n^d] \leq \sum_{n=1}^{\infty} \left[\cup_{i=1}^{j_n} A_i\right] \to 1,$$

and hence by the Borel-Cantelli lemma,

$$\mathbb{P}[\liminf_{n\to\infty} \Delta_n'/(\theta r_n^d) < \beta] = 0, \qquad \beta < H_+^{-1}(1/\alpha).$$

Combined with (4.8) this gives us the result. $\square$

5 Connectivity

Let $\mathcal{K}$ be the class of connected graphs, and let

$$\rho_n' = \min\{r : G(\mathcal{P}_n, r) \in \mathcal{K}\}$$

which is a random variable determined by the configuration of $\mathcal{P}_n$. It is called the *connectivity threshold*. Similarly define

$$\rho_n = \min\{r : G(\mathcal{P}_n, r) \in \mathcal{K}\}.$$

In this section we prove the following result. Recall that for random variables X_n and any constant c, we say $X_n \xrightarrow{P} c$ if for all $\varepsilon > 0$ we have $\mathbb{P}[|X_n - c| > \varepsilon] \to 0$.

Theorem 5.1. *(Thconn) Assume $d = 2$. Then*

$$n\theta(\rho_n')^2/\log n \xrightarrow{P} 1. \qquad (0815e) \tag{5.1}$$

Remarks.

(i) The retriction to $d = 2$ arises because boundary effects become more important in higher dimensions (and $d = 1$ is different because 1-space is 'less connected'.).

(ii) The convergence (5.1) actually holds with almost sure convergence, and these hold with ρ_n' replaced by ρ_n, but proving these extensions is beyond the scope of these lectures.

(iii) A further extension of (5.1) is the following convergence in distribution result: for any $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}[n\theta(\rho_n')^2 - \log n \le t] \to \exp(-e^{-t}).$$

Proving this is also beyond our scope here.

Recall that δ_n' denotes the minimum degree of $G(\mathcal{P}_n)$. If $\delta_n' = 0$ then clearly $G(\mathcal{P}_n) \notin \mathcal{K}'$.

Theorem 5.2. *(Thmin) If $n\theta r_n^d/\log n \to \alpha < 1$ then $\mathbb{P}[\delta_n' > 0] \to 1$ and*

Proof. Let $\varepsilon > 0$ be chosen so that $\alpha(1+\varepsilon)^d < 1-\varepsilon$. Choose a maximal collection of points $x_i \in [0,1]^2, 1 \le i \le k_n$ such that the balls $B_i^+ := B(x_i, r_n(1+\varepsilon)), 1 \le i \le k_n$ are disjoint and contained in $[0,1]^d$. Note that $r_n^d = \Theta((\log n)/n)$ and $k_n = \Theta(n/\log n)$.

Then let $B_i^- := B(x_i, r_n\varepsilon)$, for $1 \le i \le k_n$, and define the events

$$E_i := \{\mathcal{P}_n(B_i^+) = \mathcal{P}(B_i^-) = 1\}, \quad 1 \le i \le k_n.$$

Then, writing $|\cdot|$ for Lebesgue measure, we have

$$\begin{aligned}\mathbb{P}[E_i] &= e^{-n|B_i^-|}(n|B_i|) \times e^{-n|B_i^+ \setminus B_i^-|} \\ &= \exp(-n\theta(r_n(1+\varepsilon))^d) \times n\theta(r_n\varepsilon)^d \\ &\ge \exp(-(1-\varepsilon)\log n) \times \Theta(\log n) = \Theta(n^{\varepsilon-1}\log n)\end{aligned}$$

and so there is a constant such that

$$\begin{aligned}\mathbb{P}[\cap_{i=1}^{r_n} E_i^c] \le (1 - cn^{\varepsilon-1}\log n)^{k_n} &\le \exp(-k_n c n^{\varepsilon-1}\log n) \\ &\le \exp(-\Theta(n^\varepsilon)) \to 0\end{aligned}$$

and therefore

$$\mathbb{P}[\delta_n' > 0] \ge \mathbb{P}[\cup_{i=1}^{k_n} E_i] \to 1$$

as claimed □.

Corollary 5.3. *Given $\varepsilon > 0$ we have $\mathbb{P}[n\theta(\rho_n')^d/\log n > 1-\varepsilon] \to 1$.*

This follows from Theorem 5.2 and the fact that if the minimum degree of a graph is zero, then it is not connected.

$\mathbb{P}[G(\mathcal{P}_n, r_n) \in \mathcal{K}] \to 0$.

For the rest of this section we assume $d = 2$.

To complete the proof of Theorem 5.1, it suffices to prove the following:

Theorem 5.4. *Suppose $n\theta r_n^d/\log n \to \alpha > 1$. Then $\mathbb{P}[G(\mathcal{P}_n, r_n \in \mathcal{K}_n) \to 1]$.*

The proof of this is long, and requires a series of lemmas. It proceeds by discretization of space.

Let $\varepsilon \in (0, 1/2)$, to be chosen below. Assume $d = 2$ and r_n is as above. Divide $[0,1]^2$ into squares of side εr_n (we shall ignore the fact that $1/\varepsilon r_n$ is not an integer).

Let $\mathcal{L}_n$ be the set of centres of these square (a finite lattice). then $|\mathcal{L}_n| = \Theta(n/\log n)$.

List the squares as $Q_i, 1 \le i \le |\mathcal{L}_n|$, and the corresponding centres of squares (i.e., the elements of $\mathcal{L}_n$) as $q_i, 1 \le i \le |\mathcal{L}_n|$.

Let us say $q_i \in \mathcal{L}_n$ is *occupied* if $\mathcal{P}_n(Q_i) > 0$. Let $\mathcal{O}_n$ be the (random) set of sites $q_i \in \mathcal{L}_n$ that are occupied.

Lemma 5.5. (conlem1) *If $G(\mathcal{P}_n, r_n)$ is disconnected, then so is $G(\mathcal{O}_n, r_n(1-2\varepsilon))$.*

Proof. If $q_i, q_j \in \mathcal{L}_n$ with $\|q_i - q_j\| \leq r_n(1-2\varepsilon)$, then for any $X \in \mathcal{P}_n \cap Q_i$ and $Y \in \mathcal{P}_n \cap Q_j$, by the triangle inequality we have

$$\|X - Y\| \leq \|X - q_i\| + \|q_i - q_j\| + \|Y - q_j\| \leq r_n\varepsilon + r_n(1-2\varepsilon) + r_n\varepsilon = r_n$$

and therefore if $G(\mathcal{O}_n, r_n(1-2\varepsilon))$ is connected, so is $G(\mathcal{P}_n, r_n)$. □

Let $\mathcal{A}$ denote the set of $\sigma \subset \mathcal{L}_n$ with m elements such that $G(\sigma, r_n(1-\varepsilon))$ is connected (sometimes called 'lattice animals').

Let $\mathcal{A}^2_{n,m}$ be the set of $\sigma \in \mathcal{A}_{n,m}$ such that $\mathrm{dist}(\sigma, \partial[0,1]^2) > 2r_n$, i.e. all elements of σ are distant at least $2r_n$ from the boundary of $[0,1]^2$.

Let $\mathcal{A}^1_{n,m}$ be the set of $\sigma \in \mathcal{A}_{n,m}$ such that σ is distant less than $2r_n$ from *just one edge* of $[0,1]^2$.

Let $\mathcal{A}^0_{n,m} := \mathcal{A}^0_{n,m} \setminus (\mathcal{A}^2_{n,m} \cup \mathcal{A}^1_{n,m})$, the set of $\sigma \in \mathcal{A}_{n,m}$ such that σ is distant less than $2r_n$ from *two edges* of $[0,1]^2$ (i.e. near a corner of $[0,1]^2$).

Lemma 5.6. (countlem) *Given $m \in \mathbb{N}$, there is constant $C = C(m)$ such that*

$$|\mathcal{A}_{n,m} \leq C(n/\log n), \qquad |\mathcal{A}^1_{n,m} \leq C(n/\log n)^{1/2}, \qquad |\mathcal{A}^0_{n,m} \leq C$$

for all n.

Proof. Fix m. Consider how many ways there are to choose $\sigma \in \mathcal{A}_{n,m}$.

There are at most r_n^{-2} choices, and hence $O(n/\log n)$ choices, for the first element of σ in the lexicographic ordering. Having chosen the first element of σ, there are a bounded number of ways to choose the rest of σ.

Consider how many ways there are to choose $\sigma \in \mathcal{A}^1_{n,m}$. In this case there are $O(r_n^{-1}) = O((n/\log n)^{1/2})$ ways to choose the first element of σ, and then a bounded number of ways to choose the rest of σ.

Finally consider how many ways there are to choose $\sigma \in \mathcal{A}^0_{n,m}$. In this case there are $O(1)$ ways to choose the first element of σ, and then a bounded number of ways to choose the rest of σ. □

References

[1] R. Arratia, L. Goldstein, and L. Gordon (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **17**, 9-25.

[2] L. Chen and Q.-M. Shao (2004), Normal approximation under local dependence, *Ann. Probab.* **32**, 1985-2028

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