The Annals of Applied Probability
2016, Vol. 26, No. 5, 3078–3109
DOI: 10.1214/15-AAP1170
© Institute of Mathematical Statistics, 2016

ALMOST OPTIMAL SPARSIFICATION OF RANDOM GEOMETRIC GRAPHS

BY NICOLAS BROUTIN*,[1], LUC DEVROYE†,[2] AND GÁBOR LUGOSI‡,[3]

Inria, McGill University†, ICREA‡ and Pompeu Fabra University‡*

A random geometric irrigation graph $\Gamma_n(r_n, \xi)$ has n vertices identified by n independent uniformly distributed points $X_1, \dots, X_n$ in the unit square $[0, 1]^2$. Each point X_i selects ξ_i neighbors at random, without replacement, among those points X_j $(j \neq i)$ for which $\|X_i - X_j\| < r_n$, and the selected vertices are connected to X_i by an edge. The number ξ_i of the neighbors is an integer-valued random variable, chosen independently with identical distribution for each X_i such that ξ_i satisfies $\xi_i \geq 1$. We prove that when $r_n = \gamma_n \sqrt{\log n / n}$ for $\gamma_n \to \infty$ with $\gamma_n = o(n^{1/6} / \log^{5/6} n)$, the random geometric irrigation graph experiences *explosive percolation* in the sense that if $\mathbf{E}\xi_i = 1$, then the largest connected component has $o(n)$ vertices but if $\mathbf{E}\xi_i > 1$, then the number of vertices of the largest connected component is, with high probability, $n - o(n)$. This offers a natural noncentralized sparsification of a random geometric graph that is mostly connected.

1. Introduction.

We study the following model of random geometric "irrigation" graphs. Let $\mathbf{X} = \{X_1, \dots, X_n\}$ be a set of uniformly distributed random points in $[0, 1]^2$. Given a positive number $r_n > 0$, we may define the random geometric graph $G_n(r_n)$ with vertex set $[n] := \{1, \dots, n\}$ in which vertex i and vertex j are connected if and only if the distance of X_i and X_j does not exceed the threshold r_n [18, 25]. To avoid technicalities arising from irregularities around the borders of the unit square, we consider $[0, 1]^2$ as a torus. Formally, we measure distance of $x = (x_1, x_2)$, $y = (y_1, y_2) \in [0, 1]^2$ by

$$d(x, y) = \left(\sum_{i=1}^{2} \min\bigl(|x_i - y_i|, 1 - |x_i - y_i|\bigr)^2 \right)^{1/2}.$$

It is well known that the connectivity threshold for the graph $G_n(r_n)$ is $r_n^\star = \sqrt{\log n / (n\pi)}$ (see, e.g., Penrose [25]). This means that, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\bigl(G_n(r_n) \text{ is connected}\bigr) = \begin{cases} 0 & \text{if } r_n \leq (1 - \epsilon) r_n^\star, \\ 1 & \text{if } r_n \geq (1 + \epsilon) r_n^\star. \end{cases}$$

Received November 2014; revised September 2015.

[1] Supported by Inria EA RNA EQA-021303.

[2] Supported by NSERC.

[3] Supported by the Spanish Ministry of Science and Technology Grant MTM2012-37195.

MSC2010 subject classifications. 60C05, 05C80.

Key words and phrases. Random geometric graph, connectivity, irrigation graph.

In this paper, we consider values of r_n well above the connectivity threshold. So $G(r_n)$ is connected with high probability. In some applications, it is desirable to obtain, in a decentralized manner, a graph that is connected but sparse. To this aim, one may sparsify the graph in a distributed way by selecting, randomly and independently for each vertex u, a subset of the edges adjacent to u, and then consider the subgraph containing those edges only. Such random subgraphs are sometimes called *irrigation graphs* or *Bluetooth networks* [9, 12, 13, 16, 28]. A related model of *soft random geometric graphs*, resulting from bond percolation on the geometric graph $G_n(r_n)$ is studied by Penrose [27]. In this paper, we study the following slight generalization of the irrigation graph model.

The irrigation graph. We consider a positive integer-valued random variable ξ. We assume that $\xi \geq 1$ with probability one. The *random irrigation graph* $\Gamma_n = \Gamma_n(r_n, \xi)$ is obtained as a random subgraph of $G_n(r_n)$ as follows. For every $x \in [0,1]^2$, define $\rho(x) = |B(x, r_n) \cap \mathbf{X}|$ to be the number of points of $\mathbf{X}$ that are visible from x, where $B(x,r) = \{y \in [0,1]^2 : d(x,y) \leq r\}$. With every point $X_u \in \mathbf{X}$, we associate ξ_u, an independent copy of the random variable ξ. Then given that $X_u \in \mathbf{X}$ and ξ_u, let $\mathbf{Y}(X_u) := (Y_i(X_u), 1 \leq i \leq \xi_u \wedge \rho(X_u))$ be a subset of elements of $\mathbf{X} \cap B(X_u, r_n)$ chosen uniformly at random, without replacement. [Note that this definition allows a vertex to select itself. Such a selection does not create any edge. In a slight modification of the model, the selection is from the set $\mathbf{X} \cap B(X_u, r_n) \setminus \{X_u\}$. Since self-selection is unlikely, all our asymptotic results remain unchanged in the modified model.]

We then define $\Gamma_n^+ = \Gamma_n^+(r_n, \xi)$ as the digraph on $[n]$ in which two vertices $u, v \in [n]$ are connected by an oriented edge (u,v) if $X_v = Y_i(X_u)$ for some $1 \leq i \leq \xi_u \wedge \rho(X_u)$. Finally, we define $\Gamma_n = \Gamma_n(r_n, \xi)$ as the graph on $[n]$ in which $\{u,v\}$ is an edge if either (u,v) or (v,u) is an oriented edge of Γ_n^+.

We study the size of the largest connected component of the random graph $\Gamma_n(r_n, \xi)$ for large values of n. In the entire paper, the size always refers to the number of vertices. We say that a property of the graph holds *with high probability* (w.h.p.) when the probability that the property does not hold is bounded by a function of n that goes to zero as $n \to \infty$.

Connectivity of random geometric irrigation graphs. Irrigation subgraphs of random geometric graphs have some desirable connectivity properties. In particular, the graph remains connected with a significant reduction of the number of edges when compared to the underlying random geometric graph. Connectivity properties of $\Gamma_n(r_n, c_n)$ (i.e., when $\xi = c_n$ is deterministic, possibly depending on n) have been studied by various authors. Dubhashi et al. [13] showed that when $r_n = r > 0$ is independent of n, $\Gamma_n(r, 2)$ is connected with high probability. In this setting where r is bounded away from zero, the underlying random geometric graph $G(r_n)$ is an expander; the geometry only comes into play when $r_n \to 0$ as $n \to \infty$. In this regime, Crescenzi et al. [9] proved that there exist constants γ_1, γ_2 such that if $r_n \geq \gamma_1\sqrt{\log n/n}$ and $c_n \geq \gamma_2 \log(1/r_n)$, then $\Gamma_n(r_n, c_n)$ is connected with high probability. The correct scaling for the connectivity threshold for $r_n \sim \gamma\sqrt{\log n/n}$ for sufficiently large γ was obtained by Broutin et al. [6] who proved that the connectivity threshold for the irrigation graphs with $r_n \sim \gamma\sqrt{\log n/n}$ is

$$c_n^\star := \sqrt{\frac{2\log n}{\log\log n}},$$

independently of the value of γ. More precisely, for any $\epsilon \in (0, 1)$, one has

$$\lim_{n\to\infty} \mathbf{P}\big(\Gamma_n(r_n, c_n) \text{ is connected}\big) = \begin{cases} 0 & \text{if } c_n \leq (1-\epsilon)c_n^\star, \\ 1 & \text{if } c_n \geq (1+\epsilon)c_n^\star. \end{cases} \tag{1}$$

Thus, the irrigation subgraph of a random geometric graph preserves connectivity with high probability while keeping only $O(nc_n^\star)$ edges, which is much less than the $\Theta(n\log n)$ edges of the initial random geometric graph. However, the obtained random irrigation subgraph is not authentically sparse as the average degree still grows with n.

One way to obtain connected sparse random geometric irrigation graphs is to increase the size r_n of the "visibility window" slightly. Indeed, we show elsewhere [7] that by taking r_n larger (but still quite small), as $r_n \sim n^{-1/2+\epsilon}$ for some fixed $\epsilon > 0$, there exists a constant $c = c(\epsilon)$ such that $\Gamma_n(r_n, c)$ is connected with high probability.

Otherwise, one needs to relax the constraint of connectivity, and see how this affects the graph. In this paper we study the emergence of a "giant" component (i.e., a connected component of linear size) of random geometric irrigation graphs when $r_n \sim \gamma\sqrt{\log n/n}$ for a sufficiently large constant γ [i.e., of the same order as the connectivity threshold of the underlying random geometric graph $G_n(r_n)$]. The main result shows that already when $\mathbf{E}\xi > 1$, the graph $\Gamma_n(r_n, \xi)$ has a connected component containing almost all vertices. Interestingly, there is not only a phase transition around a critical value in the edge density but the phase transition is *discontinuous*. More precisely, we show that when $\mathbf{E}\xi = 1$ (or equivalently $\xi = 1$, that is, when the average degree is about 2), the largest component of $\Gamma_n(r_n, \xi)$ has $o(n)$ vertices, while for any $\epsilon > 0$, if $\mathbf{E}\xi = 1 + \epsilon$, then with high probability, $\Gamma_n(r_n, \xi)$ has a component containing $n - o(n)$ vertices. The phenomenon when there is a discontinuous phase transition was coined "explosive percolation" and has received quite a lot of attention recently [24]. In explosive percolation, the size of the largest component, divided by the number of vertices, considered as a function of the average degree, suffers a discontinuous jump from zero to a positive value. In the present case, we have even more: the jump is from zero to the maximal value of one. Therefore, the random graph process experiences a *super-explosive phase transition* or *instant percolation*. The main results of the paper are summarized in the following theorems.

THEOREM 1. *Assume that* $\mathbf{E}\xi > 1$. *For every* $\varepsilon \in (0, 1)$ *there exists a constant* $\gamma > 0$ *such that for* $r_n \geq \gamma\sqrt{\log n/n}$,

$$\mathbf{P}\big(\mathscr{C}_1(\Gamma_n(r_n, \xi)) \geq (1 - \varepsilon)n\big) \underset{n\to\infty}{\longrightarrow} 1,$$

where $\mathscr{C}_1(\Gamma_n(r_n, \xi))$ *denotes the number of vertices in the largest connected component of the graph* $\Gamma_n(r_n, \xi)$.

We observe that, by monotonicity, it suffices to prove Theorem 1 in the case where $\xi \in \{1, 2\}$ with probability one. From now on, we assume that the support of ξ is $\{1, 2\}$. We also prove that when $\xi = 1$, then the largest connected component of the irrigation graph $\Gamma_n(r_n, \xi)$ is sublinear with probability tending to one:

THEOREM 2. *Suppose that* $r_n = o((n \log n)^{-1/3})$. *Then, for any* $\varepsilon > 0$

$$\mathbf{P}\big(\mathscr{C}_1(\Gamma_n(r_n, 1)) \geq \varepsilon n\big) \underset{n\to\infty}{\longrightarrow} 0.$$

The two theorems may be combined to prove the following "instant-percolation" result.

COROLLARY 1. *Suppose* $r_n/\sqrt{\log n/n} \to \infty$ *and* $r_n = o((\log n/n)^{1/3})$. *Then* $\Gamma_n(r_n, \xi)$ *experiences* super-explosive percolation *in the sense that*:

(i) *if* $\mathbf{E}\xi = 1$, *then* $\mathscr{C}_1(\Gamma_n(r_n, \xi)) = o(n)$ *in probability*;
(ii) *if* $\mathbf{E}\xi > 1$, *then* $n - \mathscr{C}_1(\Gamma_n(r_n, \xi)) = o(n)$ *in probability.*

Note that for classical models of random graphs, including both random geometric graphs [25] and Erdős–Rényi random graphs [4, 21], the proportion of vertices in the largest connected component is bounded away from one w.h.p. when the average degree is bounded by a constant. Furthermore, for these graphs, the size of the largest connected component is continuous in the sense that the (limiting) proportion of vertices in the largest connected component vanishes as the average degree tends to the threshold value. The behavior of random geometric irrigation graphs is very different, since the largest connected component contains $n - o(n)$ vertices as soon as the expected degree is greater than two (Theorem 1). In particular, from a practical point of view, the irrigation graph provides an *almost optimal* and *distributed* algorithm for sparsification of the underlying graph. (Here "distributed" refers to the fact that every vertex makes its choices independently, as in distributed algorithms.) Indeed the largest connected component contains $n - o(n)$ nodes, and we achieve this level of connectivity with only $n(1 + \epsilon)$ edges while any such graph must contain at least $n - o(n)$ edges. Note that this relies on the fact that each node chooses at least one neighbor ($\xi \geq 1$), for otherwise there would be a linear number of isolated nodes. (Indeed, in such a case, there would clearly be a linear number of vertices of out-degree zero; the concentration of the number of vertices in balls of radius r_n and a "duality" argument should convince the reader that any such vertex has constant probability to receive no edge from another vertex.)

Recall that when the underlying graph is the complete graph K_n (i.e., when $r \geq \sqrt{2}$), then the irrigation graph model corresponds to the c-out graphs studied by Fenner and Frieze [15] if $\xi = c \in \mathbb{N}$ almost surely. With $c = 1$, a random 1-out subgraph of K_n is a just a random mapping [17, 22] (a uniformly random function from $[n]$ to $[n]$). In particular, for $c = 1$ and with high probability, a random 1-out subgraph of K_n contains a connected component of linear size but is not connected. In particular, this example shows that although the condition that $r_n = o((\log n/n)^{1/3})$ may not be optimal, some upper bound on r_n is clearly required for $\mathscr{C}_1(\Gamma_n(r_n, 1))$ to be $o(n)$ as $n \to \infty$. For $c \geq 2$, a random c-out subgraph of K_n is 2-vertex and 2-edge connected with high probability [15]. (See also [2, 3]: although a bit cryptic, Theorem 3 of Bender [2] applies to unions of two random mappings and shows that such graphs are asymptotically connected.) One may easily verify that if we write $K_n(\xi)$ for the random irrigation subgraph of K_n in which vertex i chooses ξ_i random neighbors and $\mathbf{E}[\xi] > 1$, then as $n \to \infty$,

$$\mathbf{P}\big(K_n(\xi) \text{ is connected}\big) \to 1,$$

as an easy generalization of Fenner and Frieze [15].

2. Preliminaries: Discretization and regularity of the point set. The proof relies heavily on different levels of discretization of the torus into smaller sub-squares, as shown in Figure 1. The largest of these sub-squares are called *cells* and have side length about $kr_n/2$ where k is a fixed large odd natural number. More precisely, let $k \geq 1$ be odd and define

$$m := \left\lceil \frac{2}{kr_n} \right\rceil \quad \text{and} \quad r_n' := \frac{2}{km}. \tag{2}$$

The unit square is then partitioned into m^2 congruent cells of side length $1/m = r_n' k/2$. Note that $(1 - kr_n)r_n \leq r_n' \leq r_n$ for all n large enough.

A cell Q is further partitioned into $k^2 d^2$ square *boxes*, each of side length $1/(mkd) = r_n'/(2d)$, for some natural number $d \geq 1$. We make d odd, so that kd is odd as well, and denote by $C(Q)$ the central box of cell Q. A typical square box is denoted by S, and we let $\mathscr{S}(Q)$ be the collection of all boxes in cell Q.

Note that there are two independent sources of randomness in the definition of the random graph $\Gamma_n(r_n, \xi)$. One comes from the random underlying geometric graph $G_n(r_n)$ (i.e., the collection $\mathbf{X}$ of random points), and the other from the choice of the neighbors of each vertex. We will always work conditionally on the locations of the points in $\mathbf{X}$. The first step is to guarantee that, with high probability, the random set $\mathbf{X}$ satisfies certain regularity properties. In the rest of the proof, we assume that the point set $\mathbf{X}$ satisfies the required regularity property, fix $\mathbf{X}$ and work conditionally.

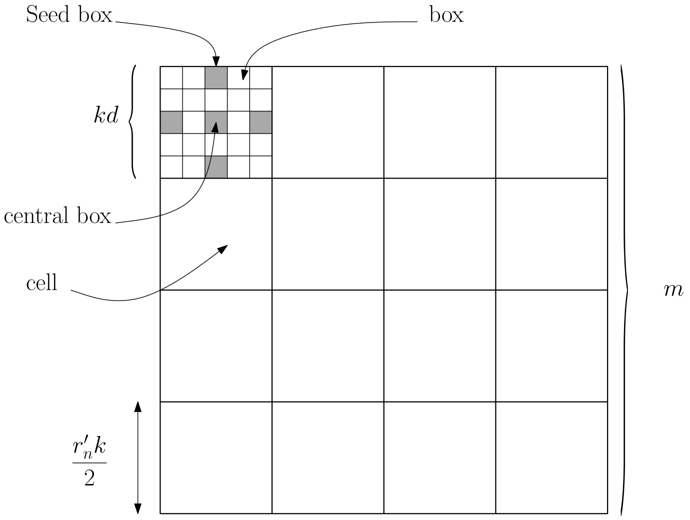

FIG. 1. *The different levels of discretization of the torus* $[0,1]^2$ *are shown here with* $k = 5$, $d = 1$, *and* $m = 4$. *The torus is subdivided into* m^2 *congruent squares, called cells, and each cell is further divided into* $k^2 d^2$ *small squares, called boxes. The central box and the seed boxes of one of the cells are marked.*

In the course of the proofs, we condition on the locations of the points $X_1, \dots, X_n$ and assume that they are sufficiently regularly distributed. The probability that this happens is estimated in the following simple lemma that relies on standard estimates of large deviations for binomial random variables.

Fix odd positive integers k and d and consider the partitioning of $[0,1]^2$ into cells and boxes as described above. For a cell Q, and a box $S \in \mathscr{S}(Q)$, we have

$$\mathbf{E}[|\mathbf{X} \cap S|] = \frac{n}{(mkd)^2} = \frac{nr_n'^2}{4d^2}.$$

Fix $\delta \in (0,1)$. A cell Q is called δ-*good* if for every $S \in \mathscr{S}(Q)$, one has

$$\frac{(1-\delta)nr_n'^2}{4d^2} \le |\mathbf{X} \cap S| \le \frac{(1+\delta)nr_n'^2}{4d^2}.$$

LEMMA 1. *For every* $\delta \in (0,1)$, *there exists* $\gamma > 0$ *such that if* $r_n \ge \gamma\sqrt{\log n/n}$, *then for all n large enough,*

$$\inf_Q \mathbf{P}(Q \textit{ is } \delta\textit{-good}) \ge 1 - 2(mkd)^2 n^{-\gamma^2\delta^2/(24d^2)}.$$

In particular, if $\gamma^2 > 24d^2/\delta^2$ *then*

$$\lim_{n\to\infty} \mathbf{P}(\textit{every cell } Q \textit{ is } \delta\textit{-good}) = 1.$$

See the Appendix for the proof.

3. An overwhelming giant: Structure of the proof.

3.1. *General approach and setting.* Our approach consists in exhibiting a large connected component by exposing the edges, or equivalently the choices of the points, in a specific order so as to maintain a strong control on what happens. The general strategy has two phases: first, a *push*-like phase in which we aim at exposing edges that form a connected graph that is fairly dense almost everywhere; we call this subgraph the *web*. Then, we rely on a *pull*-like phase in which we expose edges from the points that are not yet part of the web and are trying to hook up to it.

The push phase. The design and analysis of the push phase is the most delicate part of the construction. It is difficult to build a connected component with positive density while keeping some control on the construction. For instance, following the directed edges in Γ_n^+ from a single point, say x, in a breadth-first manner produces an exploration of Γ_n^+ that *resembles* a branching process. That exploration needs to look at neighborhoods of x of radius $\Omega(\log\log n)$ in order to reach the $\epsilon \log n$ total population necessary to have positive density in at least one ball of radius r_n. However, by the time the $\Theta(\log\log n)$ neighborhoods have been explored, the spread of the cloud of points discovered extends as far as $\Theta(\log\log n)$ away from x in most directions: in other words, doing this would waste many edges, and make it difficult to control the dependence between the events of reaching positive density in different regions of the square. An important consequence is that is it not reasonable to expect that two points that are close are connected locally: we will prove that points are indeed connected with high probability, but the path linking them does wander far away from them.

To take this observation into account, in a first stage we only build a sort of skeleton of what will later be our large connected component. That skeleton, which we call the web, *does not* try to connect points locally and its aim is to provide an almost ubiquitous network to which points will be able to hook up easily. The construction of this web uses arguments from percolation theory and relies on the subdivision of the unit square into m^2 cells described above. It is crucial to keep in mind that for the construction we are about to describe to work the web should be connected in a directed sense.

The cells define naturally an $m \times m$ grid as a square portion of $\mathbb{Z}^2$, which we view as a directed graph. To avoid confusion with vertices and edges of the graph $\Gamma_n(r_n, \xi)$, we call the vertices of the grid graph *nodes* and its edges *links*. More precisely, let $[m] := \{1, \dots, m\}$. We then consider the digraph Λ_m^+ on the node set $[m]^2$ whose links are the pairs (u, v) whose ℓ_1 distance equals one; the oriented link from u to v is denoted by uv. Write E_m^+ for the link set, so that $\Lambda_m^+ = ([m]^2, E_m^+)$.

The construction of the web uses two main building blocks: we define events on the nodes and the links of Λ_m^+ such that:

- a *node event* is the event that, starting from a vertex in the central box of the cell, if one tracks the selected neighbors of the vertex staying in the cell, then the selected neighbors of these neighbors in the cell and so on up to a number of hops k^2, then the resulting component populates the cell in a uniform manner—see Proposition 1 for the precise statement;
- a *link event* allows the connected component built within the cell to propagate to a neighboring cell. We show that both node events and link event happen with high probability. See Lemma 2.

It is important to emphasize that in proving that node and link events occur with high probability, we make use of a coupling with suitably defined branching random walks that are independent of the precise location of the points at which such events are rooted. Although this does not suffice to make all node and link events become independent, it helps us control this dependence and allows us to set up a joint site/bond percolation argument on $\mathbb{Z}^2$ that proves the existence of a directed connected component that covers most cells. The node and link events are described precisely and the bounds on their probabilities are stated in Sections 3.2 and 3.3, respectively. Finally, in Section 3.4, we show how to combine the node and link events in order to construct the web using a coupling with a percolation process. The proofs of the estimates of the probabilities of the node events are somewhat intricate and we present them in Section 4.

The pull phase. The analysis of the pull phase relies on proving that any vertex not yet explored in the process of building the web is in the same component as the web, with high probability. In order to prove this, one may construct another web starting from such a vertex, which succeeds with high probability by the arguments of the push phase. Then it is not difficult to show that the two webs are connected with high probability. The details are developed in Section 3.5.

3.2. *Populating a cell*: *Node events*. In proving the existence of the web (i.e., a connected component that has vertices in almost every cell), we fix $\delta > 0$ and any point set $\mathbf{X}$ for which every box is δ-good and work conditionally. Thus, the only randomness comes from the choices of the edges. We reveal edges of the digraph Γ_n^+ in a sequential manner. In order to make sure that certain events are independent, once the ξ_i out-edges of a vertex X_i have been revealed, the vertex becomes *forbidden* and excluded from any events considered later. We keep control of the number and density of forbidden points during the entire process.

In Section 3.4, we describe the order in which cells are examined. In this section, we look into a single cell Q and describe an event—the so-called "node event"—that, conditionally on the set of forbidden vertices, only depends on edge choices of vertices within the cell. All we need is a starting vertex $x \in \mathbf{X}$ in the central box $C(Q)$ of the cell and a set F of forbidden vertices. Both x and F may depend on the evolution of process before the cell is examined. However, by construction (detailed below), we guarantee that the set of forbidden vertices F only has a bounded number of elements in each box and, therefore, does not have a significant impact on the outcome of the node event. Similarly, the starting vertex x originates from an earlier stage of the process but its exact location is unimportant, again by the definition of the node event, as detailed below.

Consider a cell Q and a point $x \in \mathbf{X} \cap Q$. For $i \geq 0$, let $\tilde{\Delta}_x(i)$ denote the collection of points of $\mathbf{X} \cap Q$ that can be reached from x by following i directed edges of Γ_n^+ without ever using a point lying outside of Q. Let $F \subset \mathbf{X}$ denote the set of *forbidden* points containing the $y \in \mathbf{X}$ whose choices have already been exposed. Let $\Delta_x(i)$ be the subset of points of $\tilde{\Delta}_x(i)$ that can be reached from x without ever using a point in F.

Recall that the cell Q is partitioned into $k^2 d^2$ square *boxes* of side length $r_n'/(2d)$ and that $\mathscr{S} = \mathscr{S}(Q)$ denotes the collection of boxes of Q. The next proposition shows that, with high probability, any cell Q with starting point $x \in \mathbf{X} \cap Q$ is such that $\Delta_x(k^2)$ populates Q in the following way: for every $S \in \mathscr{S}(Q)$, we have $|\Delta_x(k^2) \cap S| \geq \mathbf{E}[\xi]^{k^2/2}$. We refer to this event as $N_x(Q, F)$ and the corresponding local connected component is called a *bush*. The proof is delayed until Section 4.

PROPOSITION 1. *For a cell Q, vertex $x \in \mathbf{X} \cap C(Q)$ and a set F of forbidden vertices, define the* node event

$$N_x(Q, F) = \{\forall S \in \mathscr{S}(Q) : |\Delta_x(k^2) \cap S| \geq \mathbf{E}[\xi]^{k^2/2}\}.$$

For every $\eta > 0$, there exist constants $d_0, k_0 \geq 1$ and $\delta_0 > 0$ such that for fixed $d \geq d_0$ and $k \geq k_0$, there exists n_0 for which the following holds: for $n \geq n_0$, provided that the cell is δ-good and that $\sup_{S \in \mathscr{S}(Q)} |F \cap S| < \delta_0 n r_n^2$, then for all $x \in \mathbf{X} \cap C(Q)$,

$$\mathbf{P}(N_x(Q, F) | \mathbf{X}) \geq 1 - \eta.$$

We note that although it may seem that the events $N_x(Q, F)$, for distinct cells Q would be independent (they depend on the choices of disjoint sets of vertices), it is not the case. Indeed, the events do interact through the set F, which is random, but that this dependence can be handled.

3.3. *Seeding a new cell: Link events.* We define an event that permits us to extend a bush confined to a cell Q and to find a directed path from it to a point x' in the central box of a neighboring cell Q'.

For a given cell Q, the set of $(kd)^2$ boxes $\mathscr{S}(Q)$ is naturally indexed by

$$\{-\lfloor kd/2 \rfloor, \ldots, \lfloor kd/2 \rfloor\}^2.$$

Among the boxes $S \in \mathscr{S}(Q)$, let $\mathcal{I}(Q)$ denote the collection of the four boxes that correspond to the coordinates $(-\lfloor kd/2 \rfloor, 0)$, $(\lfloor kd/2 \rfloor, 0)$, $(0, -\lfloor kd/2 \rfloor)$ and

$(0, \lfloor kd/2 \rfloor)$ (Figure 1). It is from points in these boxes that we try to "infect" neighboring cells, and we refer to these boxes as *seed boxes* or *infection boxes*. For two adjacent cells Q and Q', we let $I(Q, Q')$ denote the seed box lying in Q against the face shared by Q and Q'. Suppose, as before, that there is a set $F \subset \mathbf{X}$ of forbidden points. For a point $y \in \mathbf{X} \cap I(Q, Q')$, let $\Delta_y^\circ(i)$ denote the points of $\mathbf{X}$ that can be reached from y using i directed edges of Γ_n^+ without using any point lying outside of Q' or in F, except for y itself. Let $J_y(Q, Q')$ be the event that $\Delta_y^\circ(\lceil kd/2 \rceil)$ contains a point lying in the central box of Q':

$$J_y(Q, Q') = \{\Delta_y^\circ(\lceil kd/2 \rceil) \cap C(Q') \neq \varnothing\}.$$

Then, for $R \subset \mathbf{X} \cap I(Q, Q')$, we let $J_R(Q, Q') = \bigcup_{y \in R} J_y(Q, Q')$. The event $J_R(Q, Q')$ is called a *link event*.

LEMMA 2. *Let Q and Q' be two adjacent cells. Suppose that Q' is δ-good for $\delta \in (0, 1/4)$ and that $\sup_{S \in \mathscr{S}(Q)} |F \cap S| < \delta_0 n r_n^2$ for $\delta_0 < 1/(4d)^2$. Then, for every k, and d there exists n_0 such that for every $n \geq n_0$, and for any set $R \subseteq \mathbf{X} \cap I(Q, Q')$, we have*

$$\mathbf{P}(J_R(Q, Q')|\mathbf{X}) \geq 1 - \exp\left(-\frac{|R|}{(10\beta d^2)^{kd}}\right),$$

where $\beta = (1 + \delta)(1/(2d) + k/(16d^2))$.

PROOF. Write $h = \lceil kd/2 \rceil$. Let $L_0 = I(Q, Q'), L_1, \dots, L_h = C(Q')$ denote the sequence of boxes on the straight line from $I(Q, Q')$ to $C(Q')$. For $J_R(Q, Q')$ to occur, it suffices that for some $y \in R$, one has $|\Delta_y^\circ(i) \cap L_i| \geq 1$, for every $i = 1, \dots, h$; call E_y° the corresponding event. The E_y°, $y \in R$, are not independent because the sets $\Delta_y^\circ(i)$, $i \geq 1$, might not be disjoint. However, on $\{\bigcap_{y \in R} \Delta_y^\circ(i) = \varnothing\}$, the events E_y°, $y \in R$, are independent. Consider the ordering of the points in R induced by the ordering in $\mathbf{X}$, and write $x < x'$ if $x = X_i$ and $x' = X_j$ for $i < j$. To simplify the proof, we only consider a single path from any given point $y \in R$. Consider the path defined by $P_0(y) = y$, and for $i \geq 1$, $P_i(y) = Y_1(P_{i-1}(y))$; this is well-defined since $\xi_i \geq 1$ with probability one. Let E_y be event that for every $i = 1, 2, \dots, h$, one has $P_i(y) \in L_i$. Then we have $E_y \subset E_y^\circ$.

Note that since Q' is δ-good, for any box $S \in \mathscr{S}(Q)$ we have, for sufficiently large n,

$$|\mathbf{X} \cap S| \geq \frac{(1 - \delta) n {r'_n}^2}{4d^2} \quad \text{and} \quad |S \cap \mathbf{X} \cap F^c| \geq \frac{n {r'_n}^2}{8d^2}$$

since $\delta < 1/4$ and $\delta_0 < 1/(4d)^2$. Furthermore, at most $|R|h$ of the points of $\mathbf{X} \cap S \cap F^c$ lie in $\bigcup_{y \in R} P_i(y)$, for some $i = 1, 2, \dots, h$. Now, for $y \in R$, let $\tau_y := \inf\{i \geq 1 : P_i(y) \notin L_i\}$. Let $\mathcal{G}_y^-$ be the sigma-algebra generated by $\{P_i(y) :$

$0 \le i < \tau_y\}$, $y' < y$. Since every cell is δ-good, for every $x \in [0,1]^2$, we have $|\mathbf{X} \cap B(x, r_n)| \le \beta n r_n^2$. Then, for every $y \in R$, and all n large enough, we have

$$\mathbf{P}(E_y | \mathcal{G}_y^-) \ge \left(\frac{1}{10\beta d^2}\right)^h \ge \left(\frac{1}{10\beta d^2}\right)^{kd} =: \eta, \tag{3}$$

where the last expression serves as the definition for the constant η. Here, we used the fact that $10\beta d^2 \ge 1$. It follows that $|\{y \in R : E_y\}|$ dominates a binomial random variable $\mathrm{Bin}(|R|; \eta)$ with parameters $|R|$ and η:

$$\begin{aligned}\mathbf{P}(\exists y \in R : E_y^\circ) &\ge \mathbf{P}(\exists y \in R : E_y) \ge \mathbf{P}(\mathrm{Bin}(|R|; \eta) > 0)\\ &\ge 1 - e^{-\eta |R|}.\end{aligned}$$

Replacing η by its expression in (3) yields the claim. □

3.4. *Building the web*: *The percolation process.* In this section, percolation arguments are used to show how the node and link events can be used to build the "web", a connected component that visits most cells. The construction is based on an algorithm to decide which edges to expose depending on what we have seen so far. Once again, fix a point set $\mathbf{X}$ such that every cell is δ-good and work conditionally.

Defining a partial percolation configuration on the square grid of cells. We encode an exploration process on the digraph Λ_m^+ of cells by defining a *partial* and *joint site/bond* percolation process using the node and link events defined above. At the same time, we keep track of the set of forbidden vertices F discussed in Sections 3.3 and 4.

For $u \in [m]^2$, we let $Q_u \subset [0,1]^2$ denote the corresponding cell. The nodes of $[m]^2$ are ordered lexicographically: for $u = (u_1, u_2)$, $v = (v_1, v_2)$ we write $u \preceq v$ if $u_1 \le v_1$ or if $u_1 = v_1$ and $u_2 \le v_2$. We proceed with an exploration process in the lexicographic order, maintaining, at every step $i = 0, 1, 2, \dots$ of the process, a partition of $[m]^2$ into three sets of nodes $[m]^2 = A_i \cup E_i \cup U_i$, where we call the nodes in A_i *active*, the ones in E_i *explored*, and those of U_i *unseen*. Initially, all the nodes are unseen and, therefore, $U_0 = [m]^2$, $A_0 = \varnothing$, $E_0 = \varnothing$. The sets A_i, E_i, U_i, $i \ge 0$, are designed in such a way that:

- at any time $i \ge 0$, any node $u \in A_i$ has a distinguished vertex x_u in the center box $C(Q_u)$ for which we can check if the node event $N_{x_u}(Q_u, F)$ (defined in Proposition 1) occurs; the set of forbidden vertices that is used to assess this event is F_i to be defined shortly.
- The nodes $u \in E_i$ are the ones that have been active from some time $j < i$ and for which the node event $N_{x_u}(Q_u, F)$ has already been observed.

We now move on to the precise description of the algorithm and of the sets $A_i, E_i, U_i \subset [m]^2$, and $F_i \subset \mathbf{X}$. Initially, we set $F_1 = \varnothing$. Then we proceed as follows, for $i \geq 1$. If $E_i = [m]^2$, then we have already "tested" a node event for each node and we are done, and we now suppose that $E_i \neq [m]^2$. Then there must be some node in either A_i or U_i.

(i) Suppose first that $A_i \neq \varnothing$. Then let u_i be the node of A_i that is lowest in the lexicographic order. By construction, there is a distinguished vertex $x_{u_i} \in C(Q_{u_i})$. Say that the node u_i is *open* and set $\tilde{\sigma}(u_i) = 1$ if the node event $N_{x_{u_i}}(Q_i, F_i)$ succeeds. If this is the case, all four seed boxes in Q_{u_i} contain a set of points of the bush constructed in Q_{u_i} of cardinality at least $\mathbf{E}[\xi]^{k^2/2}$ that are all connected to x_{u_i} within Q_{u_i}. Consider all the oriented links $u_i v$, where $v \in U_i$, and let $R_{u_i v}$ be the set of points that are lying in the seed box $S(Q_{u_i}, Q_v)$ of Q_{u_i} that is adjacent to Q_v. For any such link $u_i v$, we declare the oriented link open and set $\tilde{\sigma}(u_i v) = 1$ if the link event $J_{R_{u_i v}}(Q_{u_i}, Q_v)$ (defined just before Lemma 2) succeeds. [Note that we liberally use the notation $\tilde{\sigma}(\cdot)$ to indicate either openness of a node u by $\tilde{\sigma}(u)$ or the openness of an oriented link uv by $\tilde{\sigma}(uv)$].

Let $V_i = \{v \in U_i : \tilde{\sigma}(u_i v) = 1\}$. For every $v \in V_i$, since $J_{R_{u_i v}}(Q_{u_i}, Q_v)$ succeeds, we have, by construction, a nonempty set of points of the center box $C(Q_v)$ that are connected to $R_{u_i, v}$ by directed links in Γ_n^+; we let x_v be the one of these points that has the lowest index in $\mathbf{X}$. Then update the sets by putting $E_{i+1} = E_i \cup \{u_i\}$, $A_{i+1} = A_i \cup V_i \setminus \{u_i\}$, $U_{i+1} = U_i \setminus V_i$. As for the set of forbidden vertices, let f_{i+1} be the collection of points $X_u \in \mathbf{X}$ whose choices $Y_j(X_u)$, $1 \leq j \leq \xi_u$, have been exposed when determining the node event $G_{x_{u_i}}(Q_{u_i})$ and the potential following link events. Then let $F_{i+1} = F_i \cup f_{i+1}$.

(ii) If, on the other hand, $A_i = \varnothing$, then $U_i \neq \varnothing$. Note that if this happens, it means that we have not succeeded in finding a point $x \in C(Q_{u_i})$ that is connected to the points previously explored and we need to start the exploration of a new connected component of Γ_n^+. Let $u_i \in U_i$ be the node with lowest lexicographic order. Then set $A_{i+1} = \{u_i\}$, $E_{i+1} = E_i$ and $U_{i+1} = U_i \setminus \{u_i\}$. We then let x_{u_i} be the point of $\mathbf{X} \cap C(Q_{u_i})$ that has the lowest index in $\mathbf{X}$. Such a point exists by the assumption of δ-goodness and because the number of forbidden points in each cell is bounded (see Lemma 3 below).

Note that the distinguished point x_u of a cell Q_u is chosen when the corresponding vertex is activated, which happens once and only once for every node $u \in [m]^2$.

In order to use Proposition 1 and Lemma 2 for estimating the probability of node events and link events, we need to make sure that the number of forbidden points stays under control.

LEMMA 3. *If k is sufficiently large, then for every cell Q, during the entire process, we have*

$$|F \cap Q| \leq 2^{2k^2}.$$

PROOF. To reveal a node event $N_x(Q, F)$, one only needs to expose $\Delta_x(k^2)$ for a single point $x \in Q$. This requires to look at the edge choices of at most $k^2 2^{k^2}$ vertices, all of which lie in Q. The points exposed during the evaluation of a link event account for a total of at most $4 \cdot k^2 2^{k^2} \cdot kd2^{kd}$. The claim follows easily. □

Completing the percolation configuration. Once the exploration process is finished, every node has been declared open or not, and we have assigned a value to every $\tilde{\sigma}(u)$, $u \in [m]^2$. However, we have not defined the status of all the oriented links uv. See Figure 2. In particular, $\tilde{\sigma}(uv)$ has only been defined if $\tilde{\sigma}(u) = 1$ and if $\tilde{\sigma}(v)$ had not been set to one before. For every oriented link uv, let $\theta(uv)$ be the indicator that a link event has been observed for uv. Let H_m^+ denote the open subgraph of Λ_m^+, that consists of nodes u and directed links uv for which $\tilde{\sigma}(u) = 1$ and $\tilde{\sigma}(uv) = 1$, respectively. A subset K of nodes in $[m]^2$ is called an *oriented connected component* of H_m^+ if $\tilde{\sigma}(u) = 1$ for all $u \in K$ and for all $u, v \in K$ there is an oriented open path between u and v, that is, a sequence $u = u_1, u_2, \dots, u_\ell = v$ of nodes in K such that $\tilde{\sigma}(u_i u_{i+1}) = 1$ for all $i = 1, \dots, \ell - 1$.

In order to prove that H_m^+ contains an oriented connected component containing most nodes—and, therefore, proving the existence of the web—we embed H_m^+ in an *unoriented* complete mixed site/bond percolation configuration on the digraph Λ_m^+. Then we use results from the theory of percolation to assert the existence of a connected component containing most vertices.

In a general mixed site/bond percolation configuration, every node is open independently with a certain probability p, and every undirected link is also open independently with probability q. Fix $\eta \in (0, 1)$ and choose the parameters k, d, λ

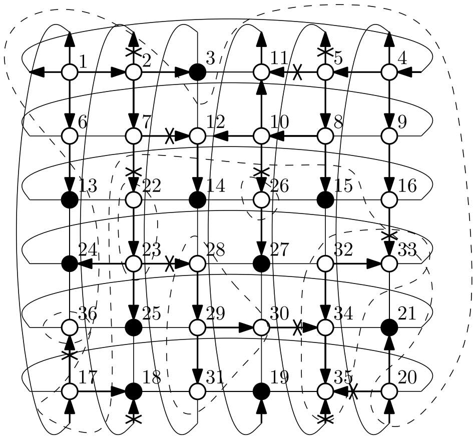

FIG. 2. *The partial percolation configuration after exploring all node and link events with the obtained oriented connected components. White nodes are those for which the node event $N_x(Q, F)$ succeeds. Crossed arrows represent failed link events. The numbers near the nodes indicate the order in which the node events are tested.*

and δ such that each node event occurs with probability at least $1 - \eta$ and each (oriented) link event occurs with probability at least $1 - \eta/2$. In order to define the mixed site/bond percolation configuration, we first assign states to the oriented links uv for which $\theta(uv) = 0$. Let $(\tilde{\sigma}(uv) : uv \in E_m^+, \theta(uv) = 0)$ be a collection of i.i.d. Bernoulli random variables with success probability $1 - \eta/2$. Now we declare an unoriented link uv open if $\tilde{\sigma}(uv) = \tilde{\sigma}(vu) = 1$. Observe that although we have assigned a configuration $(\tilde{\sigma}(u), u \in [m]^2; \tilde{\sigma}(e), e \in E_m^+)$ to the digraph Λ_m^+, this collection of random variables is not independent. For instance, for any two nodes, the $\tilde{\sigma}(u)$ and $\tilde{\sigma}(v)$ are dependent for they interact through the set of F of forbidden nodes which is random. However,

$$\inf_{1\le i\le m^2} \mathbf{P}\bigl(\tilde{\sigma}(u_i) = 1|\mathbf{X}, F_i\bigr) \ge 1 - \eta \quad \text{and}$$

$$\inf_{1\le i\le m^2} \inf_{u_i v\in E_m^+} \mathbf{P}\bigl(\tilde{\sigma}(u_i v) = 1|\mathbf{X}, F_i\bigr) \ge 1 - \eta/2,$$

which implies that there exist two independent collections of i.i.d. Bernoulli random variables $(\sigma(u), u \in [m]^2)$ and $(\sigma(uv), uv \in E_m^+)$ with success probabilities $1 - \eta$ and $1 - \eta/2$, respectively, such that almost surely $\tilde{\sigma}(u) \ge \sigma(u)$ for $u \in [m]^2$ and $\tilde{\sigma}(uv) \ge \sigma(uv)$ for $uv \in E_m^+$. The configuration defined by σ is a proper mixed site-bond percolation configuration.

By construction, in the configuration σ, every node and every unoriented link of $[m]^2$ is equipped with an independent Bernoulli random variable with success probability at least $1 - \eta$. Each node and each link is open if the corresponding Bernoulli variable equals 1. This is the *mixed site/bond percolation* model considered, for example, by [20]. In such a configuration, we say that two nodes $u, v \in [m]^2$ are bond-connected in the configuration if there exists a sequence of open nodes $u = u_1, u_2, \dots, u_\ell = v$ for which every link $u_i u_{i+1}$, $1 \le i < \ell$ is also open. This equivalence relation naturally defines bond-connected components. Clearly, each bond connected component is also an oriented open component in H_m^+ and, therefore, it suffices to show that the mixed site/bond percolation configuration has a bond-connected component containing almost all nodes, with high probability.

In order to prove this, we use results of [11] for high-density site percolation by reducing the mixed site/bond percolation problem to pure site percolation as follows.

LEMMA 4. *Consider mixed site/bond percolation on $[m]^2$ as defined above where each node is open with probability p and each link is open with probability q, independently. The number of nodes of the largest bond-connected component is stochastically dominated by the number of nodes of the largest open component in site percolation on $[m]^2$ where each node is open with probability pq^2.*

PROOF. Split each link in the mixed model into two half-links, and let each half-link be independently open with probability $\sqrt{q}$. We say that a link is open if both half-links are open.

Next, for a node v in the mixed model, we call event $D(v)$ the event that the node and its four adjacent half-links are open. This occurs with probability $r := pq^2$. Now, consider a coupled site percolation model, also on the $m \times m$ torus, in which the node v is open if $D(v)$ occurs. These are independent events. So, we have a site percolation model with node probability $r = pq^2$. It is clear that if a path exists in the site percolation model then a path exists in the mixed model, so the mixed model percolates (strictly) better. □

Now it follows from [11] that in our mixed site/bond percolation model where nodes and links are open with probability at least $1 - \eta$, the following holds: for every $\epsilon > 0$ there exists $\eta > 0$ such that for all m large enough,

$$\mathbf{P}\bigl(\text{there exists a bond-connected component of size} > (1-\epsilon)m^2\bigr) > 1 - \epsilon.$$

Now, for us the constant η is controlled by k, d, δ and γ. Putting everything together, we have proved the existence of the web.

PROPOSITION 2. *Let $\epsilon > 0$. There exist k_0, d_0, δ, γ such that if $k > k_0$, $d > d_0$ and $r_n > \gamma\sqrt{\log n/n}$, then for all n large enough, if all cells are δ-good, then, with probability (conditional on $\mathbf{X}$) at least $1 - \epsilon$, there exists a connected component of $\Gamma(r_n, \xi)$ such that at least $(1-\epsilon)$-fraction of all boxes contain at least $\mathbf{E}[\xi]^{k^2/2}$ vertices of the component.*

3.5. *Finale*: *Gathering most remaining points*. In the previous sections, we saw that after exploring only a constant number of points per cell [at most $m^2 2^{2k^2} \le 2^{2k^2+2}/(k^2 r_n^2)$ in total by Lemma 3] with high probability, we can construct a connected component—the so-called web—of the graph $\Gamma(r_n, \xi)$ that contains at least $\mathbf{E}[\xi]^{k^2/2}$ points in a vast majority of boxes. Recall that each box is a square of side length $r_n'/(2d)$ where d is a fixed but large odd integer.

It remains to prove that most other vertices belong to the same component as the web, with high probability. To this end, first we show that any not yet explored vertex is contained in the same component as the web, with high probability. As before, we fix a sufficiently small $\delta > 0$ and fix a point set $\mathbf{X}$ such that every cell is δ-good. Suppose that the exploration process of the previous sections has been carried out, revealing the edge choices of at most 2^{2k^2} points per cell (and thus also at most this many per box). If n is so large that $\delta^2\gamma \log n/(4d^2) > 2^{2k^2}$, then even after removing all vertices whose choices have already been exposed, every cell remains 2δ-good. Let $x_i \in \mathbf{X}$ be one of the still unseen vertices. We shift the coordinate system so that the box containing x_i becomes the central box of the first cell. Since the boxes in the new coordinate system were also boxes in the original coordinates, every cell is still 2δ-good in the new system.

Now we build a second web, with the aim that it contains x_i with probability close to one. We start the same exploration process from the vertex x_i as in the construction of the web but now we place all vertices of the first web in the set of forbidden points. If δ is sufficiently small, then Proposition 2 applies and, with probability at least $1-\epsilon$, we obtain another web that has at least $\mathbf{E}[\xi]^{k^2/2}$ vertices in at least $(1-\epsilon)$-fraction of the boxes. The newly built web may not contain the vertex x_i. However, by the homogeneity of the mixed site/bond percolation process, each cell is equally likely to be contained in the newly built web and, therefore, with probability at least $(1-\epsilon)^2$ vertex x_i is contained in a component that has at least $\mathbf{E}[\xi]^{k^2/2}$ vertices in at least a $(1-\epsilon)$-fraction of the boxes. It is clear from the proof of Proposition 1 that, in fact, each of these boxes contains at least $\mathbf{E}[\xi]^{k^2/2}$ points whose edge choices have not been revealed in the process of building the second web. Thus, with probability at least $(1-\epsilon)^3$, at least $(1-2\epsilon)$-fraction of the boxes contain at least $\mathbf{E}[\xi]^{k^2/2}$ points of the first web and at least $\mathbf{E}[\xi]^{k^2/2}$ points of the second web that contains the vertex x_i. Now we may reveal the edge choices of the vertices of the second web that have not been explored. Since the diameter of a box is less than r_n, the probability that the two webs do not connect—if they have not been connected already—is at most

$$(4)\qquad \left(1-\frac{\mathbf{E}[\xi]^{k^2/2}}{(1-2\delta)nr_n^2/(4d^2)}\right)^{m^2(1-2\epsilon)}=o(1)$$

whenever $r_n=o(n^{-1/4})$.

Recall that F denotes the number of forbidden vertices, that is, the vertices that have been seen at some point in the process of constructing the first web. Let $\mathscr{B}$ be the “bad” event that some cell is not δ-good. Then

$$(5)\quad \mathbf{P}(\mathscr{C}_1(\Gamma_n(r_n,\xi))<(1-\varepsilon)n)\le\mathbf{P}(\mathscr{C}_1(\Gamma_n(r_n,\xi))<(1-\varepsilon)n|\mathscr{B}^c)+\mathbf{P}(\mathscr{B}).$$

Writing $\mathcal{W}$ for the set of points of $\mathbf{X}$ that eventually lie within the connected component of the first web we constructed, we have

$$\begin{aligned}\mathbf{P}(\mathscr{C}_1(\Gamma_n(r_n,\xi))<(1-\varepsilon)n|\mathscr{B}^c)&\le\mathbf{P}\left(\sum_{i=1}^n\mathbf{1}\{X_i\notin\mathcal{W},X_i\notin F\}+|F|>\varepsilon n\middle|\mathscr{B}^c\right)\\&\le\frac{n}{n\epsilon+m^2 2^{2k^2}}\mathbf{P}(X_1\notin\mathcal{W}|X_1\notin F,\mathscr{B}^c),\end{aligned}$$

where the last line follows from Markov’s inequality and Lemma 3. Together with (4), (5) and Lemma 1, this proves that for any $\varepsilon>0$, the largest connected component of $\Gamma_n(r_n,\xi)$ contains at least $(1-\varepsilon)n$ vertices for all n large enough and, therefore, completes the proof of Theorem 1. Notice that the gathering of most remaining points given the existence of the web is extremely likely to happen, and that the constraints on parameters are essentially the ones required to prove existence of the web (Proposition 2).

4. Getting out of the central box: Proof of Proposition 1.

4.1. *Constructing a branching random walk.* Most of the work consists in estimating the probability of the local events, while ensuring independence. In this section, we consider a single cell Q of side length $kr'_n/2$. As we have already explained, the local bushes are constructed by a process that resembles a *branching random walk* in the underlying geometric graph. The main differences with an actual branching random walk are that:

- the potential individuals are the elements of $\mathbf{X}$, and so they are fixed conditionally on $\mathbf{X}$,
- an element of $\mathbf{X}$ only gets to choose its neighbors once; in particular, if a vertex X_i is chosen that has already used up its ξ_i choices, the corresponding branch of the exploration must stop (if we were to continue the exploration, it would trace steps that have already been discovered).

We note that closely related arguments relying on the comparison with branching random walks have been used by Penrose [26] in the context of continuum percolation in high dimension and by Häggström and Meester [19] to treat the case of nearest neighbor graphs and hard sphere models. For these two cases, the branching random walk approximation becomes relevant as the dimension increases; here, the dimension remains fixed and equal to two, but the parameter that allows us to take advantage of the approximation is the large number of boxes $(kd)^2$ per cell.

The entire argument in this section is conditional on the location of the points, assuming the regularity property that the cell Q is δ-good. Recall that a cell Q of side-length $kr'_n/2$ is called δ-good if the number of points within every box S lies within a multiplicative $[1-\delta, 1+\delta]$ range of its expected value $\mathbf{E}|\mathbf{X} \cap S|$. By Lemma 1, for any $\epsilon > 0$, the probability that every cell is δ-good is at least $1-\epsilon$ for all n large enough (provided the constant γ in $r_n = \gamma\sqrt{\log n/n}$ is large enough).

Fix a cell Q, in which we want to analyze the node event. Then, for every $i \in [n]$ such that $X_i \notin Q$, we work with an independent copy ξ'_i of ξ_i. (That is, we resample ξ_i for every point that is used outside of the cell Q; this way, we are certain that the outcome does not depend on the actual values of ξ_i.) Since such points X_i are not considered when analyzing the node event on $\mathbf{X}$, this has no effect on the event $N_x(Q, F)$, for $x \in \mathbf{X} \cap C(Q)$. However, this makes the proofs a little smoother since we can look at all k^2 neighborhoods without worrying (at least in a first stage) whether the points are in Q or not. Note the important fact that this is only used for the *analysis*, and that the actual exploration is not carried out when the points leave the cell Q. In particular, this thought experiment does not affect the number of forbidden vertices.

Discretizing the steps. We define a *skimmed* version of the neighborhood exploration in which we drop some of the points in order to guarantee simplified dynamics. The simplification uses the refined discretization of the space into boxes. Let $\mathscr{S}$ denote the collection of all boxes (open squares). For a point $x \in [0,1]^2$, we let $S(x) \subset Q$ denote the box containing the point x (this is well-defined for every point of $Q \cap \mathbf{X}$ with probability one).

For a given point $x \in Q \subseteq [0,1]^2$, let A_x° denote the union of the boxes that are fully contained in the ball of radius r_n centered at x. (Note that although $x \in Q$, A_x° may contain boxes lying outside of Q.) Then, for any box $S \in \mathscr{S}$ and $x \in S$ define

$$A_x = \bigcap_{y \in S} A_y^\circ$$

(Figure 3). Write $\mathcal{A}_x$ for the collection of boxes whose union is A_x, and let a denote the number of boxes that compose $\mathcal{A}_x$, for $x \in [0,1]^2$.

LEMMA 5. *For any $\delta \in (0, 1/4)$ there exist constants k_0, d_0 and n_0 such that for all $k \geq k_0$, $d \geq d_0$ and $n \geq n_0$, we have*

$$|a - 4\pi d^2| \leq \delta d^2.$$

PROOF. First, every box that is fully contained in $B(x, r_n)$ accounts for an area of $(r_n'/(2d))^2$, so we must have $a(r_n'/(2d))^2 \leq \pi r_n^2$. Now, since $(1 - kr_n)r_n \leq$

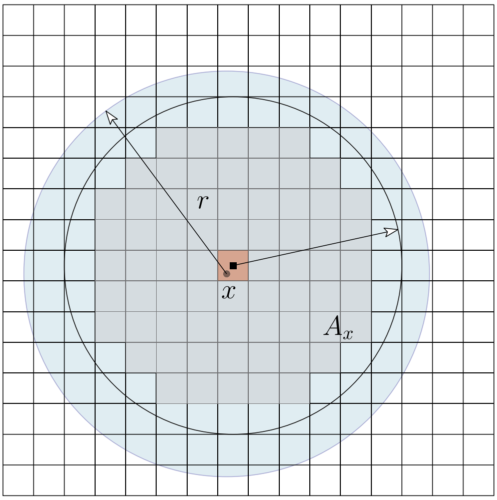

FIG. 3. *Any circle of radius r_n centered at x fully contains A_x which is a copy of a* fixed *collection of boxes.*

$r_n' \le r_n$ [see just below the definition in (2)], it follows that

$$a \le 4\pi d^2 \left(\frac{r_n}{r_n'}\right)^2 \le 4\pi d^2 \cdot \frac{1}{1 - kr_n} \le 4\pi d^2(1+\delta),$$

provided that $kr_n \le \delta/2 < 1/4$.

On the other hand, the boxes that intersect the ball $B(x, r_n)$ but are not fully contained in it must touch the boundary of $B(x, r_n)$. So any such box must lie entirely inside the annulus $B(x, r_n + r_n'\sqrt{2}/(2d)) \setminus B(x, r_n - r_n'\sqrt{2}/(2d))$. In particular, the number of these boxes is at most

$$\frac{\pi}{(r_n'/(2d))^2}\left\{\left(r_n + \frac{r_n'\sqrt{2}}{d\sqrt{2}}\right)^2 - \left(r_n - \frac{r_n'}{d\sqrt{2}}\right)^2\right\} \le 8\pi\sqrt{2}d\frac{r_n}{r_n'}.$$

Since $(1 - kr_n)r_n \le r_n' \le r_n$, it follows that

$$\begin{aligned} a &\ge 4\pi d^2 \left(\frac{r_n}{r_n'}\right)^2 - 8\pi d\sqrt{2}\frac{r_n}{r_n'} \\ &\ge 4\pi d^2 - 16\pi d\sqrt{2}, \end{aligned}$$

for all k and r_n such that $2kr_n \le 1$. The result follows readily. □

Extracting a discrete branching random walk. Next, we obtain lower bounds for the sizes of neighborhoods of points x in the irrigation graph. It is here that the discretization in boxes is important since it ensures that the process $S(y)$, $y \in \Delta_x$, of boxes containing the points which may be reached from x using directed edges *dominates* a branching random walk. In order to properly extract this branching random walk on the set of boxes:

(a) we artificially reduce the number of points in each box so that the distributions of the number of offspring of the spatial increments are *homogeneous*, which fixes the spatial component; this has to be done dynamically, since each time a point is discovered, one extra point should be removed in every other box to maintain this property;

(b) we then kill some more branches of the process to ensure that the offspring of every individual has the same distribution, so that the underlying genealogy is a Galton–Watson tree.

We now define the discrete branching random walk as a process indexed by the infinite plane tree

$$\mathcal{U} = \bigcup_{n \ge 0} \{1, 2, 3, \ldots\}^n,$$

where the individuals in the nth generation are represented by a word of length n on the alphabet $\{1, 2, \ldots\}$. The tree $\mathcal{U}$ is seen as rooted at the empty word $\varnothing$. The descendants of an individual u are represented by the words have u as a prefix. The children of $u \in \mathcal{U}$ are ui, $i \geq 1$. If $v = ui$ for some $u \in \mathcal{U}$ and $i \geq 1$, u is the parent of v and is denoted by $p(v) = u$. For $u, v \in \mathcal{U}$, we write $u \preceq v$ if u is an ancestor of v, potentially $u = v$. Consider any δ-good cell Q, and a point $x \in Q$. The construction is done in stages:

- we first define $(Z_u^\bullet, u \in \mathcal{U})$, corresponding to the points in $\mathbf{X}$ that can be reached from x;
- then we define $(Z_u^\circ, u \in \mathcal{U})$, as a subset of $(Z_u^\bullet, u \in \mathcal{U})$ for which all the Z_u° are distinct and the spatial increments are homogeneous [see (a) above];
- finally, we define $(Z_u, u \in \mathcal{U})$ which is a subset of $(Z_u^\circ, u \in \mathcal{U})$ for which the progenies of all nodes are identically distributed [see (b) above].

Exploring the neighborhoods in Γ_n^+. Let us now proceed with the details of the definition of the tree-indexed process $(Z_u^\bullet, u \in \mathcal{U})$ corresponding to the exploration of the neighborhoods of x in the directed irrigation graphs Γ_n^+. To properly prune the tree, we introduce a cemetery state ∂, that is assigned to the words of $\mathcal{U}$ that do not correspond to a vertex of $\mathbf{X}$. We set $Z_\varnothing^\bullet = x$; then if $Z_u^\bullet = z \in \mathbf{X}$ for some $u \in \mathcal{U}$, we define $Z_{ui}^\bullet = Y_i(z)$, for $1 \leq i \leq \xi_z$, and $Z_{ui}^\bullet = \partial$ for $i > \xi_z$ and similarly for any word w with ui as a prefix. Then, for an integer $m \geq 0$, $\{Z_u^\bullet : |u| = m, Z_u^\bullet \neq \partial\}$ is precisely the set of points of $\mathbf{X}$ that can be reached from x using a path of length exactly m. However, some points may appear more than once in the process.

Getting each point only once and the spatial components. We now *skim* the process $(Z_u^\bullet, u \in \mathcal{U})$ in order to extract a subprocess $(Z_u^\circ, u \in \mathcal{U})$ for which all the points are distinct, and for which we can guarantee that the process induced on the set of boxes visited by the points is a branching random walk. Although it is not crucial, it is natural to skim the tree in lexicographic order on $\mathcal{U}$.

The skimming is done by maintaining a set of *valid* points, which ensures that a point chosen at random among the valid points in a certain subset of the boxes is contained in a uniformly random box in that subset. Initially, we have a set of *valid* points $V \subseteq \mathbf{X} \setminus F$, that consists of points whose choices have not yet been exposed, but maybe not all such points. We choose V in such a way that for every box $S \in \mathscr{S}$, the number of elements of $V \cap S$ is the same, and we denote by c this the common cardinality. We do this in such a way that the set V has maximal cardinality. We observe that, as the new points are discovered and assigned to $(Z_u^\circ, u \in \mathcal{U})$, the set of valid points have to be updated to maintain the property that each box contains the same number of them.

Let $(w_i, i \geq 0)$ be the breadth-first ordering of the elements of the set $\{u : Z_u^\bullet \neq \partial, |u| \leq k^2\}$. Set $V_{w_0} = V_\varnothing = V$. If $x \notin V_\varnothing$, we kill the entire tree by setting $Z_\varnothing^\circ = \partial$ as well as for all the words $u \in \mathcal{U}$. Otherwise, $x \in V_\varnothing$, and we set $Z_\varnothing^\circ = x$. Then we update the set of valid points: for each box that does not contain x, one valid point must be removed. More precisely, for each box $S \in \mathscr{S} \setminus \{S(x)\}$, let $X_{(0)}(S)$ be the point of $V_\varnothing \cap S$ which has minimum index in $\mathbf{X}$, if such a point exists, or $X_{(\varnothing)}(S) = \partial$ otherwise. Then we let V_{w_1} be obtained from $V_\varnothing = V_{w_0}$ by removing x together with all the points $X_\varnothing(S)$, $S \in \mathscr{S}$:

$$V_{w_1} := V_\varnothing \setminus \big(\big\{X_{(\varnothing)}(S) : S \in \mathscr{S}\big\} \cup \{x\}\big).$$

This ensures that the number of points in $V_{w_1} \cap S$ is the same for every box $S \in \mathscr{S}$, and equal to $c - 1$, since precisely one point has been removed from every box.

For all the subsequent steps, we only keep a point z if (1) it is in the current set of valid points, and (2) it falls in the region A_y, where $y \in \mathbf{X}$ denotes the point from which we arrive at z. So suppose now that we have defined $Z^\circ_{w_j}$ for all $j < i$, and V_{w_j} for $j \le i$. Suppose also that, for every $S \in \mathscr{S}$, $V_{w_i} \cap S$ has cardinality exactly $c - i$. Recall that $p(u)$ denotes the parent of a node $u \in \mathcal{U}$. Then, if $Z^\bullet_{w_i} \in V_{w_i} \cap A_{Z_{p(w_i)}}$, we set $Z^\circ_{w_i} = Z^\circ_{w_i}$. Otherwise, that steps fails, and we define $Z^\circ_{w_i} = \partial$, as well as for all the nodes in the subtree of $\mathcal{U}$ rooted at w_i. Now, regardless of the success/failure of this step (i.e., if $Z^\circ_{w_i} \ne \partial$ or $Z^\circ_{w_i} = \partial$), we update the set of valid points so that $V_{w_{i+1}}$ has exactly $c - (i + 1)$ points in every box $S \in \mathscr{S}$. Let $S_i \in \mathscr{S}$ be the box containing $Z^\bullet_{w_i}$. Then, for every box $S \in \mathscr{S}$, let $X_{(i)}(S)$ be the point in $V_{w_i} \cap S$ which has minimum index in $\mathbf{X}$. Let $V_{w_{i+1}}$ be obtained from V_{w_i} by removing all the $X_{(i)}(S)$, for the boxes $S \ne S_i$, and either $X_{(i)}(S_i)$ or $Z^\bullet_{w_i}$ itself according to whether $Z^\circ_{w_i} = \partial$ or not. More formally, we have

$$V_{w_{i+1}} = V_{w_i} \setminus \big\{X_{(i)}(S) : S \in \mathscr{S}, S \ne S_i\big\} \setminus \begin{cases} \{X_{(i)}(S_i)\} & \text{if } Z^\circ_{w_i} = \partial, \\ \{Z^\circ_{w_i}\} & \text{if } Z^\circ_{w_i} \ne \partial. \end{cases}$$

This way, we have that for every $S \in \mathscr{S}$, $V_{w_{i+1}} \cap S$ has cardinality precisely $c - (i + 1)$.

Skimming the underlying genealogy. The process of interest is the tree-indexed process of boxes containing the points discovered by the exploration of increasing neighborhoods in Γ_n^+, $(S(Z^\circ_u), u \in \mathcal{U})$. Note that for $u, v \in \mathcal{U}$, with u the parent of v in $\mathcal{U}$, $u = p(v)$, conditional on $Z^\circ_u, Z^\circ_v \ne \partial$, and say $S(Z^\circ_u) = s$, the dynamic updates in the set of valid points and the pruning of branches imply that the box $S(Z^\circ_v)$ which contains Z_v is uniformly random in $\mathcal{A}_s$. So for every sequence of words $(v_i, i \ge 0)$ in $\mathcal{U}$ with $|v_i| = i$, conditional on $Z_{v_\ell} \ne \partial$, the process $(S(Z^\circ_{v_i}))_{0 \le i \le \ell}$ is a random walk with i.i.d. increments. The only reason why the entire process $(S(Z^\circ_u), u \in \mathcal{U})$ is not a branching random walk is that the individuals do not all jump to ∂ with the same probability (in other words the individuals do not all have the same offspring distribution) either because of the inhomogeneity of the point set $\mathbf{X}$, or because of the changing number of valid points. We now construct the process $(Z_u, |u| \le k^2)$ by homogenizing the offspring distribution; this is simply done by (1) proving that the progeny distributions of the nodes in $(Z^\circ_u, |u| \le k^2)$ all stochastically dominate a common progeny distribution ν,

(2) killing extra branches at random for all the offspring to be distributed as ν. Recall that a is the number of boxes in the collection $\mathcal{A}_z$, for every point z, and that $\rho(z)$ denotes the number of points of $\mathbf{X}$ lying within distance r_n of z. For $u \in \mathcal{U}$, and $s \geq 0$, define

$$\alpha_i := \frac{a(c-i)}{\rho(Z^\circ_{p(w_i)})}; \tag{6}$$

then α_i is the probability that $Z^\circ_{w_i} \neq \partial$ conditional on the event that for the parent $p(w_i)$ of w_i in $\mathcal{U}$, $Z^\circ_{p(w_i)} \neq \partial$. Let also $\alpha := \inf\{\alpha_i : 1 \leq i \leq i_m\}$, where $i_m := \#\{u : Z^\bullet_u \neq \partial, |u| \leq k^2\}$. Let U_i, $i \geq 1$, be a collection of i.i.d. random variables uniformly distributed on $[0, 1]$, and finally define

$$Z_{w_i} = \begin{cases} Z^\circ_{w_i} & \text{if } U_i \leq (1-\alpha)/(1-\alpha_i), \\ \partial & \text{otherwise.} \end{cases}$$

Then, for every u, $|u| < k^2$, writing $\zeta_u := \#\{ui : Z_{ui} \neq \partial\}$ for the number of offspring of u, $(\zeta_u : |u| < k^2)$ is a collection of i.i.d. random variables; write ζ for the typical copy of this random variable. More precisely, ζ_u is distributed like a binomial random variable with parameters ξ_u and α. Now, of special interest is the mean of the offspring distribution, which is controlled by the value of $\alpha \in (0, 1)$, and we must ensure that it can be made close enough to one by choice of the parameters. By (6), Lemma 3, Lemma 5 and the fact that all cells are assumed δ-good, we have

$$\alpha \geq \frac{(4\pi - \delta)d^2(\eta - \delta)\log n - 2^{2k^2}}{(4\pi + \delta)d^2(\eta + \delta)\log n}. \tag{7}$$

It follows that, if we write $\mathbf{E}[\xi] = 1 + \epsilon$ for $\epsilon > 0$, it is possible to choose δ, d_1, n_1 large enough such that for $d \geq d_1$ and $n \geq n_1$, we have

$$\mathbf{E}[\zeta] \geq 1 + \epsilon/2,$$

so that the underlying Galton–Watson process is supercritical.

4.2. *Analyzing the discrete branching random walk.* In this section, we slightly abuse notation and identify the set of boxes and their representation as the discrete torus. Furthermore, since for n large enough, the difference between the torus and Z^2 cannot be felt by a walk of k^2 steps, we talk about $\mathbb{Z}^2$. In particular, we let $\mathcal{A}$ denote the subset of $\mathbb{Z}^2$ corresponding to the boxes in $\mathcal{A}_0$, which is the set of potential spatial increments of our walks.

We now consider the (truncated) *branching random walk* $(Z_u, |u| \leq k^2)$ taking values in $\mathbb{Z}^2$, that we complete into a branching random walk by generating the missing individuals using an independent family of random variables for the offspring and the spatial displacements. By definition, an individual u located at Z_u gives birth to ζ_u individuals, such that the displacements are i.i.d. uniform in $\mathcal{A}$.

Furthermore, every individual behaves in the same way and independently of the others. For $S \in \mathscr{S}$ and $i \geq 0$, define

$$M_i(S) := \#\{u \in \mathcal{U} : |u| = i, Z_u \in S\},$$

the number of individuals $u \in \mathcal{U}$ in generation i such that $Z_u \in S$.

LEMMA 6. *Let $q > 0$ be the extinction probability of the Galton–Watson process underlying the branching random walk $(Z_u)_{u\in\mathcal{U}}$. Then, for all k large enough, we have*

$$\mathbf{P}(\#\{v \in \mathcal{U} : |v| = k^2, Z_v \in S\} \leq \mathbf{E}[\zeta]^{2k^2/3}) \leq 2q.$$

Before proving Lemma 6, we show that the extinction probability q in the bound may be made as small as we want by choice of the constants. By the bound in (7), this reduces to showing that the extinction probability goes to zero as $\alpha \to 1$.

LEMMA 7. *Let q be the extinction probability of a Galton–Watson process with offspring distribution $\zeta = \mathrm{Bin}(\xi, \alpha)$ such that $\mathbf{E}[\xi]\alpha > 1$ and $\xi \geq 1$. Then*

$$q \leq \frac{1 - \alpha}{1 - \mathbf{E}[(1-\alpha)^{\xi}] - \mathbf{E}[\xi\alpha(1-\alpha)^{\xi-1}]}.$$

PROOF. To prove this, we use the standard fact that q is the smallest $x \in [0, 1]$ such that $x = \mathbf{E}[x^{\zeta}]$ [1]. Note the simple fact that if $f(x)$ and $g(x)$ are probability generating functions, then if $f(x) \leq g(x)$ for all $x \in [0, 1]$ the corresponding extinction probabilities q_f and q_g satisfy $q_f \leq q_g$. So it suffices to find an upper bound on $\mathbf{E}[x^{\zeta}]$, which gives us a computable (and small) extinction probability. Writing $p_i = \mathbf{P}(\zeta = i)$, and $p_{\geq 2} = 1 - p_0 - p_1$, we have, for every $x \in [0, 1]$,

$$\mathbf{E}[x^{\zeta}] \leq p_0 + (1 - p_0 - p_{\geq 2})x + p_{\geq 2}x^2.$$

It follows readily that

$$q \leq \frac{(p_0 + p_{\geq 2}) - |p_0 - p_{\geq 2}|}{2p_{\geq 2}} = \frac{\min\{p_0, p_{\geq 2}\}}{p_{\geq 2}} \leq \frac{p_0}{p_{\geq 2}}.$$

Here, $p_{\geq 2} = 1 - \mathbf{E}[(1-\alpha)^{\xi}] - \mathbf{E}[\xi\alpha(1-\alpha)^{\xi-1}]$ and since $\xi \geq 1$, we have $p_0 \leq 1 - \alpha$, which completes the proof. □

The proof of Lemma 6 goes in two steps. First, one shows that for some $\delta > 0$, the branching random walk has at least $(1 + \epsilon/2)^{\delta k}$ individuals in the δkth generation, that all lie within distance $k/4$ of the center of the cell. We call such individuals *decent*. The decent individuals are the starting points of independent branching random walks. In order to prove the claim, we show that, with probability no smaller than a polynomial in $1/k$, a single of these decent individuals produces enough descendants for $\#\{v \in \mathcal{U} : |v| = k^2, Z_u \in S\} \geq \mathbf{E}[\zeta]^{2k^2/3}$ to occur.

PROOF OF LEMMA 6. Consider the genealogical tree of the branching random walk $(Z_u)_{u\in\mathcal{U}}$, and write $(M_i)_{i\ge 0}$, for the associated Galton–Watson process. So we have

$$M_i = \#\{Z_u : u \in \mathcal{U}, |u| = i\}.$$

As we already mentioned, we have $\mathbf{E}[\zeta] > 1$ and the process is supercritical. Furthermore, the offspring distribution is bounded ($\zeta \le 2$) so that Doob's limit law [1] implies that, as $m \to \infty$, we have as $\ell \to \infty$,

$$\frac{M_\ell}{\mathbf{E}[M_1]^\ell} \to W$$

in distribution, for some random variable W that is absolutely continuous, except possibly at 0. Furthermore, the limit random variable satisfies $\mathbf{P}(W = 0) = q$, where q is the extinction probability of the Galton–Watson process $(M_i)_{i\ge 0}$.

Since $q \in (0, 1)$, we can find a $\beta > 0$ such that $\mathbf{P}(2\beta < W < 1/(2\beta)|W > 0) > 1 - q$. It follows that

$$\liminf_{\ell\to\infty} \mathbf{P}\biggl(\frac{M_\ell}{\mathbf{E}[M_1]^\ell} \in \biggl[\beta, \frac{1}{\beta}\biggr]\biggr) \ge \mathbf{P}\biggl(W \in \biggl[2\beta, \frac{1}{2\beta}\biggr]\biggr) > (1 - q)^2,$$

and in particular, for any $\mu \in (0, 1/2)$ and k large enough,

$$\mathbf{P}\bigl(M_{\lfloor\mu k\rfloor} \ge \beta\mathbf{E}[M_1]^{\mu k-1}\bigr) \ge (1 - q)^2.$$

Recall that an individual v is *decent* if $\|Z_v\| \le k/4$, where $\|\cdot\|$ denotes the Euclidean distance. However, the spatial increments are bounded, and for every v such that $|v| = \lfloor\mu k\rfloor$, we have

$$\|Z_v\| \le \mu \le \mu k 2d.$$

It follows that for $\mu \in (0, 1/(8d))$, *every* individual v with $|v| = \lfloor\mu k\rfloor$ is decent. Fix now such a μ. Writing D_m for the number of decent individuals at level m, we have

$$\mathbf{P}\bigl(D_{\lfloor\mu k\rfloor} < \beta\mathbf{E}[M_1]^{\mu k-1}\bigr) \le 1 - (1 - q)^2. \tag{8}$$

For every decent individual at depth $\lfloor\mu k\rfloor$, there is a subtree that might well give us enough individuals at generation k^2 all lying in S. In order to ensure some level of concentration, we only consider the individuals u, $|u| = \lfloor\mu k\rfloor$, for which the corresponding Doob limit W_u in the subtree rooted at u satisfies $2\beta < W_u < 1/(2\beta)$. For $\ell \ge 0$ and u such that $|u| \le \ell$ write

$$M_\ell(u) := \#\{v : u \preceq v, |v| = \ell\}.$$

Then, for all k large enough,

$$\mathbf{E}\bigl[M_{k^2}(u)|2\beta < W_u < 1/(2\beta), u \text{ decent}\bigr] \ge \beta\mathbf{E}[M_1]^{k^2-\lfloor\mu k\rfloor} \cdot k^{-c}$$

for some $c > 0$ whose existence is guaranteed by Lemma 8 below. However, for every such individual u, on the event $\{2\beta < W_u < 1/(2\beta)\}$, we have for all k large enough

$$M_{k^2}(u, S) := \#\{v : u \preceq v, |v| = k^2, Z_v \in S\} \le \beta^{-1}\mathbf{E}[M_1]^{k^2-\lfloor \mu k\rfloor}.$$

It follows that

$$\mathbf{P}\biggl(M_{k^2}(u, S) \ge \frac{\beta}{2}\mathbf{E}[M_1]^{k^2-\lfloor \mu k\rfloor}k^{-c}\Big| 2\beta < W_u < 1/(2\beta), u \text{ decent}\biggr) \ge \frac{\beta}{2}\cdot k^{-c},$$

hence

$$\mathbf{P}\biggl(M_{k^2}(u, S) \ge \frac{\beta}{2}\mathbf{E}[M_1]^{k^2-\lfloor \mu k\rfloor}k^{-c}\Big| u \text{ decent}\biggr) \ge \frac{\beta}{2}k^{-c}(1-q)^2. \tag{9}$$

Finally, combining (8) and (9), we see that, for k large enough, the probability that we do not have at least $\mathbf{E}[M_1]^{2k^2/3}$ individuals of the k^2th generation that lie in B is at most

$$\begin{aligned}
&\mathbf{P}(D_{\lfloor \mu k\rfloor} < \beta\mathbf{E}[M_1]^{\mu k-1}) + \mathbf{P}(M_{k^2}(u, S) < \mathbf{E}[M_1]^{2k^2/3}|u \text{ decent})^{\beta\mathbf{E}[M_1]^{\mu k-1}}\\
&\quad\le 1-(1-q)^2 + 2^{-k\mu} + \biggl(1-\frac{\beta}{2}k^{-c}(1-q)^2\biggr)^{\beta\mathbf{E}[M_1]^{\mu k}}\\
&\quad\le 2q,
\end{aligned}$$

for k sufficiently large. □

It remains to prove the following key ingredient of the proof of Lemma 6.

LEMMA 8. *Let $(R_i)_{i\ge 0}$ be a random walk on $\mathbb{Z}^2$ where the increments are i.i.d. uniformly random in $\mathcal{A}$ (where $\mathcal{A}$ is defined above just before Lemma 6). Then, for any $\mu \in (0, 1/2)$, there exists a constant $c > 0$, such that for any $x \in \{-\lfloor kd/4\rfloor, \dots, \lfloor kd/4\rfloor\}^2$ and $y \in \{-\lfloor kd/2\rfloor, \dots, \lfloor kd/2\rfloor\}^2$,*

$$\mathbf{P}(R_{k^2-\mu k} = y; R_i \in \mathscr{S}, 0 \le i \le k^2-\mu k|R_0 = x) \ge k^{-c},$$

for all k large enough.

PROOF. Let $\mathscr{S} = [-\lfloor kd/2\rfloor, \lfloor kd/2\rfloor]^2$ be the scaled version of the cell. The argument relies on the strong embedding theorem of Komlós et al. [23] or, more precisely, its multidimensional version by Zaitsev [29] (see also [14]): there exists a coupling of $(R_i)_{0\le i\le k^2}$ with a Brownian motion $(\Xi_t)_{0\le t\le k^2}$ such that for every $c_2 > 0$ there exists a $c_1 > 0$ such that, for every k large enough,

$$\mathbf{P}\Bigl(\max_{0\le i\le k^2} \|R_i - \Xi_i\| \ge c_1 \log k\Bigr) \le k^{-c_2}, \tag{10}$$

where $\|\cdot\|$ denotes Euclidean norm in $\mathbb{R}^2$. We now consider such a coupling, and for a constant c_1 to be chosen later, let $E = E(c_1)$ be the event that $\|R_i - \Xi_i\| \le c_1 \log k$ for every $0 \le i \le k^2$. Let $y \in \{-\lfloor kd/2\rfloor, \dots, \lfloor kd/2\rfloor\}^2$, and recall that $S(y)$ denotes the corresponding box in $\mathscr{S}$. On E, if it turns out that $\Xi_{k^2-\mu k-m} \in S(y)$, then $R_{k^2-\mu k-m}$ is reasonably close to y and there is a decent chance that it ends up at y at time $k^2 - \mu k$. More precisely, if for some integer $m \le k$ we have $\Xi_{k^2-\mu k-m} \in S(y)$, then $\|R_{k^2-\mu k-m} - y\| \le c_1 \log k$, and we let $H = H(c_1, m)$ denote the latter event. Then we have

$$\begin{aligned}
&\mathbf{P}\bigl(H, R_i \in \mathscr{S}, 0 \le i \le k^2 - \mu k - m\bigr)\\
&\quad\ge \mathbf{P}\bigl(H, R_i \in \mathscr{S}, 0 \le i \le k^2 - \mu k - m, E\bigr)\\
&\quad\ge \mathbf{P}\bigl(\Xi_{k^2-\mu k-m} = S(y), \Xi_i \in \mathscr{S}, 0 \le i \le k^2 - \mu k - m, E\bigr)\\
&\quad\ge \mathbf{P}\bigl(\Xi_{k^2-\mu k-m} = S(y), \Xi_i \in \mathscr{S}, 0 \le i \le k^2 - \mu k - m\bigr) - \mathbf{P}\bigl(E^c\bigr).
\end{aligned}$$

Now, by the local limit theorem, for all k large enough, one has

$$\inf_{0\le m\le k} \mathbf{P}\Bigl(\Xi_{k^2-\mu k-m} \in S(y); \inf_{1\le i\le k^2} d\bigl(\Xi_i, \mathscr{S}^c\bigr) \le c_1 \log k\Bigr) \ge k^{-2},$$

where $d(x, \mathscr{S}^c)$ denotes the distance from $x \in \mathbb{R}^2$ to the set $\mathscr{S}^c$. Choosing c_1 be the constant such that $c_2 = 3$ in (10), we obtain

$$\inf_{0\le m\le k} \mathbf{P}\bigl(H, R_i \in \mathscr{S}, 0 \le i \le k^2 - \mu k - m\bigr) \ge k^{-3}, \tag{11}$$

for all k large enough. In particular, with $m = \lfloor c_1 \log k/(2d)\rfloor$ it is possible for the random walk to go to y within the m steps, while staying within $\mathscr{S}$. It follows that, with $a := |\mathcal{A}|$ the number of potential increments at every step,

$$\mathbf{P}\bigl(R_{k^2-\mu k} = y, R_{k^2-\mu k-i} \in \mathscr{S}, 0 \le i \le m | H(c_1, m)\bigr) \ge a^{-m}. \tag{12}$$

Putting (11) and (12) together completes the proof for $c = 3 + c_1 \log a$. □

5. An upper bound on the size of the largest component for $c = 1$.

In this section, we prove Theorem 2 about the size of the largest component of $\Gamma_n(r_n, 1)$. Write $\mathscr{C}_1 = \mathscr{C}_1(\Gamma_n(r_n, 1))$ for the number of vertices of the largest connected component. Although Theorem 2 is suboptimal, the condition on r_n cannot be replaced altogether, because it is easy to show that for fixed $r_n > 0$ large enough, $\mathscr{C}_1 = \Theta(n)$ with high probability when $\xi = 1$ almost surely. Indeed, as we already mentioned, if $r_n \ge \sqrt{2}$, the underlying random geometric graph is the complete graph, so that Γ_n^+ is the graph of a random mapping. Such a random mapping has a largest connected component of linear size; see Theorem 3 in [17]. This is also the case for sequences r_n that tend to 0 slowly with n.

The main technical result is the following tail bound on the size of the largest connected component.

LEMMA 9. *Let* $r_n > 0, t \geq 1, \epsilon > 0$. *Then*

$$\mathbf{P}\big(\mathscr{C}_1 \geq 2 + \big(1 + tnr_n^2\big)^3 (1+\epsilon)^2 \log^2 n\big) \leq n^{-\epsilon + 1/(1+tnr_n^2)} + n^2 e^{(n-2)\pi r_n^2 (t-1-t\log t)}.$$

PROOF. For $\xi = 1$, the structure of the graph is that of a mapping and Γ_n is of a collection of connected components each of which consist of either a tree or a unique cycle from which some trees are pending. In order to bound the number of vertices $\mathscr{C}_1$ of the largest connected component, we first bound the length of the longest directed path in Γ_n^+. Since the edges bind vertices that are at most r_n apart, this bounds the extent of the connected components hence their sizes.

Recall that $\rho_{r_n}(x) = |B(x, r_n) \cap \mathbf{X}|$ denotes the number of X_i's in $B(x, r_n)$. We first show that for $t > 1$ (and thus, $t - 1 - t\log t < 0$), we have

$$\mathbf{P}\bigg(\max_{1\leq i\leq n} \sup_{s \geq r_n} \frac{\rho_s(X_i) - 2}{ns^2} \geq t\bigg) \leq n^2 e^{(n-2)\pi r_n^2 (t-1-t\log t)}.$$

Observe that the supremum in this inequality is reached for $\rho_s(X_i)$ for some $s = \|X_i - X_j\|$. Also, $\rho_s(x)$ is distributed as a binomial random variable with parameters n and πs^2 (we are in the torus), and $\rho_{\|X_i - X_j\|}(X_i)$ is approximately equal to $2 + \mathrm{Bin}(n-2; \pi\|X_i - X_j\|^2)$. By Chernoff's bound [8] (see also [10, 21]), for $u > 1$,

$$\mathbf{P}\big(\mathrm{Bin}(k; p) \geq ukp\big) \leq e^{kp(u-1-u\log u)},$$

so that here, we have

$$\mathbf{P}\big(\mathrm{Bin}\big(n-2, \pi s^2\big) \geq 2 + u(n-2)\pi s^2\big) \leq e^{(n-2)\pi s^2 (u-1-u\log u)}.$$

Thus,

$$\begin{aligned}\mathbf{P}\bigg(\max_{1\leq i\leq n} \sup_{s \geq r_n} \frac{\rho_s(X_i) - 2}{ns^2} \geq t\bigg) &\leq \binom{n}{2} \sup_{s\geq r_n} e^{(n-2)\pi s^2 (t-1-t\log t)} \\ &\leq n^2 e^{(n-2)\pi r_n^2 (t-1-t\log t)}.\end{aligned}$$

Introduce the event

$$A := \bigg\{\max_{1\leq i\leq n} \sup_{s \geq r_n} \frac{\rho_s(X_i) - 2}{ns^2} < t\bigg\}.$$

Starting from a vertex i, we can follow the directed links in Γ_n^+, forming a *maximal* path P_i of distinct vertices. The last vertex j in this path must be pointing toward a vertex k of P_i (potentially itself). From each vertex in P_i the probability of linking to a k higher up in the path is at least

$$\frac{1}{\rho_{r_n}(X_i) - 1} \geq \frac{1}{1 + tnr_n^2}$$

if A occurs. Writing $|P_i|$ for the number of vertices of P_i, we see that, since the choices of links are independent,

$$\mathbf{P}(|P_i| > \ell) \le \left(1 - \frac{1}{1 + tnr_n^2}\right)^{\ell}.$$

By the union bound, conditional on $X_1, \dots, X_n$ such that A holds,

$$\mathbf{P}\Big(\max_{1\le i\le n} |P_i| > \ell\Big) \le n\left(1 - \frac{1}{1 + tnr_n^2}\right)^{\ell} \le n \exp\left(-\frac{\ell}{1 + tnr_n^2}\right).$$

Now, if the maximum length of a directed path $\max_i |P_i|$ is no more than ℓ, then every vertex is within ℓ edges of *any* vertex of the unique cycle of the connected component. Thus, if this occurs, then every connected component is contained within a ball $B(X_j, r_n\ell)$ for some $1 \le j \le n$. It follows that

$$\begin{aligned}\mathbf{P}(\mathscr{C}_1 \ge 2 + n(r_n\ell)^2 t) &\le \mathbf{P}(A^c) + \mathbf{P}(A, \mathscr{C}_1 \ge 2 + n(r_n\ell)^2 t)\\ &\le \mathbf{P}(A^c) + \mathbf{P}\Big(A, \max_{1\le i\le n} |P_i| > \ell\Big)\\ &\le n^2 e^{(n-2)\pi r_n^2 (t-1-t\log t)} + ne^{-\ell/(1+tnr_n^2)}.\end{aligned}$$

For fixed $\epsilon > 0$, take $\ell = \lfloor (1 + tnr_n^2)(1+\epsilon)\log n\rfloor$. We conclude that

$$\mathbf{P}\big(\mathscr{C}_1 \ge 2 + (1 + tnr_n^2)^3 (1+\epsilon)^2 \log^2 n\big) \le n^{-\epsilon + 1/(1+tnr_n^2)} + n^2 e^{(n-2)\pi r_n^2(t-1-t\log t)},$$

which completes the proof of the lemma. □

PROOF OF THEOREM 2. Lemma 9 can be used for various ranges of r_n. In the entire proof, we use it with $\epsilon = 2$ to ensure that the first term in the upper bound there is $o(1)$. We split the region $r_n \in [0, o((n\log n)^{-1/3})]$ into two, and first consider

$$r_n \le \sqrt{\frac{\log n}{\pi n}}.$$

In this range, we define t as the solution of

$$t\log t + 1 - t = \frac{3\log n}{\pi n r_n^2}.$$

Observe that since the right-hand side is at least $3 > 1$, there is indeed a unique solution. Note that this solution could have an infinite limit supremum, but its limit infimum is larger than one (so that one can use Lemma 9 with this value for t). Moreover, one has

$$t = \Theta\left(\frac{3\log n}{\pi n r_n^2 \log((3\log n)/(\pi n r_n^2))}\right),$$

so that

$$(1+tnr_n^2)^3(\log n)^2 = \Theta\left(\frac{\log^5 n}{\log^3((3\log n)/(\pi n r_n^2))}\right) \le \Theta(\log^5 n).$$

By Lemma 9, in this range of r_n, we have $\mathscr{C}_1 \le C\log^5 n$ with probability tending to one as $n \to \infty$, where C is a fixed constant (uniform over all sequences r_n in this range).

Next, consider

$$r_n \ge \sqrt{\frac{\log n}{\pi n}}.$$

Define $t_0 = 3.59112167\ldots$ as the unique solution greater than one of $t_0 \log t_0 = t_0 + 1$. With this choice, if $t > t_0$, the upper bound in the inequality of Lemma 9 is $o(1)$. Note that

$$(1+tnr_n^2)^3(\log n)^2 = \Theta(n^3(\log n)^2 r_n^6),$$

which is $o(n)$ if $r_n = o((n\log n)^{-1/3})$. Overall, we have proved that $\mathscr{C}_1 = o(n)$ as long as $r_n = o((n\log n)^{-1/3})$. □

6. Concluding remarks and open questions.

From a practical point of view, the sparsification done via irrigation graphs is especially interesting since an average degree of $(1+\epsilon)$ guarantees that the majority of the nodes are part of the network. It is proved in [6] that catching all the outsiders would require an average degree of about $\Theta(\sqrt{\log n/\log\log n})$, so that it might not be worth the cost.

Theorem 2 is suboptimal in the range it allows for r, and it would be interesting to find a wider range of r for which one does not have a connected component of linear size. It is not quite clear that there is a threshold since the property that there exists a connected component of size at least cn is not clearly monotonic in r for fixed ξ. It would be of interest to know whether the property that a giant exists with high probability is monotonic in r_n (for fixed ξ): is it the case that if a giant exists w.h.p. for a given r_n and fixed ξ, then a giant also exists w.h.p. for any sequence r_n' with $r_n' \ge r_n$ and the same fixed ξ? Assuming this is the case, it would be interesting to study where the threshold $r^\star = r^\star(\xi)$ is for the existence of a giant when $\xi = 1$, but also for other (constant) values.

The question of the spanning ratio of the giant component is another interesting one. Of course, for ξ such that $\mathbf{E}\xi \ge 1+\epsilon$, the largest connected component has unbounded spanning ratio if we consider the definition

$$\max_{1\le i,j\le n} \frac{\|X_i - X_j\|}{d_\Gamma(i,j)},$$

where d_Γ denotes the graph distance in $\Gamma_n(r_n,\xi)$. However, even if we disallow the pairs of points that are either disconnected or too close, that is, for which

$\|X_i - X_j\| \leq r$, it is not clear that the ratio becomes bounded. Indeed, our construction only guarantees that most points in the same cells get connected via two webs that hook up potentially far from that cell. In [6], it is shown that the spanning ratio of $\Gamma(r_n, c_n)$ is bounded w.h.p. when $r_n \geq \gamma\sqrt{\log n/n}$ and $c_n \geq \mu\sqrt{\log n}$ for sufficiently large constants γ and μ.

Our techniques only show that when $\mathbf{E}[\xi] > 1$ the largest connected component spans most of the vertices, but we have no control on the number of vertices that are left over. The question of the size of the second largest connected component may possibly be tackled by guessing which configurations are most "economical" in terms of avoiding to connect to the outside world, as in [6].

Finally, in order for $\mathbf{X}$ to be sufficiently regular, we assume that $nr_n^2 \geq \gamma \log n$ for γ sufficiently large. Our techniques would fail for values of r_n that are closer to the connectivity threshold of the random geometric graph. However, this regime is especially intriguing and it would be interesting to see what happens for the size of the largest component of the irrigation graph for such values of r_n. Also, there is also no fundamental reason why one should restrict oneself to values of r_n above the connectivity threshold: what happens even for r_n below the connectivity threshold. Of course, one cannot expect anymore that for $\mathbf{E}[\xi] > 1$, the size of the largest connected component is almost n: does it span almost the entire giant component of the random geometric graph or is it asymptotically smaller?

APPENDIX: PROOF OF UNIFORMITY LEMMA

PROOF OF LEMMA 1. For any box S, the number points $|\mathbf{X} \cap S|$ is distributed like a binomial random variable with parameters n and $r'^2/(4d^2)$. By a classical concentration bound for binomial random variables (see, e.g., [5, 21]), we have for $\delta \in (0, 1)$ and $p \in (0, 1)$,

$$\mathbf{P}\big(|\mathrm{Bin}(n, p) - np| \geq \delta np\big) \leq 2e^{-np\delta^2/3}. \tag{13}$$

Now, every cell Q contains k^2d^2 boxes, and by the union bound we have, for all n large enough,

$$\begin{aligned}\mathbf{P}(Q \text{ is not } \delta\text{-good}) &\leq 2k^2d^2e^{-nr_n'^2\delta^2/(3\cdot 4d^2)}\\ &\leq 2k^2d^2n^{-\gamma^2\delta^2/(24d^2)},\end{aligned}$$

since $\sqrt{2}r_n' \geq r_n$ for any $k \geq 1$ and all n large enough. Furthermore, if there exists one cell that is not δ-good, then one of the $(mkd)^2$ boxes has a number of points that is out of range, so that as $n \to \infty$,

$$\begin{aligned}\mathbf{P}(\exists Q : Q \text{ is not } \delta\text{-good}) &\leq 2(mkd)^2n^{-\gamma^2\delta^2/(24d^2)}\\ &\leq n^{1-\gamma^2\delta^2/(24d^2)+o(1)},\end{aligned}$$

which tends to zero provided that $\gamma^2 \geq 24d^2/\delta^2$. □

Acknowledgments. Part of this work has been conducted while attending a workshop at the Banff International Research Station for the workshop on Models of Sparse Graphs and Network Algorithms. We are grateful to the referees for their careful reading and their suggestions.

REFERENCES

[1] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. *Die Grundlehren der mathematischen Wissenschaften* **196**. Springer, New York. MR0373040

[2] BENDER, E. A. (1974). Asymptotic methods in enumeration. *SIAM Rev.* **16** 485–515. MR0376369

[3] BENDER, E. A. (1975). An asymptotic expansion for the coefficients of some formal power series. *J. Lond. Math. Soc.* (2) **9** 451–458. MR0398846

[4] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

[5] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193

[6] BROUTIN, N., DEVROYE, L., FRAIMAN, N. and LUGOSI, G. (2014). Connectivity threshold of Bluetooth graphs. *Random Structures Algorithms* **44** 45–66. MR3143590

[7] BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2015). Connectivity of sparse Bluetooth networks. *Electron. Commun. Probab.* **20** Art. ID 48. MR3358970

[8] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23** 493–507. MR0057518

[9] CRESCENZI, P., NOCENTINI, C., PIETRACAPRINA, A. and PUCCI, G. (2009). On the connectivity of Bluetooth-based ad hoc networks. *Concurrency and Computation*: *Practice and Experience* **21** 875–887.

[10] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036

[11] DEUSCHEL, J.-D. and PISZTORA, A. (1996). Surface order large deviations for high-density percolation. *Probab. Theory Related Fields* **104** 467–482. MR1384041

[12] DUBHASHI, D., HÄGGSTRÖM, O., MAMBRINI, G., PANCONESI, A. and PETRIOLI, C. (2007). Blue pleiades, a new solution for device discovery and scatternet formation in multi-hop Bluetooth networks. *Wireless Networks* **13** 107–125.

[13] DUBHASHI, D., JOHANSSON, C., HÄGGSTRÖM, O., PANCONESI, A. and SOZIO, M. (2005). Irrigating ad hoc networks in constant time. In *Proceedings of the Seventeenth Annual ACM Symposium on Parallelism in Algorithms and Architectures* 106–115. ACM, New York.

[14] EINMAHL, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68. MR0996984

[15] FENNER, T. I. and FRIEZE, A. M. (1982). On the connectivity of random m-orientable graphs and digraphs. *Combinatorica* **2** 347–359. MR0708149

[16] FERRAGUTO, F., MAMBRINI, G., PANCONESI, A. and PETRIOLI, C. (2004). A new approach to device discovery and scatternet formation in Bluetooth networks. In *Proceedings of the* 18*th International Parallel and Distributed Processing Symposium*.

[17] FLAJOLET, P. and ODLYZKO, A. M. (1990). Random mapping statistics. In *Advances in Cryptology—EUROCRYPT* '89 (*Houthalen*, 1989). *Lecture Notes in Computer Science* **434** 329–354. Springer, Berlin. MR1083961

[18] GILBERT, E. N. (1961). Random plane networks. *J. Soc. Indust. Appl. Math.* **9** 533–543. MR0132566

[19] HÄGGSTRÖM, O. and MEESTER, R. (1996). Nearest neighbor and hard sphere models in continuum percolation. *Random Structures Algorithms* **9** 295–315. MR1606845

[20] HAMMERSLEY, J. M. (1980). A generalization of McDiarmid's theorem for mixed Bernoulli percolation. *Math. Proc. Cambridge Philos. Soc.* **88** 167–170. MR0569643

[21] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. Wiley-Interscience, New York. MR1782847

[22] KOLCHIN, V. F. (1986). *Random Mappings*. Optimization Software, Inc., Publications Division, New York. MR0865130

[23] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412

[24] PANAGIOTOU, K., SPÖHEL, R., STEGER, A. and THOMAS, H. (2011). Explosive percolation in Erdős–Rényi-like random graph processes. *Electron. Notes Discrete Math.* **38** 699–704.

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

[26] PENROSE, M. D. (1996). Continuum percolation and Euclidean minimal spanning trees in high dimensions. *Ann. Appl. Probab.* **6** 528–544. MR1398056

[27] PENROSE, M. D. (2016). Connectivity of soft random geometric graphs. *Ann. Appl. Probab.* **26** 986–1028. MR3476631

[28] PETTARIN, A., PIETRACAPRINA, A. and PUCCI, G. (2009). On the expansion and diameter of Bluetooth-like topologies. In *Algorithms—ESA* 2009. *Lecture Notes in Computer Science* **5757** 528–539. Springer, Berlin. MR2557780

[29] ZAITSEV, A. Y. (1998). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM Probab. Stat.* **2** 41–108 (electronic). MR1616527

N. BROUTIN
INRIA PARIS—ROCQUENCOURT
DOMAINE DE VOLUCEAU
78153 LE CHESNAY
FRANCE
E-MAIL: nicolas.broutin@inria.fr

L. DEVROYE
MCGILL UNIVERSITY
3480 UNIVERSITY STREET
MONTREAL, QUÉBEC H3A 0E9
CANADA
E-MAIL: luc.devroye@gmail.com

G. LUGOSI
DEPARTMENT OF ECONOMICS
POMPEU FABRA UNIVERSITY
RAMON TRIAS FARGAS 25-27
08005, BARCELONA
SPAIN
E-MAIL: gabor.lugosi@gmail.com