Connectivity Properties of Bluetooth Wireless Networks

Devdatt Dubhashi[*] Olle Häggström[†] Alessandro Panconesi[‡]

November 4, 2003

Abstract

We propose a model for geometric random graphs motivated by wireless communication networks adopting the Bluetooth standard. We obtain asymptotic and high probability results concerning connectivity and the emergence of connected giant components in such random graphs. In particular, global connectivity is possible to obtain even as the typical number of links connecting a given node remains constant. Our results show how a very simple, constant-time distributed algorithm can compute connected spanning subgraphs of low average degree in random unit disk graphs, with high probability.

1 Introduction

In the study of wireless communication networks, the following setup is natural. A number n of nodes are distributed in some geographical region. Links can be established between any pair of nodes within a given distance r from each other.

> *What is a good strategy for connecting nodes to each other, in order to obtain good connectivity properties of the resulting global network?* (1.1)

We shall for simplicity take the geographical region to be the unit square $[0,1]^2$, and also make the probabilistic model assumption that the positions of the n nodes are random: independent and uniformly distributed on $[0,1]^2$.

An obvious answer to question (1.1) is to establish links between *all* pairs of nodes within distance r from each other. This may, however, be very costly, so one would like to find a localized, distributed strategy (i.e., not requiring global optimization or coordination) which keeps the number of links incident to each node small. We propose the following strategy.

Definition 1 *Fix* $r > 0$ *and a positive integer* c*. We take* $G^n_{r,c} = (V^n_{r,c}, E^n_{r,c})$ *to denote the geometric random graph defined as follows.*

- *The vertex set* $V^n_{r,c}$ *consists of* n *points, picked independetly according to the uniform distribution on* $[0,1]^2$.
- *Each node* $v \in V^n_{r,c}$ *connects to* c *nodes chosen uniformly at random among those within distance* r*. If the number of such nodes is less than* c *then* v *connects to all of them. This is done independently for all nodes* v*.*

---

[*] Department of Computing Science, Chalmers University, SE 412 96 Göteborg, Sweden, `dubhashi@cs.chalmers.se`

[†] Department of Mathematics, Chalmers University, SE 412 96 Göteborg, Sweden, `olleh@math.chalmers.se`

[‡] Informatica, La Sapienza, via Salaria 113, 00198, Roma, Italy, `ale@dsi.uniroma1.it`

The resulting graph $G^n_{r,c}$ *is called the* **Bluetooth graph** *with parameters* r*,* c*, and* n*.*

We emphasize that the degree (i.e., the number of links incident to) a given node v may exceed c, because in addition to the c links created by v, there may be further links resulting from other nodes choosing to connect to v. However, the total number of links cannot exceed cn, and therefore the average degree among the n nodes cannot exceed $2c$.

The model is motivated by the study of networks adopting the Bluetooth standard, one of the most promising technologies for multihop wireless networking. As shown by Ferraguto et al. [3], $G^n_{r,c}$ approximates very well the graphs generated by certain protocols for the Bluetooth so-called device discovery phase. When a Bluetooth network starts operating each device explores its neighborhood in order to establish reliable communication links with devices within transmision range. The goal of device discovery is to set up enough links in order to have a connected network (see [3] and references therein). Since device discovery is both time and energy consuming it is not possible to set up all possible links. The problem then becomes that of finding a local strategy to set up a subset of all possible links in order to have high probability of global connectivity. In [3] a device discovery strategy where each device connects to c neighbours is studied, and the Bluetooth graph of Definition 1 makes this notion matematically precise. This approach leads to device discovery protocols outperforming all existing solutions (in practical implementations a couple of additional requirements are enforced that make the protocol more effective but much harder to analyze).

The nature of the networks generated by the real Bluetooth device discovery protocol is difficult to analyze due to the complicated nature of that protocol. However, a rather extensive set of experiments – see [3] – show that $G^n_{r,c}$, simple as it is, is a very good model. The same study shows that $G^n_{r,c}$ and hence, the real BT graph, is, above a certain point density, very likely to have a huge giant component, consisting of about 95% of the vertices, and, for a higher density, is very likely to be connected.

The emergence of a giant component is interesting and potentially useful. For instance in coverage applications it opens up the possibility of being able to monitor the area of interest with a connected network using a significantly smaller number of devices.

In this paper we present a probabilistic analysis that validates these empirical observations. Our first result says, roughly, that the Bluetooth graph $G^n_{r,c}$ with large n exhibits a giant component. We now state it precisely. We write $\mathbf{P}$ for probability (and will sometimes for clarity write $\mathbf{P}_n$ rather than $\mathbf{P}$ to emphasize that the probability model depends on the number of nodes n).

Definition 2 *Let* $s \in (0, 1]$ *be some fixed constant. An* s-giant component *of an undirected graph* G *is a connected subgraph of* G *containing at least* ns *of the vertices.*

Proposition 3 *Fix* $r > 0$ *and* $c \geq 2$*. Then there exists a constant* $s > 0$ *such that*

$$\lim_{n\to\infty} \mathbf{P}(G^n_{r,c} \textit{ has an s-giant component}) = 1 . \tag{1.2}$$

The next result shows that this is even true with $s = 1$, i.e., the graph becomes connected:

Theorem 4 *Fix* $r > 0$ *and* $c \geq 2$*. Then*

$$\lim_{n\to\infty} \mathbf{P}(G^n_{r,c} \textit{ is connected}) = 1 . \tag{1.3}$$

Of course, Theorem 4 implies Proposition 3. The reason we state Proposition 3 separately is that, as we shall see, it forms a natural intermediate step towards proving Theorem 4. In fact, our proofs show that both statements holds with high probability. The probability of the complementary events goes to 0 as $\Theta(n^{-\epsilon})$ for some $\epsilon \in (0, 1)$.

This shows that the very simple (constant time!) distributed algorithm embodied in Definition 1 computes with high probability a connected spanning subgraph of constant average degree. In fact, most of the nodes will have very low degree. Spanning graphs like these are considered to be a useful routing infrastructure to distribute traffic evenly. This is one of the reasons why the minimum degree spanning tree problem has received much attention recently (see for instance [8] and, for a distributed implementation, [2]).

We shall consider one more model, which is natural in the context of networks that are generated by Bluetooth device discovery protocols, but which have the additional complication that links may fail.

Definition 5 *Fix $r > 0$, $c \geq 2$, and $p \in (0, 1)$. The graph $G^n_{r,c,p} = (V^n_{r,c,p}, E^n_{r,c,p})$ is defined as the geometric random graph obtained by the following procedure.*

- *First, generate a Bluetooth graph $G^n_{r,c} = (V^n_{r,c}, E^n_{r,c})$ as in Definition 1.*
- *Then, for each edge $e \in E^n_{r,c}$ independently, delete e with probability $1 - p$ (thus keeping it with probability p).*

The resulting graph $G^n_{r,c,p}$ is called a **thinned Bluetooth graph** *with parameters r, c, p and n.*

Our main result for thinned Bluetooth graphs is the following, which shows that, provided p is not too small, the giant component result (Proposition 3) extends to the thinned Bluetooth model. However, there is a striking qualitative difference in that the connectedness result (Theorem 4) does *not* obtain:

Theorem 6 *Fix $r > 0$, $c \geq 2$, and $p > \frac{1}{c}$. Then there exists a constant $s > 0$ such that*

$$\lim_{n\to\infty} \mathbf{P}(G^n_{r,c,p} \text{ has an } s\text{-giant component}) = 1 . \tag{1.4}$$

On the other hand,

$$\lim_{n\to\infty} \mathbf{P}(G^n_{r,c,p} \text{ is connected}) = 0 . \tag{1.5}$$

We remark that an easy branching process comparison shows that the emergence of a giant component *fails* when p is sufficiently small. In fact, our main technique for proving the results stated above is also comparisons with branching processes, but these are somewhat more sophisticated. This difference comes from the fact that for the small p result, the comparison works by showing that a certain branching process *dominates* the connected component containing a given vertex, while for our main results, the comparison goes the other way.

The use of branching process comparisons to establish connectivity properties is quite standard in percolation theory (see, e.g., Meester and Roy [11]), but deserves to become better known in the study of wireless networks.

Our main results – Proposition 3, Theorem 4, and the first half of Theorem 6 – can all be seen as instances of what appears to be a general principle in percolation theory: if the selection mechanism with which nodes connect to other nodes is sufficiently “spread out”, i.e. if it chooses randomly among very many, then rather few links per node will suffice to obtain good global connectivity. See [12, 10, 5] for other results in this direction. In contrast, it appears that if the device discovery mechanism in Definition 1 is replaced by one where each node connects to its c *nearest* neighbors, then $c = 2$ is not enough to obtain a giant component (as indicated by the simulations reported in [6]), while no finite c suffices for connectedness [4].

In the next section, we prove Proposition 3, while Sections 3 and 4 are devoted to the proofs of Theorems 4 and 6, respectively. Finally, in Section 5, we mention some possibilities for extensions and future work.

2 Bluetooth: giant component

For the purpose of proving Proposition 3, and also later, Lemma 7 below will be useful. Fix an integer k such that

$$k > \frac{\sqrt{5}}{r} \tag{2.6}$$

and partition $[0,1]^2$ into k^2 subsquares of size $\frac{1}{k} \times \frac{1}{k}$ in the obvious way. (One point of this choice of k is that it ensures that any two points sitting in adjacent subsquares are within distance r from each other; this will be needed in the proof of Theorem 4.)

Lemma 7 *For any fixed k, we have*

$$\lim_{n\to\infty} \mathbf{P}_n\left(\textit{each of the } k^2 \textit{ subsquares contains at least } \tfrac{n}{2k^2} \textit{ points}\right) = 1\,. \tag{2.7}$$

Proof: Fix a square S and let X denote the number of points in S. Then, $\mu := EX = \frac{n}{k^2}$ and, by the Chernoff-Hoeffding bound,

$$\mathbf{P}_n(X < \frac{n}{2k^2}) \leq e^{-n/8k^2}.$$

Thus, the probability that some square has less than the required number of points is at most $k^2 e^{-n/8k^2}$. $\square$

To investigate connected components of the Bluetooth graph, we shall employ the following method, which we will call the **sequential discovery procedure** (this is simply a breadth-first exploration). First, select a node v_0 at random (among all n nodes). Then consider the c edges chosen by v_0 (in the device discovery procedure of Definition 1), and denote the endpoints (other than v_0) of these edges by $v_1, \ldots, v_c$. Then continue with the edges chosen by v_1, and so on, in a breadth first search manner. Each time a new node is encountered, the node reached by it is included in our list of nodes, and the choice is deemed a **success**. Sometimes, the edge leads to a node already seen in this procedure, in which case the choice is said to be a **failure**. At any point of this search procedure, we may stop, and those vertices encountered whose outgoing edges have not been investigated (yet), are called **fresh** nodes.

At various stages of our arguments, we will invoke a comparison between the sequential discovery procedure and a (Galton–Watson) **branching process**. Such a branching process (see, e.g., Harris [7] or Asmussen and Hering [1]) begins with m_0 individuals. Each of these begets, independently of the others, a number of offspring, which has some given distribution f on the non-negative integers. Each of these children then has a number of children for themselves, again independent with distribution f. And so on, again ina BFS manner. One of two things will happen: either the branching process dies out after a finite number of generations, or it survives (forever). Excluding the trivial case where f puts unit mass on 1, it is well known that the branching process has positive probability of surviving if and only if f's first moment is strictly greater than 1.

We shall be particularly concerned with a branching process whose offspring distribution is the binomial distribution $\mathrm{Bin}(2,p)$. This can be compared to the sequential discovery procedure for $c = 2$ in the following way. Suppose that we can show that up until some given stage S of the sequential discovery procedure, each choice of a new node to connect to has probability at least p (conditionally on everything seen so far) of being a success. Then we can make a joint construction (a so-called coupling; see [9]) of the sequential discovery procedure and the $\mathrm{Bin}(2,p)$ branching process in such a way that each individual in the branching process corresponds to a particular node (not shared by any of the other individuals of the branching process) of the sequential procedure, up until the given stage S. We say in this case that the sequential procedure up until stage S **stochastically dominates** the branching processs (for stochastic domination see for example [9]). We will show that the sequential discovery procedure first generates almost surely a set of $\log n$ points and that from then on each point u begets offsprings with distribution $\mathrm{Bin}(2, p_u)$, with $p_u \geq \frac{3}{4}$. It follows from a standard application of stochastic domination that the survival probability of the sequential discovery procedure is at least that of $\log n$ independent branching processes with distribution $\mathrm{Bin}(2, \frac{3}{4})$.

Proof of Proposition 3: We prove the result for $c = 2$ only, which is obviously enough since adding edges is not going to destroy a giant component.

Run the sequential discovery procedure until the outgoing edges of $\log(n)$ nodes are investigated (or until there are no more fresh nodes, in which case we are stuck).

By Lemma 7, we may assume that the event in (2.7) happens, and condition on that event. By the choice (2.6) of k, this means that each time a node selects another node to connect to, there are at least $\frac{n}{2k^2}$ nodes to choose from. And each time, there are at most $2\log(n)$ nodes that have already been seen, so each edge has probability at most

$$\frac{4k^2 \log(n)}{n} \tag{2.8}$$

of hitting a node that has already been seen. Hence, the probability that *at least one* of the $2\log(n)$ choices is a failure, is at most

$$2\log(n)\frac{4k^2\log(n)}{n} = \frac{8k^2(\log(n))^2}{n}, \tag{2.9}$$

which tends to 0 as $n \to \infty$. Hence, the probability that *all* choices, up until the outgoing edges of $\log(n)$ nodes are investigated, are successful, tends to 1 as $n \to \infty$.

Hence, we have shown that with probability approaching 1 as $n \to \infty$, we get a connected component with at least $2\log(n)$ nodes. But this is not enough to prove Proposition 3, which asserts a component whose size is *linear* in n.

We can, however, continue the sequential discovery procedure from the $\log(n)$ fresh nodes that we have (assuming that all choices so far have been succesful). Let us continue the sequential procedure until the stage S when either a total of $\frac{n}{8k^2}$ nodes have been found (or no fresh nodes remain). Before stage S, each new discovery has, by an analogous argument as that used to establish (2.8), probability at most

$$\frac{n/8k^2}{n/2k^2} = \frac{1}{4}$$

of not being successful. It follows that the sequential discovery procedure starting from the $\log(n)$ nodes until stage S stochastically dominates a $\mathrm{Bin}(2, \frac{3}{4})$ branching process with the same initial number of individuals. We therefore get, conditionally on no failures associated with the first $2\log(n)$ nodes,

$$\begin{aligned}
&\mathbf{P}(\text{the sequential procedure fails to survive until } \tfrac{n}{8k^2} \text{ nodes are found}) \\
&\quad\leq \mathbf{P}(\text{a Bin}(2, \tfrac{3}{4}) \text{ branching process starting with } \log(n) \text{ individuals dies out}) \\
&\quad= (\mathbf{P}(\text{a Bin}(2, \tfrac{3}{4}) \text{ branching process starting with 1 individual dies out}))^{\log(n)} \\
&\quad= (1-\alpha)^{\log n}
\end{aligned} \tag{2.10}$$

where $\alpha > 0$ is the survival probability of a $\mathrm{Bin}(2, \frac{3}{4})$ branching process starting from a single individual (an easy calculation shows that $\alpha = \frac{8}{9}$, but we only need the fact that $\alpha > 0$, which follows from the fact that the offspring distribution has expectation $\frac{3}{2} > 1$). The sum of (2.9) and (2.10) tends to 0 as $n \to \infty$, whence (1.2) follows with $s = \frac{1}{8k^2}$, and we are done. $\square$

Note that by Lemma 7, (2.9) and (2.10) the probability of not having a giant component is $\Theta(n^{-\epsilon})$ for $\epsilon > 0$.

3 Bluetooth: connectedness

In this section we go on to prove the connectedness of $G^n_{r,c}$ asserted in Theorem 4. We begin by proving the following strengthening of Proposition 3.

Proposition 8 *Fix $r > 0$ and $c \geq 2$. Then there exists a constant $s > 0$ such that*

$$\lim_{n\to\infty} \mathbf{P}(\textit{every node of } G^n_{r,c} \textit{ is in some } s\textit{-giant component}) = 1 \,.$$

Proof: Again, it suffices to consider $c = 2$. As in the previous section, let α denote the survival probability of a $\text{Bin}(2, \frac{3}{4})$ branching process starting from a single individual.

We proceed using the sequential discovery procedure as in the proof of Proposition 3, with the following modification. Instead of initially running it until the outgoing edges of $\log(n)$ nodes have been checked, run it until the outgoing edges of $a\log(n)$ nodes have been checked, where a is a fixed number chosen so that

$$a > \log\left(\frac{1}{1-\alpha}\right) .$$

The estimate in (2.9) then becomes replaced by $\frac{8k^2a^2(\log(n))^2}{n}$. However, since the result we are trying to prove concerns all n points simultaneously, we need to improve on this estimate (which, when multiplied by n, fails to approach 0). To do this, note we can afford to have *one* failed edge during the discovery of the outgoing edges of the first $a\log(n)$ nodes without very much damage (there will still be $a\log(n)$ fresh edges at the end of this search). To estimate the probability that at least two choices fail, note that there are less than $\frac{(a\log(n))^2}{2}$ pairs of times during the procedure at which the choices can fail, and for each such pair the probability of failure in both is at most $\left(\frac{2a\log(n)}{n/2k^2}\right)^2$ (assuming as before the event in Lemma 7). The probability that at least two of the $2a\log(n)$ choices are failures is therefore at most

$$\frac{(a\log(n))^2}{2}\left(\frac{2a\log(n)}{n/2k^2}\right)^2 = \frac{8k^4a^4(\log(n))^4}{n^2}\,, \tag{3.11}$$

which tends to 0 at a rate which (as we shall see) is fast enough for our purposes.

Again imitating the proof of Proposition 3, we go on to run the sequential discovery procedure until a total of $\frac{n}{8k^2}$ nodes have been found. Since we begin with $a\log(n)$ fresh nodes, the analogue of (2.10) becomes

$$\begin{aligned} \mathbf{P}(\text{the sequential procedure fails to survive until } \tfrac{n}{8k^2} \text{ nodes are found}) &\leq (1-\alpha)^{a\log(n)} \\ &= n^{-b} \qquad (3.12) \end{aligned}$$

where $b = -a\log(1-\alpha)$, and $b > 1$ by the choice of a.

On the event in Lemma 7 (whose probability tends to 1), we can bound the probability that *some* node fails to sit in an s-giant component (with $s = \frac{1}{8k^2}$) by adding the estimates in (3.11) and (3.12) and multiplying by the number of nodes n. This yields

$$\frac{8k^4a^4(\log(n))^4}{n} + n^{1-b} \tag{3.13}$$

which still tends to 0 as $n \to \infty$, so the proof is complete. $\square$

Proof of Theorem 4: As usual, we need only consider the $c = 2$ case. Note that in view of Proposition 8 with the estimate

$$s = \frac{1}{8k^2} \tag{3.14}$$

that comes out of its proof, the only thing that can cause (1.3) to go wrong is if there exists an $\varepsilon > 0$ such that

$$\limsup_{n\to\infty} \mathbf{P}(G^n_{r,c} \text{ contains at least two distinct } \tfrac{1}{8k^2}\text{-giant components}) \geq \varepsilon\,. \tag{3.15}$$

Now consider the experiment of generating $G^n_{r,c}$ and then picking two of its nodes at random; let A denote the event that these two nodes end up in the same connected component. By conditioning on the first of these nodes, we see that (3.15) implies that

$$\limsup_{n\to\infty} \mathbf{P}(\neg A) \;\geq\; \frac{\varepsilon}{8k^2}\,.$$

In order to prove the theorem, it therefore suffices to show that

$$\lim_{n\to\infty} \mathbf{P}(\neg A) \;=\; 0\,. \tag{3.16}$$

Let us denote the two nodes chosen at random by v_0 and v_1. By Proposition 8 and the estimate (3.14), we may assume that v_0 is in a connected component of at size least $\frac{n}{8k^2}$. Then, by the pigeonhole principle, at least one of the k^2 subsquares of $[0,1]^2$ introduced in Section 2 contains at least $\frac{n}{8k^4}$ nodes of that connected component. Let us pick such a subsquare and denote it by B.

Next, fix an integer m, and run the sequential discovery procedure starting from the other node v_1, with the following restriction. As soon as an edge fails to lead to a new node, we give up. Assuming this does not happen, we run the procedure until the outgoing edges of exactly $m-1$ nodes have been investigated; this leaves us with exactly m fresh nodes. Let $w_1, \ldots w_m$ denote the fresh nodes after having checked the outgoing edges of $m-1$ nodes in the sequential procedure. Pick one of these vertices, w_i, and denote the subsquare it sits in by $B_{i,0}$. We can then find a sequence of subsquares $B_{i,1}, B_{i,2}, \ldots, B_{i,\ell}$, $\ell \leq 2k$, such that

(i) for each $j \in \{0, 1, \ldots, \ell-1\}$, the subsquares $B_{i,j}$ and $B_{i,j+1}$ are adjacent, and

(ii) $B_{i,\ell} = B$.

Fix such a sequence, and consider the "naked-branch" sequential discovery procedure starting from w_i, and denote the nodes found along this branch by $w_{i,1}, w_{i,2}, \ldots, w_{i,\ell}$. Given the event in Lemma 7 (which we may assume happens), the probability that $w_{i,1}$ ends up in $B_{i,1}$ is at least $\frac{n/2k^2}{n} = \frac{1}{2k^2}$ (due to our choice (2.6) of k). Given that, the conditional probability that $w_{i,2}$ ends up in $B_{i,2}$ is at least $\frac{1}{2k^2}$. And so on. Finally, given that $w_{i,\ell-1}$ is in $B_{i,\ell-1}$, the conditional probability that $w_{i,\ell}$ is in the connected component of v_1, is at least $\frac{n/8k^4}{n} = \frac{1}{8k^4}$. Multiplying these conditional probabilities yields that $w_{i,\ell}$ has probability at least

$$\left(\frac{1}{2k^2}\right)^{\ell-1} \frac{1}{8k^4} \geq \left(\frac{1}{2k^2}\right)^{2k-1} \frac{1}{8k^4} \tag{3.17}$$

of being in the connected component of v_0.

On the event that no checked edges result in failures (which we assume), the m different "naked branches" move independently, so (3.17) implies that the probability that none of them hit the connected component of v_0 is at most

$$\left(1 - \left(\frac{1}{2k-1}\right)^{k-1} \frac{1}{8k^4}\right)^m .$$

We have thus shown that

$$\limsup_{n\to\infty} \mathbf{P}(\neg A) \leq \left(1 - \left(\frac{1}{2k^2}\right)^{2k-1} \frac{1}{8k^4}\right)^m . \tag{3.18}$$

Now, m was arbitrary, and the right hand side of (3.18) can be made as small as we wish by picking m large. Hence (3.16) is established and the proof is complete. □

A careful examination of the estimates of failure probabilities in the proof above show that the probability of not having connectivity is at most $\Theta(n^{-\epsilon})$, for $\epsilon > 0$.

4 Thinned Bluetooth

In this section, we go on to consider the thinned Bluetooth graph, and in particular to prove Theorem 6. We begin with the easy part.

Proof of Theorem 6, part (1.5): First construct the (non-thinned) Bluetooth graph $G^n_{r,c}$. This graph contains at most cn edges, so the nodes have on average degree at most $2c$. It follows that at least $\frac{n}{2}$ nodes have degree at most $4c$. Among $\frac{n}{2}$ such nodes, we can find a subset A of size at least $\frac{n}{2(4c+1)}$, such that no two of the nodes in A share an edge.

When we now delete edges from $G^n_{r,c}$ to obtain the thinned Bluetooth graph $G^n_{r,c,p}$, each node in A gets all its edges deleted with probability at least $(1-p)^{4c}$, independently for different nodes in A. Hence, the probability that *at least one* node in A becomes isolated in $G^n_{r,c,p}$ is at least

$$1 - \left(1 - (1-p)^{4c}\right)^{\frac{n}{2(4c+1)}} ,$$

which tends to 1 as $n \to \infty$, and (1.5) follows. □

Proof of Theorem 6, part (1.4): Fix an arbitrary $\varepsilon > 0$, and let $p' = \frac{1/c+p}{2}$. Let α' denote the survival probability of a $\mathrm{Bin}(c, p')$ branching process starting from a single individual, and note that $\alpha' > 0$ since $\mathrm{Bin}(c, p')$ has first moment $cp' > 1$. Choose an integer a in such a way that

$$(1-\alpha')^{c^a} < \varepsilon ,$$

and also an integer b such that

$$\left(1 - (p')^{(c+c^2+\cdots+c^a)}\right)^b < \varepsilon .$$

Finally, define the positive number d as

$$d = \frac{p - \frac{1}{c}}{4k^2} .$$

Now start a sequential discovery procedure beginning with b randomly chosen initial nodes, running for a generations. Each node can potentially connect to c others, but the connection is considered a failure if either the edge is deleted in the thinning step of Definition 5, or the node discovered has already been seen in the sequential procedure.

During this part of the sequential discovery procedure, the probability of failure of an edge is bounded by $(1-p)$ (for the event of being removed in the thinning) plus $\frac{2k^2b(1+c++c^2...+c^a)}{n}$ (for the event of encountering a vertex that has already been seen). We may suppose that n is large enough so that

$$p - \frac{2k^2b(1+c++c^2\ldots+c^a)}{n} \geq p'$$

so that (assuming as usual the event in Lemma 7) the probability of success of an edge remains at least p'.

Hence, each one of the b initial nodes has probability of seeing only successes in the first a generations of its part of the sequential discovery process is at least

$$(p')^{(c+c^2+\cdots+c^a)},$$

and the event D that *at least one* of the b initial nodes yields a such completely successful generations, has probability at least

$$\begin{aligned} \mathbf{P}(D) &\geq 1-\left(1-(p')^{(c+c^2+\cdots+c^a)}\right)^b \\ &> 1-\varepsilon \end{aligned} \tag{4.19}$$

where the last inequality is due to the choice of b.

On the event D, pick one of those spectacularly successful initial nodes at random, and continue the sequential discovery procedure from the c^a fresh nodes of its connected component until a total of dn nodes have been encountered (or no fresh nodes remain). The success probability of each new edge during this process is at least

$$p - \frac{dn}{n/2k^2}$$

which, by the choice of d, equals p'. Hence, we can make the by now familiar comparison with a branching process – in this case a branching process with $\text{Bin}(c,p')$ offspring distribution. We get

$$\begin{aligned} &\mathbf{P}(\text{the sequential procedure fails to survive until } dn \text{ nodes are found}) \\ &\quad\leq \mathbf{P}(\text{a Bin}(c,p') \text{ branching process starting with } c^a \text{ individuals dies out}) \\ &\quad= (\mathbf{P}(\text{a Bin}(c,p') \text{ branching process starting with 1 individual dies out}))^{c^a} \\ &\quad= (1-\alpha')^{c^a} \\ &\quad< \varepsilon\,, \end{aligned} \tag{4.20}$$

where the last inequality is due to the choice of a. Combining the results of (4.19) and (4.20), we get that

$$\limsup_{n\to\infty} \mathbf{P}\left(G^n_{r,c,p} \text{ fails to contain connected component of size at least } dn-(b-1)\sum_{i=0}^{a} c^i\right) < 2\varepsilon\,.$$

Hence

$$\limsup_{n\to\infty} \mathbf{P}\left(G^n_{r,c,p} \text{ fails to contain connected component of size at least } \frac{dn}{2}\right) < 2\varepsilon\,,$$

and since $\varepsilon > 0$ was arbitrary, we have established (1.4) with $s = \frac{d}{2}$. □

5 Final remarks

There are a number of possible ways to proceed from where we are, in order to produce further results. We mention some of these:

1. One may note that the only place in which we have used the fact that the positions of the n nodes are independent and uniformly distributed on $[0,1]^2$, is in the proof of Lemma 7. Hence, one may replace the chosen model for placement of the nodes by any other model which yields a sufficiently even distribution of nodes to obtain Lemma 7 or something along those lines. A large number of models for which this may be attempted can be found, e.g., in [13].
2. Likewise, there is scope for great flexibility in replacing the region $[0,1]^2$ in which the nodes are distributed, by others. One could even move to other dimensions; this is a mathematically natural extension but perhaps less interesting for the engineering application. For instance, the proofs of all our main results extend easily to the case where $[0,1]^2$ is replaced by $[0,1]^d$ for $d = 1$ or any $d \geq 3$.
3. The condition $p > \frac{1}{c}$ in Theorem 6 is probably not sharp.
4. One could consider other kinds of asymptotics, where r and c are still kept fixed, but the size of the region grows as the number of points n increases. As an extreme case, one could take the region to be all of $\mathbf{R}^2$, and points to be distributed according to a homogeneous Poisson process with intensity λ. This would take us into the realm of percolation theory, where the number one question would be to ask for the existence of an *infinite* connected component. Fixing r and c, one would expect the probability of existence of such an infinite component to depend on λ in such a way that it is 0 for λ less than some critical value λ_c, and 1 for $\lambda > \lambda_c$. Somewhat counterintuitively, it seems difficult to prove the monotonicity in λ required for such a statement (adding nodes may not always improve connectivity).

References

[1] S. Asmussen and H. Hering, *Branching Processes*, Birkhäuser, Boston, 1983.

[2] S. Banerjee, C. Kommareddy, K. Kar, B. Bhattacharjee and S. Khuller, *Construction of an efficient overlay multicast infrastructure for real-time applications*, Proceedings of Infocom 2003

[3] F. Ferraguto, G. Mambrini, A. Panconesi and C. Petrioli, *Blue Pleiades, a new solution for device discovery and scatternet formation in multi-hop Bluetooth networks*, submitted to ACM Winet.

[4] Feng Xue and P. R. Kumar, *The number of neighbors needed for connectivity of wireless networks.* To appear in Wireless Networks.

[5] M. Franceschetti, L. Booth, M. Cook, R. Meester, and J. Bruck, *Percolation of multi-hop wireless networks*, preprint, `http://www.cs.vu.nl/~rmeester/pre.html`

[6] O. Häggström and R. Meester, "Nearest neighbour and hard sphere models in continuum percolation", *Random Structures and Algorithms*, 9:3, pp. 295–315, 1996.

[7] T.E. Harris, *The Theory of Branching Processes*, Springer, Berlin, 1963.

[8] J.Könemann *Approximation Algorithms for Minimum-Cost Low-Degree Subgraphs.* PhD Thesis in Algorithms, Combinatorics, and Optimization, Carnegie Mellon University, 2003.

[9] T. Lindvall, *Lectures on the Coupling Method*, Wiley, New York, 1992.

[10] R. Meester, M. Penrose and A. Sarkar, "The random connection model in high dimensions", *Statistics and Probability Letters*, 35:2, pp. 145–153, 1997.

[11] R. Meester and R. Roy, *Continuum Percolation*, Cambridge University Press, 1996.

[12] M. Penrose, "On the spread-out limit for bond and continuum percolation", *Annals of Applied Probability* 3:1, pp. 253–276, 1993.

[13] D. Stoyan, W.S. Kendall and J. Mecke, *Stochastic Geometry and its Applications*, 2nd ed., Wiley, New York, 1995.