Connectivity threshold of Bluetooth graphs

Nicolas Broutin
INRIA

Luc Devroye
McGill University *

Nicolas Fraiman
McGill University

Gábor Lugosi
ICREA and Pompeu Fabra University †

May 31, 2012

Abstract

We study the connectivity properties of random Bluetooth graphs that model certain "ad hoc" wireless networks. The graphs are obtained as "irrigation subgraphs" of the well-known random geometric graph model. There are two parameters that control the model: the radius r that determines the "visible neighbors" of each vertex and the number of edges c that each vertex is allowed to send to these. The randomness comes from the underlying distribution of vertices in space and from the choices of each vertex. We prove that no connectivity can take place with high probability for a range of parameters r, c and completely characterize the connectivity threshold (in c) for values of r close the critical value for connectivity in the underlying random geometric graph.

1 Introduction

It is sometimes necessary to *sparsify* a network: given a connected graph, one wants to extract a sparser yet connected subgraph. In general, the protocol should be distributed, in that it should not involve any global optimization or coordination for obvious scaling reasons. The problem arises for instance in the formation of Bluetooth ad-hoc or sensor networks [24], but also in settings related to information dissemination (broadcast or rumour spreading) [4, 8].

In this paper, we consider the following simple and distributed algorithm for graph sparsification. Let $G_n = (V, E)$ be a finite undirected graph on $|V| = n$ vertices and edge set E. A *random irrigation subgraph* $S_n = (V, \widehat{E})$ of G_n is obtained as follows: Let $2 \leq c_n < n$ be a positive integer. For every vertex $v \in V$, we pick randomly and independently, without replacement, c_n edges from E, each adjacent to v. These edges form the set of edges $\widehat{E} \subset E$ of the graph S_n (if the degree of v in G_n is less than c_n, all edges adjacent to v belong to $\widehat{E}$). The main question is how large c_n needs to be so that the graph S_n is connected, with high probability. Naturally, the answer depends on what the underlying graph G_n is.

When $G_n = K_n$ is the complete graph then for constant $c_n = c \geq 2$, Fenner and Frieze [11] proved that S_n is c-connected (for both vertex- and edge-connectedness) with high probability.

*The second author's research was sponsored by NSERC Grant A3456 and FQRNT Grant 90-ER-0291.

†The fourth author acknowledges support by the Spanish Ministry of Science and Technology grant MTM2009-09063 and by the PASCAL Network of Excellence under EC grant no. 506778.

This model is also known as the random c-out graph. In a subsequent paper, Fenner and Frieze [12] considered the probability of existence of a Hamiltonian cycle. They showed that there exists $c \le 23$ such that a Hamiltonian cycle exists with probability tending to 1 as n tends to infinity. In a recent article Bohman and Frieze [1] proved that $c = 3$ suffices.

Apart from the complete graph, the most extensively studied case, and arguably the most important for applications, is when $G_n = G_n(r_n)$ is a *random geometric graph* defined as follows: Let $X_1, \ldots, X_n$ be independent, uniformly distributed random points in the unit cube $[0,1]^d$. The set of vertices of the graph $G_n(r_n)$ is $V = \{1, \ldots, n\}$ while two vertices i and j are connected by an edge if and only if the Euclidean distance between X_i and X_j does not exceed a positive parameter r_n, i.e., $E = \{(i,j) : \|X_i - X_j\| < r_n\}$ where $\|\cdot\|$ denotes the Euclidean norm. Many properties of $G_n(r_n)$ are well understood. We refer to the monograph of Penrose [21] for a survey. The graph $S_n = S_n(r_n, c_n)$ was introduced in the context of the Bluetooth network [24], and is sometimes called the *Bluetooth* or *scatternet graph* with parameters n, r_n, and c_n. The model was introduced and studied in [6, 10, 13, 19, 22].

We are interested in the behavior of the graph $S_n(r_n, c_n)$ for large values of n. When we say that a property of the graph holds *with high probability* (whp), we mean that the probability that the property does not hold is bounded by a function of n that goes to zero as $n \to \infty$. Equivalently we say that a sequence of random events E_n occurs with high probability if $\lim_{n\to\infty} \mathbf{P}\{E_n\} = 1$. There are two independent sources of randomness in the definition of the random graph $S_n(r_n, c_n)$. One comes from the random underlying geometric graph $G_n(r_n)$ and the other from the choice of the c_n neighbors of each vertex.

Since we are interested in connectivity of $S_n(r_n, c_n)$, a minimal requirement is that $G_n(r_n)$ should be connected. It is well known that the connectivity threshold of $G_n(r_n)$ is $\gamma^* \sqrt[d]{\log n/n}$ where $\gamma^* = 2\sqrt[d]{1/(2d\theta_d)}$, where $\theta_d = \mathrm{Vol}\, B(0,1)$ is the Lebesgue measure of the unit ball in $\mathbb{R}^d$. See [15, 20] or Theorem 13.2 in [21]. This means that G_n is connected with high probability if r_n is at least $\gamma \sqrt[d]{\log n/n}$ where $\gamma > \gamma^*$ while G_n is disconnected with high probability if r_n is less than $\gamma \sqrt[d]{\log n/n}$ where now $\gamma < \gamma^*$. We always consider values of r_n above this level.

When $r_n = r$ is constant, the geometry has very little influence: For instance, Dubhashi, Johansson, Häggström, Panconesi, and Sozio [9] showed that when $r_n = r$ is independent of n, $S_n(r, 2)$ is connected with high probability. The case when r_n is small is a more delicate issue, since the geometry now plays a crucial role. Crescenzi, Nocentini, Pietracaprina, and Pucci [6] proved that in dimension $d = 2$ there exist constants γ_1, γ_2 such that if $r_n \ge \gamma_1 \sqrt{\log n/n}$ and $c_n \ge \gamma_2 \log(1/r_n)$, then $S_n(r_n, c_n)$ is connected with high probability.

Arguably the most interesting values for r_n are those just above the connectivity threshold for the underlying graph $G_n(r_n)$, that is, when r_n is proportional to $\sqrt[d]{\log n/n}$. The results of Crescenzi et al. [6] show that for such values of r_n, connectivity of $S_n(r_n, c_n)$ is guaranteed, with high probability, when c_n is a sufficiently large constant multiple of $\log n$. In this paper we show that this bound can be improved substantially. For the given choice of r_n, there is a critical c_n for connectivity. It is quite easy to show that no connectivity can take place (whp) for constant c_n, and that for $c_n \ge \lambda \log n$ for a sufficiently large λ, the graph is connected whp (because the maximal cardinality of any ball of radius r is whp $O(\log n)$). The objective of this paper is to nail down the precise threshold. Our main result is the following theorem.

Theorem 1. *There exists a finite constant γ^{**}, depending on d only, such that for all $\gamma \ge \gamma^{**}$, and with*

$$r_n = \gamma \left(\frac{\log n}{n}\right)^{1/d},$$

we have, for all $\varepsilon \in (0,2)$,

$$\lim_{n\to\infty} \mathbf{P}\left\{S_n(r_n,c_n) \text{ is connected}\right\} = \begin{cases} 1 & \textit{if } c_n \geq \sqrt{\frac{(2+\varepsilon)\log n}{\log\log n}}, \\ 0 & \textit{if } c_n \leq \sqrt{\frac{(2-\varepsilon)\log n}{\log\log n}}. \end{cases}$$

The proof shows that connectivity occurs at the same threshold for the presence of $(c+1)$-cliques. It might be a bit surprising that the threshold is virtually independent of γ. The threshold (in c_n) is also independent of the dimension d. This is probably less surprising since c_n counts a number of neighbors and the number of visible vertices in a ball is of order $\log n$, independently of d, for the range of r_n we consider.

The structure of the paper is the following: In Section 2 we prove a lower bound on the critical value of c_n needed to obtain a connected graph whp given a value of r_n in the range where connectivity could be achieved. In Section 3 we show that $S_n(r_n,c_n)$ is connected whp where r_n is proportional to $\sqrt[d]{\log n/n}$ and c_n is just above the corresponding value obtained in Section 2 nailing down the precise threshold in that case. Finally in Section 4 we obtain an upper bound on the diameter of $S_n(r_n,c_n)$ for the same values of r_n as in Section 3 but with a slightly larger value of c_n. In particular, we show that if c_n is a sufficiently large constant times $\sqrt{\log n}$ then the diameter of $S_n(r_n,c_n)$ is $O(1/r_n)$ which is the same order of magnitude as for the underlying random geometric graph.

A final notational remark: To ease the reading for the rest of the paper we omit the subscript n in the parameters r and c as well as in most of the events and sets we define that depend on n.

2 A lower bound for connectivity on the whole range

The aim of this section is to prove a lower bound on the value of c needed to obtain connectivity whp for a given value of r. First we need a lemma on the regularity of uniformly distributed points. Let $N(A) = \sum_{i=1}^n \mathbf{1}_{[X_i \in A]}$ be the number of vertices in a set $A \subset [0,1]^d$. We consider $\gamma^{**} > \gamma^*$ to provide a sufficient margin of play. It is an interesting problem to consider smaller values of γ. We expect that the results also hold for that case. However, the methods that we use don't allow us to go closer to the critical radius for connectivity.

Lemma 1 (Ball density Lemma)**.** *Let* $\gamma^{**} = 4\sqrt[d]{2/\theta_d}$*, where* θ_d *is the volume of the unit ball in* $\mathbb{R}^d$*. Then for each* $\gamma > \gamma^{**}$*, there exist constants* $0 < \alpha_\bigcirc < \beta_\bigcirc < \infty$ *such that the following event, which we denote by* $D_\bigcirc$*, occurs whp:*

$$\begin{array}{rcccll} \alpha_\bigcirc n r^d & \leq & N(B(X_i,r)) & \leq & \beta_\bigcirc n r^d, & \textit{and} \\ 2^d\alpha_\bigcirc n r^d & \leq & N(B(X_i,2r)) & \leq & 2^d\beta_\bigcirc n r^d, & \textit{and} \\ 2^{-d}\alpha_\bigcirc n r^d & \leq & N(B(X_i,r/2)) & \leq & 2^{-d}\beta_\bigcirc n r^d & \end{array}$$

for every $1 \leq i \leq n$.

Proof. We use the binomial Chernoff bound: If $\xi \sim \text{Binomial}(n,p)$ and $t > 0$ then

$$\min\Big(\mathbf{P}\left\{\xi \leq tnp\right\}, \mathbf{P}\left\{\xi \geq tnp\right\}\Big) \leq \exp\left(t - np - t\log\left(\frac{t}{np}\right)\right) = \exp\Big(f(t)np\Big),$$

where we write $f(x) = x - 1 - x\log x$, for reference see [5, 17].

The expected cardinality of the set of vertices in a ball $B(x,r)$ is $\theta(x)nr^d$, where $\theta(x) \in [\theta_d/2^d, \theta_d]$ takes care of the border effect and θ_d is the volume of the unit ball in dimension d. Therefore the number of vertices $N(B(X_i,r))$ is stochastically between $\zeta_1 \sim \text{Binomial}(n, \theta_d r^d/2^d)$ and $\zeta_2 \sim \text{Binomial}(n, \theta_d r^d)$. Thus, we have for any $1 \le i \le n$

$$\mathbf{P}\left\{N(B(X_i,r)) \le \alpha_{\bigcirc} nr^d\right\} \le \mathbf{P}\left\{\zeta_1 \le \alpha_{\bigcirc} nr^d\right\} \le \exp\left(f\left(\frac{2^d\alpha_{\bigcirc}}{\theta_d}\right)\frac{\theta_d nr^d}{2^d}\right),$$
$$\mathbf{P}\left\{N(B(X_i,r)) \ge \beta_{\bigcirc} nr^d\right\} \le \mathbf{P}\left\{\zeta_2 \ge \beta_{\bigcirc} nr^d\right\} \le \exp\left(f\left(\frac{\beta_{\bigcirc}}{\theta_d}\right)\theta_d nr^d\right).$$

We choose $\alpha_{\bigcirc} < \theta_d/2^d$ so that $f(2^d\alpha_{\bigcirc}/\theta_d) = -1/2$ and $\beta_{\bigcirc} > \theta_d$ so that $f(\beta_{\bigcirc}/\theta_d) = -1/2$. Define the event $D_i = \{N(B(X_i,r)) \in [\alpha_{\bigcirc} nr^d, \beta_{\bigcirc} nr^d]\}$. We can apply a union bound to obtain

$$\mathbf{P}\left\{\bigcup_{i=1}^n D_i^c\right\} \le \sum_{i=1}^n \mathbf{P}\left\{D_i^c\right\} \le \sum_{i=1}^n 2\exp\left(-\frac{\theta_d nr^d}{2\cdot 2^d}\right) \le 2n\exp\left(-\frac{\theta_d\gamma^d}{2\cdot 2^d}\log n\right) \to 0,$$

if $\gamma > 2\sqrt[d]{2/\theta_d}$. Repeating the argument for balls of radius $2r$ and $r/2$ we need $\gamma > \gamma^{**}$ where $\gamma^{**} = 4\sqrt[d]{2/\theta_d}$. $\square$

The next theorem shows that for any value of r above the connectivity threshold of the random geometric graph one cannot hope that S_n is connected unless c is at least of the order of $\sqrt{\log n/\log(nr^d)}$. In particular, when r is just above the threshold (i.e., it is proportional to $\sqrt[d]{\log n/n}$) then c must be at least of the order of $\sqrt{\log n/\log\log n}$. We say that the vertices at distance less than r from X_i are the *visible neighbors* of i (i.e., the neighbors of i in G_n) and that $B(X_i,r)$ is the *visibility ball* of i. Note that the following result implies the lower bound of Theorem 1.

Theorem 2. *Let $\varepsilon \in (0,1)$ and $\lambda \in [1,\infty]$ be such that*

$$\gamma^{**}\left(\frac{\log n}{n}\right)^{1/d} < r < 1, \qquad \frac{\log nr^d}{\log\log n} \to \lambda \qquad \text{and} \qquad c = \left\lfloor\sqrt{(1-\varepsilon)\left(\frac{\lambda}{\lambda-1/2}\right)\frac{\log n}{\log nr^d}}\right\rfloor.$$

Then $S_n(r,c)$ is not connected whp. (In the case of $\lambda = \infty$, we define $\lambda/(\lambda-1/2) = 1$.)

Note that in the range of r considered, we do have $\lambda \ge 1$.

Proof. Note that we can assume $\log nr^d < (1-\varepsilon)\log n$ otherwise $c = 0$ so every vertex is isolated, the graph is disconnected and there is nothing to prove. We will use this fact at the end of the proof. We show that there exists an isolated $(c+1)$-clique whp. The proof is an application of the second moment method. Let $\mathcal{F}$ be the random family of subsets of $\{1,\ldots,n\}$ given by

$$\mathcal{F} = \left\{Q \subset \{1,\ldots,n\} : |Q| = c+1, \|X_i - X_j\| < r \;\; \forall i,j \in Q\right\}.$$

Denote by $I(Q)$ the indicator of the event that the vertices in Q form an isolated clique in S_n. Then $N = \sum_{Q\in\mathcal{F}} I(Q)$ is the number of isolated $(c+1)$-cliques. First we condition on all the vertices $X_1,\ldots,X_n$. The only randomness we consider are the choices of each vertex among their visible neighbors. Let $D_{\bigcirc}$ be the event described in Lemma 1 which holds whp. In the following we work conditionally on $X_1,\ldots,X_n$ assuming $D_{\bigcirc}$ holds. Throughout, we use several auxiliary functions ϕ_i with the property that $\phi_i(n) = o(\log n)$ for all $i = 1,\ldots,8$.

Define $I_1(Q)$ as the indicator for the event that no vertex $j \in Q$ chooses to link to a vertex $i \notin Q$, and $I_2(Q)$ as the indicator for the event that every $i \notin Q$ avoids choosing the vertices in Q as an endpoint of any of its c links. Clearly $I(Q) = I_1(Q)I_2(Q)$. Furthermore, conditionally on $X_1, \ldots, X_n$, the variables $I_1(Q)$ and $I_2(Q)$ are independent (because they involve the choices of disjoint sets of indices). On $D_\bigcirc$,

$$\begin{aligned}
\mathbf{E}\{I_1(Q) \mid X_1, \ldots, X_n\} &= \prod_{i\in Q} \prod_{k=0}^{c-1} \left(\frac{c-k}{N(B(X_i, r)) - k} \right) \\
&\geq \left(\prod_{k=0}^{c-1} \frac{c-k}{\beta_\bigcirc nr^d - k} \right)^{c+1} \\
&\geq \left(\frac{c!}{(\beta_\bigcirc nr^d)^c} \right)^{c+1} \\
&\geq \left(\frac{c}{e\beta_\bigcirc nr^d} \right)^{c^2+c} \\
&\geq \exp\Big((c^2+c)(\log c - \log nr^d) - \phi_1(n) \Big).
\end{aligned}$$

We can write the first equality in the display above because $c < \alpha_\bigcirc nr^d$ for all n sufficiently large. Let $J_Q = \big\{ j \in \{1, \ldots, n\} : \exists\, i \in Q,\ \|X_i - X_j\| < r \big\}$ then we also have

$$\begin{aligned}
\mathbf{E}\{I_2(Q) \mid X_1, \ldots, X_n\} &\geq \prod_{j\in J_Q} \prod_{k=0}^{c-1} \left(1 - \frac{c+1}{N(B(X_j, r)) - k} \right) \\
&\geq \left(\prod_{k=0}^{c-1} \left(1 - \frac{c+1}{\alpha_\bigcirc nr^d - k} \right) \right)^{|J_Q|} \\
&\geq \left(1 - \frac{2c}{\alpha_\bigcirc nr^d} \right)^{c|J_Q|} \\
&\geq \exp\left(2^d \beta_\bigcirc cnr^d \log\left(1 - \frac{2c}{\alpha_\bigcirc nr^d} \right) \right) \\
&\geq \exp\left(-4c^2 2^d \beta_\bigcirc / \alpha_\bigcirc \right) \\
&\geq \exp\Big(-\phi_2(n) \Big),
\end{aligned}$$

where we use the bound $|J_Q| < N(B(X_{i_0}, 2r)) < 2^d \beta_\bigcirc nr^d$ (if $j \in J_Q$ then there exists $i \in Q$ such that $\|X_{i_0} - X_j\| \leq \|X_{i_0} - X_i\| + \|X_i - X_j\| < 2r$). Also, we used the facts that $nr^d = \Omega(\log n)$ and $c^2 = O(\log n / \log nr^d) = o(\log n)$.

Moreover, on $D_\bigcirc$, we can lower bound the size of $\mathcal{F}$ by choosing i_0 and counting the sets $Q = \{i_0, i_1, \ldots, i_c\}$ such that all the points X_{i_k} are inside $B(X_{i_0}, r/2)$ since this implies that all the distances between them are less than r. Note that this counts each set $c+1$ times. So, we have

$$|\mathcal{F}| \geq \frac{n}{c+1} \binom{\lceil 2^{-d} \alpha_\bigcirc nr^d \rceil}{c} \geq \frac{n}{c+1} \left(\frac{2^{-d}\alpha_\bigcirc nr^d}{c} \right)^c \geq \exp\Big(-c(\log c - \log nr^d) + \log n - \phi_3(n) \Big).$$

Thus, the expected number of isolated $(c+1)$-cliques may be lower bounded as

$$\begin{aligned}
\mathbf{E}\left\{N \mid X_1,\ldots,X_n\right\} &= \sum_{Q\in\mathcal{F}} \mathbf{E}\left\{I(Q) \mid X_1,\ldots,X_n\right\} \\
&= \sum_{Q\in\mathcal{F}} \mathbf{E}\left\{I_1(Q) \mid X_1,\ldots,X_n\right\} \cdot \mathbf{E}\left\{I_2(Q) \mid X_1,\ldots,X_n\right\} \\
&\geq \sum_{Q\in\mathcal{F}} \exp\Big((c^2+c)(\log c - \log nr^d) - \phi_1(n) - \phi_2(n)\Big) \\
&= |\mathcal{F}| \cdot \exp\Big((c^2+c)(\log c - \log nr^d) - \phi_1(n) - \phi_2(n)\Big) \\
&\geq \exp\Big(c^2(\log c - \log nr^d) + \log n - \phi_1(n) - \phi_2(n) - \phi_3(n)\Big) \\
&= \exp\Big(c^2(\log c - \log nr^d) + \log n - \phi_4(n)\Big)
\end{aligned}$$

when $D_{\bigcirc}$ holds. Therefore, when $\lambda < \infty$,

$$\begin{aligned}
\mathbf{E}\left\{N \mid X_1,\ldots,X_n\right\} &\geq \exp\Big(c^2(\log c - \log nr^d) + \log n - \phi_4(n)\Big) \\
&= \exp\left(\frac{(1-\varepsilon)\lambda\log n}{(\lambda-1/2)\log nr^d}\left(\frac{1}{2}\log\log n - \log nr^d\right) + \log n - \phi_5(n)\right) \\
&= \exp\left(\frac{(1-\varepsilon)\lambda}{\lambda-1/2}\left(\frac{1}{2\lambda}-1\right)\log n + \log n - \phi_6(n)\right) \\
&= \exp\Big(-(1-\varepsilon)\log n + \log n - \phi_6(n)\Big) \\
&= \exp\Big(\varepsilon\log n - \phi_6(n)\Big) \to \infty,
\end{aligned}$$

since $\phi_6(n) = o(\log n)$. When $\lambda = \infty$, the proof is analogous, if we substitute $\frac{\lambda}{\lambda-1/2}$ by 1 and $\frac{1}{2\lambda}$ by 0 in the previous equation.

To finish the proof we need to upper bound the variance N to ensure that $N > 0$ with high probability. Note that if $Q \cap Q' \neq \emptyset$ and $Q \neq Q'$, then $I(Q)I(Q') = 0$ because Q and Q' cannot be isolated cliques at the same time. Now, in the case $Q \cap Q' = \emptyset$ the random variables $I_1(Q)$ and $I_1(Q')$ are independent and we obtain, for any $X_1, X_2, \ldots, X_n$ such that $D_{\bigcirc}$ holds,

$$\begin{aligned}
\mathbf{E}\left\{I(Q)I(Q') \mid X_1,\ldots,X_n\right\} &\leq \mathbf{E}\left\{I_1(Q)I_1(Q') \mid X_1,\ldots X_n\right\} \\
&\leq \mathbf{E}\left\{I_1(Q) \mid X_1,\ldots X_n\right\} \cdot \mathbf{E}\left\{I_1(Q') \mid X_1,\ldots X_n\right\} \\
&\leq \prod_{i\in Q}\prod_{k=0}^{c-1}\frac{c-k}{N(B(X_i,r))-k} \cdot \prod_{j\in Q'}\prod_{k=0}^{c-1}\frac{c-k}{N(B(X_j,r))-k} \\
&\leq \left(\prod_{k=0}^{c-1}\frac{c-k}{\alpha_{\bigcirc} nr^d - k}\right)^{2(c+1)} \\
&\leq \left(\frac{c}{\alpha_{\bigcirc} nr^d}\right)^{2(c^2+c)} \\
&\leq \exp\Big(2(c^2+c)(\log c - \log(\alpha_{\bigcirc} nr^d))\Big),
\end{aligned}$$

For the variance we have

$$\begin{aligned}
&\mathrm{Var}\{N \mid X_1,\dots,X_n\} \\
&\quad = \sum_{Q,Q'\in\mathcal{F}} \mathbf{E}\left\{I(Q)I(Q') \mid X_1,\dots,X_n\right\} - \mathbf{E}\left\{I(Q) \mid X_1,\dots,X_n\right\}\mathbf{E}\left\{I(Q') \mid X_1,\dots,X_n\right\}.
\end{aligned}$$

If the vertices in Q and Q' are far enough apart (i.e., when there is no vertex ℓ such that both $\|X_i - X_\ell\|, \|X_j - X_\ell\| < r$ for some $i \in Q$ and $j \in Q'$) then the choices involved in $I(Q)$ and $I(Q')$ are independent. Thus, we need only sum over pairs in

$$\mathcal{G} = \left\{(Q,Q') \in \mathcal{F}\times\mathcal{F} : \exists\, i_0 \in Q, j_0 \in Q' \;\; \|X_{i_0} - X_{j_0}\| < 2r\right\},$$

since all other terms vanish. Therefore,

$$\begin{aligned}
&\mathrm{Var}\{N \mid X_1,\dots,X_n\} \\
&\quad = \sum_{(Q,Q')\in\mathcal{G}} \mathbf{E}\left\{I(Q)I(Q') \mid X_1,\dots,X_n\right\} - \mathbf{E}\left\{I(Q) \mid X_1,\dots,X_n\right\}\mathbf{E}\left\{I(Q') \mid X_1,\dots,X_n\right\} \\
&\quad \le \sum_{(Q,Q')\in\mathcal{G}} \mathbf{E}\left\{I(Q)I(Q') \mid X_1,\dots,X_n\right\} \\
&\quad \le |\mathcal{G}| \cdot \exp\left(2(c^2+c)(\log c - \log(\alpha_\bigcirc n r^d))\right), \\
&\quad \le \exp\left(2c^2(\log c - \log nr^d) + (2-\varepsilon)\log n + \phi_7(n)\right), \\
&\quad = \exp\left(\varepsilon \log n + \phi_8(n)\right),
\end{aligned}$$

where we upper bound the size of $\mathcal{G}$ by choosing i_0 and j_0 and counting the sets $Q = \{i_0, i_1, \dots, i_c\}$ and $Q' = \{j_0, j_1, \dots, j_c\}$ such that all the points X_i for $i \in Q$ are inside $B(X_{i_0}, r)$ (since all of them have to be at distance r from X_{i_0}) and X_j for $j \in Q'$ are inside $B(X_{j_0}, r)$. So, on $D_\bigcirc$ we have

$$\begin{aligned}
|\mathcal{G}| &\le n\left\lceil 2^d \beta_\bigcirc n r^d\right\rceil \binom{\lceil \beta_\bigcirc n r^d\rceil}{c}^2 \\
&\le \exp\left(\log n + \log nr^d + o(\log n)\right)\left(\frac{\beta_\bigcirc n r^d}{c}\right)^{2c} \\
&\le \exp\left(-2c(\log c - \log nr^d) + (2-\varepsilon)\log n + \phi_7(n)\right).
\end{aligned}$$

The last inequality holds because $\log nr^d < (1-\varepsilon)\log n$ by the remark at the beginning of the proof. Finally, on $D_\bigcirc$, applying Chebyshev's inequality we get

$$\mathbf{P}\left\{N = 0 \mid X_1,\dots,X_n\right\} \le \frac{\mathrm{Var}\{N \mid X_1,\dots,X_n\}}{\mathbf{E}\left\{N \mid X_1 \dots X_n\right\}^2} \le \frac{\exp\left(\varepsilon\log n + \phi_8(n)\right)}{\exp\left(2\varepsilon \log n - \phi_6(n)\right)} \to 0,$$

as $n \to \infty$. This completes the proof since $D_\bigcirc$ holds whp by Lemma 1. $\square$

3 Connectivity near the critical radius

In this section we prove the remaining part of Theorem 1. We consider $r = \gamma\sqrt{\log n/n}$ with $\gamma > \gamma^{**}$. We only need to prove that S_n is connected whp when c is above the threshold since Theorem 2 implies that S_n is disconnected whp when c is below it.

Theorem 3. *Let $\varepsilon \in (0,2)$, $\gamma \geq \gamma^{**}$ and suppose that*

$$r = \gamma \left(\frac{\log n}{n}\right)^{1/d} \quad \text{and} \quad c \geq \sqrt{\frac{(2+\varepsilon)\log n}{\log\log n}},$$

Then $S_n(r,c)$ is connected whp.

We first give a high-level proof using a combinatorial argument which reduces the problem of connectivity to the occurrence of four properties that will be shown to hold in a second part.

We tile the unit cube $[0,1]^d$ into cells of side length $\lfloor 1/r \rfloor^{-1}$. A cell is interconnected and colored black if all the vertices in it are connected to each other without ever using an edge that leaves the cell. The other cells are initially colored white. Two cells are connected if they are adjacent (they share a $(d-1)$-dimensional face) and there is an edge of S_n that links a vertex in one cell to a vertex in the other cell. Two cells are $*$-connected if they share at least a corner and there is an edge of S_n binding one vertex of each cell.

Consider the following events:

(i) All cells in the grid are occupied and connected to all their neighbors. ($2d$ for cells in the inside, less than $2d$ for cells on the boundary.)

(ii) The largest $*$-connected component of white cells has cardinality at most q.

(iii) The smallest connected component of S_n is of size at least s.

(iv) Each grid cell contains at most $\lambda \log n$ vertices.

Proposition 4. *Suppose that (i)–(iv) above hold. Assuming further that q, s and λ are positive functions of n such that*

$$q = o\left(\frac{1}{r^{1-1/d}}\right) \quad \text{and} \quad \frac{s}{\lambda \log n} > q^{d/(d-1)} \ ,$$

then, for all sufficiently large n, the graph S_n is connected.

Proof. The proof uses a percolation-style argument on the grid of cells. We define a *black connector* as a connected component of black cells that links one side of the cube $[0,1]^d$ to the opposite side.

(a) There exists a black connector in the cell grid graph: Note that by a generalization of the celebrated argument of Kesten [18], either there is a black connector, or there is a white $*$-connected component of cells that prevents this connection from happening (one of the two events must occur). In dimension 2, this blocking $*$-connected component of white cells is a path that separates the two opposite faces of interest; in dimension d, the blockage must be a $(d-1)$-dimensional sheet (see also [3, 14]). In any case, the $*$-connected component of white cells, if it exists, must be of size at least r^{1-d} in order to block any black connector. Since the largest $*$-connected component of white cells has size at most q, and $q < r^{1-d}$ for n large enough, a black connector must exists. The black components of size less than $1/r$ are now recolored gray. Note that this leaves at least the black connector component, of size at least $1/r$.

(b) Next we show that all remaining black cells are connected. Note that this implies that the corresponding vertices of S_n belong to the same connected component. This collection of vertices of S_n is called the *black monster*. Assume for a contradiction that there exists two connected components of black cells that are not connected together, say K and K'. Then they must be separated by a $*$-connected component of white cells, and in particular there must exist some white cells. Now consider K, one of these two components of black cells. Let $\partial K_1, \ldots, \partial K_\ell$ be the $*$-connected components of white cells of the (vertex-) boundary of K in the grid. Each one of these boundaries separates K from one of the components of the complement of K in the grid, see Lemma 2.1 from [7]. Clearly, one and only one of $\partial K_1, \ldots, \partial K_\ell$, without loss of generality ∂K_1, suffices to separate K from K', see Figure 1.

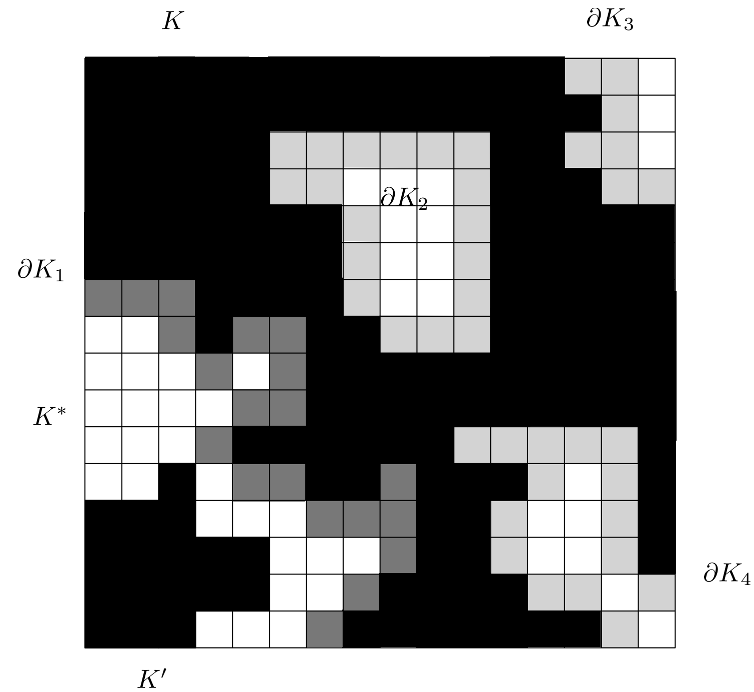

Figure 1: The two components K and K' separated by the $*$-connected piece of the boundary ∂K_1. The boundary of K is colored in gray and ∂K_1 in dark gray. The boundary of K'' is ∂K_1.

By definition, removing ∂K_1 from the grid creates some connected components of cells, one of them containing K and other containing K'. Let K'' be the one containing K'. Without loss of generality, we may assume that the size of K'' is at most $\lfloor 1/r \rfloor^d /2$ (otherwise we may replace K' by K). Note also that K'' contains at least $1/r$ cells, for K' itself contains that many cells. By the isoperimetric theorem on the *finite grid* $\{1, \ldots, \lfloor 1/r \rfloor\}^d$ due to Bollobás and Leader [2], the (vertex-) boundary of K'', $\partial K'' \subseteq \partial K_1$ (inside the finite grid) consists of at least $\Omega(r^{1/d-1})$ white cells. In particular, since ∂K_1 is $*$-connected, there exists a $*$-connected component of white cells containing at least a constant times $r^{1/d-1}$ cells. By assumption, $q = o(r^{1/d-1})$, and thus, no such separating white $*$-connected chain can exist for a sufficiently large n.

(c) Each vertex connects to at least one vertex of the black monster: To prove this, consider any vertex j, outside of the black monster, and write C for the component of S_n it belongs to. If any vertex of C lies in the black monster, then j is connected to the black monster and we are done. So we now assume that all vertices of C belong to white or gray grid cells. Adjacent vertices in C lie in the same cell, or two $*$-adjacent cells. Let K be the $*$-connected component of all grid cells visited by vertices of C. Enlarge K by adding all grid cells that reach K via a white $*$-connected chain of cells. The resulting $*$-connected component of white and gray cells is called K^*, see Figure 2.

By assumption, it contains at least $s/(\lambda \log n)$ cells, since it covers the connected component C

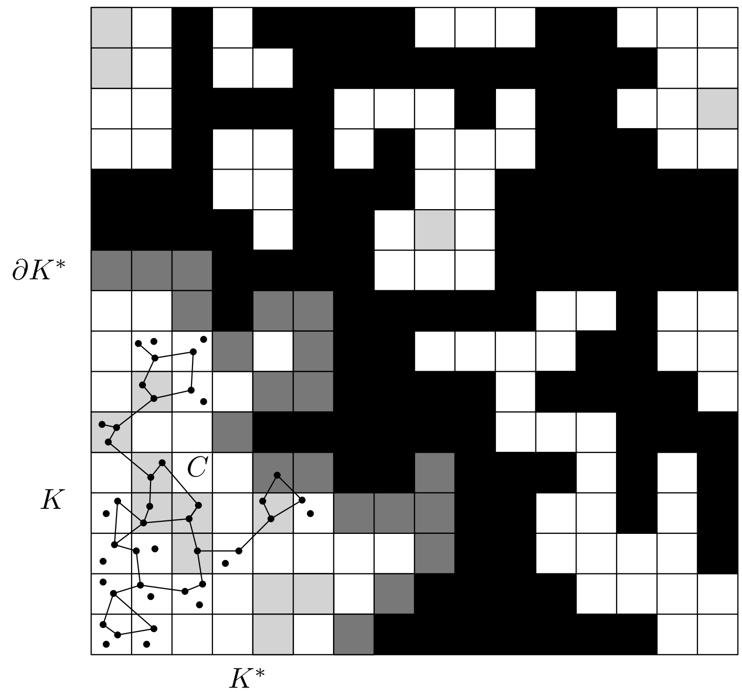

Figure 2: The component C with its corresponding $*$-connected component of occupied cells K. We enlarge K by adding all the connected white and gray cells to get K^*. The border cells of K^* are colored in dark gray.

of S_n (by properties (iii) and (iv)). So we have exhibited a fairly large $*$-connected component of cells that are not black; the only issue is that it might not be fully white, and we wish to isolate a large *white* $*$-connected component is order to invoke property (ii) for a contradiction. Call a cell of K^* a border cell if one of its $2d$ neighbors in the grid is black. Clearly, border cells must be white, because no gray cell can have a black neighbor. Now, K^* is surrounded either by border cells, or by pieces of the boundary of the cube. The argument in (b) shows that there is a component of ∂K^* containing $\Omega(|K^*|^{1-1/d})$ white cells. By property (ii), this is impossible. This finishes the proof. $\square$

To show properties (i) through (iv) we further subdivide each cell into $(2d)^d$ cubes of side length $\ell = (2d\lfloor 1/r \rfloor)^{-1}$ which we call "minicells". We need two auxiliary results, one similar to Lemma 1 for the number of vertices in each minicell, and another about the connectivity of adjacent pairs of minicells.

Lemma 2 (Cube density Lemma)**.** *Grid the cube $[0,1]^d$ using cubes of side length $\ell = (2d\lfloor 1/r \rfloor)^{-1}$. Let $\gamma^{**} = 2d\sqrt[d]{2}$. Then for each $\gamma > \gamma^{**}$, there exist constants $0 < \alpha_\square < \beta_\square < \infty$ such that the following event, which we denote by $D_\square$, occurs whp:*

$$\alpha_\square n\ell^d < N(C) < \beta_\square n\ell^d \quad \textit{for every cube } C.$$

Proof. Given a fixed cube C, the number of vertices $N(C)$ is distributed as $\text{Binomial}(n, \ell^d)$. Thus, writing $f(x) = x - 1 - x\log x$, we have

$$\mathbf{P}\left\{N(C) \le \alpha_\square n\ell^d\right\} \le \exp\left(f(\alpha_\square)n\ell^d\right),$$
$$\mathbf{P}\left\{N(C) \ge \beta_\square n\ell^d\right\} \le \exp\left(f(\beta_\square)n\ell^d\right).$$

Choose $\alpha_\square$ and $\beta_\square$ to be the solutions of $f(x) = -1/2$ smaller and greater than 1 respectively. Define the event $D(C) = \{N(C) \in [\alpha_\square n\ell^d, \beta_\square n\ell^d]\}$. We can apply a union bound over all the cells to obtain

$$\mathbf{P}\left\{\bigcup_C D(C)^c\right\} \le \sum_C \mathbf{P}\left\{D(C)^c\right\} \le \sum_C 2e^{-n\ell^d/2} \le \ell^{-d} 2e^{-n\ell^d/2} \to 0,$$

because $\ell^{-d} = O(r^{-d}) = O(n/\log n)$ and $n\ell^d \ge n(r/2d)^d \ge 2\log n$ so that $e^{-n\ell^d/2} = O(1/n)$. $\square$

Lemma 3 (Cube connectivity)**.** *With high probability, all minicells are occupied and connected to their* $2d$ *adjacent neighbors.*

Proof. From Lemma 2, when the event $D_\square$ holds all cardinalities of the minicells are at least $\alpha_\square n\ell^d$ (and at most $\beta_\square n\ell^d$) whp. We condition on any point set with this distributional property, leaving only the choices of the c neighbors as a random event. Consider two neighboring minicells C and C' in any direction. By the choice of ℓ we have $\|x-y\| < r$ for any $x \in C$ and $y \in C'$.

When $D_\bigcirc$ holds each ball $B(X_i, r)$ has cardinality at most $\beta_\bigcirc nr^d$. By independence, the probability that *all* vertices in C' miss those in C with their c choices is not more than

$$\prod_{i\in C'}\left(1 - \frac{\alpha_\square n\ell^d}{\beta_\bigcirc nr^d}\right)^c \le \left(1 - \frac{\alpha_\square}{(2d)^d\beta_\bigcirc}\right)^{c\alpha_\square n\ell^d} \le \exp\left(\frac{-\alpha_\square^2 cn\ell^d}{(2d)^d\beta_\bigcirc}\right).$$

Since there is a total of $\ell^{-d} = O(r^{-d}) = o(n)$ minicells, the union bound shows that the probability that two neighboring minicells do not connect tends to zero. $\square$

We now show (i) through (iv) in four lemmas, leaving the hardest one, (iii), for last. We show all these properties with λ a sufficiently large constant depending upon γ, $q = 2(\log n)^{2/3}$, and $s = \exp\left((\log n)^{1/3}\right)$, leaving wide margins. Properties (i) and (iv) will follow easily from their minicell related statements above.

Lemma 4 (Part (iv))**.** *Each grid cell contains at most* $\lambda \log n$ *vertices with high probability, where* $\lambda = \beta_\square(2\gamma)^d$.

Proof. By Lemma 2 we have that every minicell of side length ℓ has less than $\beta_\square n\ell^d$ vertices whp. This implies immediately that every cell contains at most $(2d)^d\beta_\square n\ell^d < \beta_\square(2\gamma)^d \log n$ vertices. $\square$

Lemma 5 (Part (i))**.** *With high probability, all cells in the grid are occupied and connected to their* $2d$ *adjacent neighbors.*

Proof. It suffices to consider two adjacent minicells in the boundary of the cells. $\square$

Lemma 6 (Part (ii))**.** *The largest $*$-connected component of white cells has cardinality at most* $q = 2(\log n)^{2/3}$ *whp.*

Proof. We start by bounding the number of $*$-connected components of cells of a fixed size k. Fix an integer $\Delta \ge 2$, and let $\mathcal{T}_\Delta$ be the infinite Δ-ary rooted tree (every vertex has Δ children). Let $N_k(\mathcal{T}_\Delta)$ be the number of subtrees of $\mathcal{T}_\Delta$ containing the root and having exacly k vertices. It is well-known (see [23], Theorem 5.3.10) that $N_k(\mathcal{T}_\Delta) = \frac{1}{k}\binom{\Delta k}{k-1} \le (e\Delta)^{k-1}$.

The number of cells $*$-adjacent to any fixed cell is at most 3^d, thus the number of $*$-connected components of size k containing a specified cell is at most $(e3^d)^{k-1}$. To see this it suffices to consider a spanning tree of the component and to note that for any graph G with maximum degree Δ, the number of subtrees with k vertices containing a fixed vertex v is not larger than the corresponding number in $\mathcal{T}_\Delta$. Overall, the number of $*$-connected components of size k is at most $n(e3^d)^k$ since there are at most $O(r^{-d}) = o(n)$ starting cells.

Assume that we can then show that the probability that a cell is white is at most p. In that case, the probability that there is a $*$-connected component of size k or larger is not more than

$$n(e3^d)^k p^k, \tag{1}$$

by the union bound and because the colors of the cells are independent, given the location of the vertices. If we can show that

$$p \le \exp\left(-(\log n)^{1/3}\right)$$

then $k = 2(\log n)^{2/3}$ suffices to make the probability bound (1) tend to zero.

We now prove that for n large enough, the probability that a specified cell is white is at most $\exp\left(-(\log n)^{1/3}\right)$. By the preceding arguments, this will complete the proof of the lemma. Recall that a cell is colored white if the graph induced by the vertices lying inside the cell is not connected.

We subdivide the cell into minicells of side length ℓ. We know from Lemma 3 that all adjacent minicells are connected whp. Then if every minicell was connected inside, the whole cell would be black. Therefore, if we can bound the probability of the subgraph inside a minicell being disconnected by $\hat{p}$ the probability that the cell is white is $p < (2d)^d \hat{p}$ by a union bound.

Consider now a fixed minicell C and take any vertex v inside. Let V' be the subset of the c neighbors of v that fall in C. Consider then all c choices of the vertices in V' that fall in C as well, and that are not in $\{v\} \cup V'$. Call that second collection V''. We show that with high probability, all the remaining vertices select at least one vertex from $\{v\} \cup V' \cup V''$. Each of the remaining vertices selects in any of its c choices a vertex in $\{v\} \cup V' \cup V''$ with probability at least

$$\frac{1 + |V'| + |V''|}{\beta_{\bigcirc} n r^d}$$

when $D_{\bigcirc}$ holds. The probability that some vertex does not select any neighbor from $\{v\} \cup V' \cup V''$ is at most

$$\begin{aligned} \sum_{w \notin \{v\} \cup V' \cup V''} \left(1 - \frac{1 + |V'| + |V''|}{\beta_{\bigcirc} n r^d}\right)^c &\le \beta_{\square} n \ell^d \times \left(1 - \frac{1 + |V'| + |V''|}{\beta_{\bigcirc} n r^d}\right)^c \\ &\le \beta_{\square} n \ell^d \exp\left(-\frac{|V''| c}{\beta_{\bigcirc} n r^d}\right). \end{aligned}$$

If all vertices select a neighbor inside $\{v\} \cup V' \cup V''$, then clearly, all vertices are connected (and within distance six of each other, pairwise: two vertices of $\{v\} \cup V' \cup V''$ are within distance four, and any two neighbors of these are within distance six), and the cell is black. As a consequence, the probability of a having a white cell given the event $D = D_{\square} \cap D_{\bigcirc}$ is thus bounded from above by

$$\mathbf{P}\left\{|V''| \le \delta^2 c^2/4 \mid D\right\} + \beta_{\square} n \ell^d \exp\left(-\frac{\delta^2 c^3}{4 \beta_{\bigcirc} n r^d}\right),$$

where $\delta > 0$ is a constant to be selected later. Note that, for any $\delta > 0$, the second term in the upper bound is smaller than $\exp\left(-(\log n)^{1/3}\right)$ for all n large enough.

Finally, then, we consider V' and V'' and condition on the event D. This implies that $N(C) \geq \alpha_{\Box} n\ell^d$. Now, for V'' to be small, one of the following events must occur: either V' is small, or V' is not small but V'' is small. Note that by definition $|V'|$ is stochastically larger than a

$$\mathrm{Binomial}\left(c, \frac{N(C)-c}{\beta_{\bigcirc} nr^d}\right).$$

Let $\delta = \alpha_{\Box}/2\beta_{\bigcirc}(2d)^d$ then for n large enough the above distribution is stochastically larger than a random variable Z distributed as $\mathrm{Binomial}(c,\delta)$. We repeat a similar argument and note that $|V''|$ is stochastically larger than a Z-fold sum of independent binomial random variables, each of parameters c and $(N(C)-c-c^2)/\beta_{\bigcirc} nr^d$. Thus, assuming D and for n large enough, $|V''|$ is stochastically larger than a $\mathrm{Binomial}(Zc,\delta)$.

So gathering the preceding observations, we obtain

$$\begin{aligned}
\mathbf{P}\left\{|V''| \leq \delta^2c^2/4 \mid D\right\} &\leq \mathbf{P}\left\{Z \leq \delta c/2\right\} + \mathbf{P}\left\{Z \geq \delta c/2,\ \mathrm{Binomial}(Zc,\delta) \leq \delta^2c^2/4\right\} \\
&\leq \mathbf{P}\left\{Z \leq \delta c/2\right\} + \mathbf{P}\left\{\mathrm{Binomial}(\left\lfloor \delta c^2/2 \right\rfloor, \delta) \leq \delta^2c^2/4\right\} \\
&\leq (2/e)^{\delta c/2} + (2/e)^{\delta^2c^2/4}.
\end{aligned}$$

This shows that for n large enough, $p \leq \exp\left(-(\log n)^{1/3}\right)$, as required. $\Box$

Lemma 7. *The smallest connected component of S_n is of size at least $s = \exp\left((\log n)^{1/3}\right)$ whp.*

Proof. It is in this critical lemma that we will use the full power of the threshold. The proof is in two steps. For that reason, we grow S_n in stages. Having fixed ε in the definition of

$$c = \sqrt{\frac{(2+\varepsilon)\log n}{\log\log n}},$$

we find an integer constant L (depending upon ε – see further on), and let all vertices select their c neighbors in rounds. In round one, each vertex selects

$$\hat{c} = \sqrt{\frac{(2+\varepsilon/2)\log n}{\log\log n}}$$

neighbors uniformly at random without replacement. Then, in each of the remaining $(c-\hat{c})/L$ rounds, each vertex chooses L further neighbors within its range r, but this time independently and with replacement, with a possibility of duplication and selection of previously selected neighbors. This makes the graph less connected (by a trivial coupling argument), and permits us to shorten the proof. Note that

$$\frac{c-\hat{c}}{L} = \Delta\sqrt{\frac{2\log n}{\log\log n}}, \qquad \text{where} \qquad \Delta = \frac{\sqrt{1+\varepsilon/2}-\sqrt{1+\varepsilon/4}}{L}.$$

After the first (main) round, we will show that the smallest component is whp at least $\delta\log n$ in size, for a specific $\delta > 0$. We then show that whp, in each of the remaining rounds, each component joins another component, and thus the minimal component size doubles in each round. After the last round, the minimal component is therefore of size at least

$$\delta\log n \times 2^{\frac{c-\hat{c}}{L}},$$

which in turn is larger than $\exp\left((\log n)^{1/3}\right)$ for all n large enough.

So, on to round one. Let N_h count the number of connected components of S_n of size exactly h obtained after round one. By definition, $N_h = 0$ for $h \le \hat{c}$. We show by the first moment method that whp the smallest component after round one is of size at least $\delta \log n$ for some $\delta > 0$.

Let $D_\bigcirc$ be the event described in Lemma 1. If $D_\bigcirc$ holds, the number of sets of h vertices that can be connected is bounded from above by $n(e\beta_\bigcirc nr^d)^h$ since we count subgraphs of the visibility graph with maximum degree $\Delta = \beta_\bigcirc nr^d$.

Given a fixed set $\{i_1, \ldots, i_h\}$ of indices, one can only form a connected component if all the h vertices choose their neighbours among the remaining vertices in the set. Assuming $h < \alpha_\bigcirc nr^d$, the probability of this is at most

$$\prod_{j=1}^{h}\left(\prod_{k=1}^{\hat{c}} \frac{h-k}{N(B(X_{i_j}, r)) - k}\right) \le \left(\prod_{k=1}^{\hat{c}} \frac{h-k}{\alpha_\bigcirc nr^d - k}\right)^h \le \left(\frac{h}{\alpha_\bigcirc nr^d}\right)^{\hat{c}h}.$$

Therefore,

$$\mathbf{E}\left\{N_h \mathbf{1}_{D_\bigcirc}\right\} \le n\left(e\beta_\bigcirc nr^d\right)^h \left(\frac{h}{\alpha_\bigcirc nr^d}\right)^{\hat{c}h}.$$

We can rewrite the upper bound as

$$f(h) = \exp\left(\log n + h\log\left(e\beta_\bigcirc nr^d\right) + \hat{c}h\log\left(\frac{h}{\alpha_\bigcirc nr^d}\right)\right).$$

Note that $f(h)$ is decreasing for $h \le \rho nr^d = \rho\gamma^d \log n$ where $\rho < \alpha_\bigcirc/e$ because

$$\frac{d}{dh}\Big(\log f(h)\Big) = \log\left(e\beta_\bigcirc nr^d\right) + \hat{c}\log\left(\frac{eh}{\alpha_\bigcirc nr^d}\right) < -1$$

for n sufficiently large since $\hat{c} = \omega(\log nr^d)$. For such ρ, and n large enough, the upper bound is thus maximal at $h = \hat{c}+1$. We have shown that

$$\mathbf{E}\left\{N_h \mathbf{1}_{D_\bigcirc}\right\} \le f(\hat{c}+1)e^{-h+\hat{c}+1} < n^{-\varepsilon/5}e^{-h+\hat{c}+1}$$

for n large enough, since we have

$$\begin{aligned}
f(\hat{c}+1) &= \exp\left(\log n + (\hat{c}+1)\log\left(e\beta_\bigcirc nr^d\right) + \hat{c}(\hat{c}+1)\log\left(\frac{\hat{c}+1}{\alpha_\bigcirc nr^d}\right)\right)\\
&= \exp\left(\hat{c}^2\left(\log(\hat{c}+1) - \log(\alpha_\bigcirc nr^d)\right) + \log n + o(\log n)\right)\\
&= \exp\left(\frac{(2+\varepsilon/2)\log n}{\log\log n}\left(\frac{1}{2}\log\log n - \log\log n\right) + \log n + o(\log n)\right)\\
&\le \exp\Big(-(1+\varepsilon/4)\log n + \log n + o(\log n)\Big)\\
&\le \exp\Big(-(\varepsilon/4 + o(1))\log n\Big),
\end{aligned}$$

This means we can take $\delta = \alpha_\bigcirc\gamma^d/e$. Define the event $E_h = [N_h > 0]$ of having a component of size h. Finally, the probability that a component of size at most $\delta \log n$ exists after round one is bounded from above by

$$\begin{aligned}
\mathbf{P}\left\{\bigcup_{h=\hat{c}+1}^{\delta\log n} E_h\right\} &\le \mathbf{P}\left\{D_{\bigcirc}^c\right\} + \sum_{h=\hat{c}+1}^{\delta\log n} \mathbf{P}\left\{E_h \cap D_{\bigcirc}\right\} \\
&\le \mathbf{P}\left\{D_{\bigcirc}^c\right\} + \sum_{h=\hat{c}+1}^{\delta\log n} \mathbf{E}\left\{N_h \mathbf{1}_{D_{\bigcirc}}\right\} \\
&\le o(1) + \left(\frac{e}{e-1}\right) n^{-\varepsilon/5} \to 0.
\end{aligned}$$

For the final act, we tile the unit cube into minicells of side length ℓ. Consider a connected component having size t after round one, where $\delta\log n \le t \le n^{1/4}$. (Note that, for n large enough, any component of size at least $n^{1/4}$ already satisfies the lower bound of $\exp\left((\log n)^{1/3}\right)$ we want to prove.) Let the vertices of this component populate the cells. The i-th cell receives n_i vertices from this component, and receives m_i vertices from all other components taken together. The cell is colored red if $n_i > m_i$ and blue otherwise. First note that not all cells can be red, since that would mean that $t = \sum_i n_i \ge n/2$. In one round, each vertex chooses L eligible vertices in its neighborhood independently and with replacement. Consider two neighboring cells i and j (in any direction or diagonally) of opposite color (i is red and j is blue). Conditional on $D = D_{\square} \cap D_{\bigcirc}$, the probability that these cells do not establish a link between the size t component and any of the other components is at most

$$\begin{aligned}
\left(1 - \frac{n_i}{\beta_{\bigcirc} n r^d}\right)^{L m_j} &\le \exp\left(-\frac{L n_i m_j}{\beta_{\bigcirc} n r^d}\right) \\
&\le \exp\left(-\frac{L(\alpha_{\square} n \ell^d/2)^2}{\beta_{\bigcirc} n r^d}\right) \qquad \text{(recall } \ell > r/2d \text{ and } r^d = \gamma^d \log n/n\text{)}, \\
&= \exp\left(-\frac{L\alpha_{\square}^2 \gamma^d}{4(2d)^{2d}\beta_{\bigcirc}} \log n\right).
\end{aligned}$$

Consider finally the situation that all cells are blue. Then the probability (still conditional on D) that no connection is established with the other components is not more than

$$\begin{aligned}
\prod_i \left(1 - \frac{n_i}{\beta_{\bigcirc} n r^d}\right)^{L m_i} &\le \exp\left(-\sum_i \frac{L n_i m_i}{\beta_{\bigcirc} n r^d}\right) \\
&\le \exp\left(-\frac{L\alpha_{\square}/2}{(2d)^d \beta_{\bigcirc}} \sum_i n_i\right) \\
&\le \exp\left(-\frac{L\alpha_{\square}\delta}{2(2d)^d \beta_{\bigcirc}} \log n\right).
\end{aligned}$$

Since there are not more than n components to start with, the probability that any component of size between $\delta\log n$ and $n^{1/4}$ fails to connect with another one is bounded from above by

$$n^{1-L\xi},$$

where $\xi = \min\{\alpha_{\square}^2 \gamma^d/4(2d)^{2d}\beta_{\bigcirc}, \alpha_{\square}\delta/2(2d)^d\beta_{\bigcirc}\}$. The probability that we fail in any of the $(c-\hat{c})/L$ rounds is at most equal to the probability that D fails plus

$$\frac{c-\hat{c}}{L} \times n^{1-L\xi} = o(1)$$

by choosing L large enough that $L\xi > 1$. Thus, whp, after we are done with all rounds, the minimal component size in S_n is at least

$$\delta \log n \times 2^{\frac{c-\hat{c}}{L}}.$$

This concludes the proof of Lemma 7. □

4 Upper bound for the spanning ratio and diameter

In the previous sections, we have identified the threshold for connectivity near the critical radius. The connectivity is of course an important property, but the order of magnitude of distances in the sparsified S_n graph should also be as small as possible. Here we show that in the same range of values of r as in Theorem 1, as soon as c is of the order of $\sqrt{\log n}$ the diameter of the graph S_n is $O(1/r)$ which is clearly best possible as even the diameter of $G_n(r)$ cannot be smaller than $\sqrt{d}/r$. This improves a result of Pettarin, Pietracaprina, and Pucci [22].

Given a connected graph embedded in the unit cube $[0,1]^d$ and two vertices u and v (points in space), let $d(u,v)$ denote the Euclidean distance between u and v when one is only allowed to travel in space along the straight lines between connected vertices in the graph (this is the intrinsic metric associated to the embedded graph). Of course $d(u,v) \ge \|u-v\|$, and one defines the spanning ratio as

$$\sup_{u,v} \frac{d(u,v)}{\|u-v\|} \ . \tag{2}$$

One would ideally want the spanning ratio to be as close to one as possible. In the present case, this definition is not very relevant, since there is a chance that points that are very close in the plane are not connected by an edge. In particular one can show that, with probability bounded away from zero, there is a pair of points at distance $\Theta(n^{-1/d})$ for which the smallest path along the edges is of length $\Theta(r)$, so that for some $\varepsilon > 0$, whp,

$$\liminf_{n\to\infty} \sup_{u,v\in S_n} \frac{d(u,v)}{\|u-v\|} \ge \varepsilon (\log n)^{1/d} \to \infty.$$

(To see this, consider the event that for a point X_i, one other point falls within distance $\delta n^{-1/d}$ and there are no other points within distance $\varepsilon(n/\log n)^{-1/d}$.) This justifies introducing the constraint that the points in the supremum in (2) be at least at distance r. Hence the following modified definition of *spanning ratio*:

$$\Gamma(S_n) := \sup_{i,j:\|X_i-X_j\|>r} \frac{d(X_i,X_j)}{\|X_i-X_j\|}.$$

The next theorem shows that the spanning ratio is within a constant factor of the optimal.

Theorem 5. *There exist a constant $\mu > 0$ such that for any $\gamma > \gamma^{**}$, if*

$$r = \gamma \left(\frac{\log n}{n}\right)^{1/d} \quad \text{and} \quad c \ge \mu\sqrt{\log n}\ ,$$

there exists a constant K independent of n such that $\Gamma(S_n) \le K$ whp. This implies the fact that the diameter of S_n is at most $K\sqrt{d}/r$.

The idea of the proof of Theorem 5 is the following. Partition the unit square into a grid of cells of side length $\ell = (1/2d)\lfloor 1/r \rfloor^{-1}$. We show that, with high probability, any two vertices i and j, such that X_i and X_j fall in the same cell, are connected by a path of length at most five. On the other hand, by Lemma 3, with high probability, any two neighboring cells contain two vertices, one in each cell, that are connected by an edge of S_n. These two facts imply the statement of the theorem. We prove the former in Lemma 9 below. The bound for the diameter follows immediately from the fact that, with high probability, starting from any vertex, a point in a neighboring cell can be reached by a path of length 6 and any cell can be reached by visiting at most $2d^2 \lfloor 1/r \rfloor$ cells.

Just like in the arguments for the lower and upper bounds for connectivity, all we need about the underlying random geometric graph G_n is that the points $X_1, \dots, X_n$ are sufficiently regularly distributed. This is formulated as follows: A *moon* is the intersection of two circles, one of radius r and the other of radius $r/2$ such that their centers are within distance $5r/4$ (see Figure 3). Denote by $M(x,y) = B(x,r) \cap B(y,r/2)$ the moon with centers x and y.

Lemma 8 (Moon density Lemma)**.** *Let $\gamma^{**} = \sqrt[d]{2/c_1}$, where c_1 is the infimum of the volume of a moon. Then for each $\gamma > \gamma^{**}$, there exist constants $0 < \alpha_{\between} < \beta_{\between} < \infty$ such that the following event, which we denote by $D_{\between}$, occurs whp:*

$$\alpha_{\between} \log n \ < \ N(M(X_i, y)) \ < \ \beta_{\between} \log n$$

for every $1 \le i \le n$ and every center of a cell y within distance $5r/4$ of X_i.

Proof. Since any moon has volume at least in $[c_1 r^d, c_2 r^d]$ the number of vertices $N(M(X_i,y))$ is stochastically between $\zeta_1 \sim \text{Binomial}(n, c_1 r^d)$ and $\zeta_2 \sim \text{Binomial}(n, c_2 r^d)$. Thus, we have for any $1 \le i \le n$

$$\mathbf{P}\left\{N(M(X_i,y)) \le \alpha_{\between} n r^d\right\} \le \mathbf{P}\left\{\zeta_1 \le \alpha_{\between} n r^d\right\} \le \exp\left(f\left(\frac{\alpha_{\between}}{c_1}\right) c_1 n r^d\right),$$
$$\mathbf{P}\left\{N(M(X_i,y)) \ge \beta_{\between} n r^d\right\} \le \mathbf{P}\left\{\zeta_2 \ge \beta_{\between} n r^d\right\} \le \exp\left(f\left(\frac{\beta_{\between}}{c_2}\right) c_2 n r^d\right).$$

Let $\alpha_{\between} < c_1$ so that $f(\alpha_{\between}/c_1) = -1/2$ and $\beta_{\between} > c_2$ so that $f(\beta_{\between}/c_2) = -1/2$. Define the events $D_i = \{N(M(X_i,y)) \in [\alpha_{\between} n r^d, \beta_{\between} n r^d] \ \ \forall y \in \mathcal{Y}_i\}$ where $\mathcal{Y}_i = \{y : \|X_i - y\| < 5r/4 \text{ and } y \text{ is a cell center}\}$. Note that there exists a constant C_d that only depends on d such that $|\mathcal{Y}_i| < C_d$. So, we can apply a union bound to obtain

$$\mathbf{P}\left\{\bigcup_{i=1}^{n} D_i^c\right\} \le \sum_{i=1}^{n} \mathbf{P}\left\{D_i^c\right\} \le \sum_{i=1}^{n} 2C_d \exp\left(-c_1 n r^d/2\right) \le 2C_d n \exp\left(-\frac{c_1 \gamma^d}{2} \log n\right) \to 0,$$

if $\gamma > \sqrt[d]{2/c_1}$ where $c_1 = \inf \operatorname{Vol} M(x,y)$. $\square$

The key lemma is the following.

Lemma 9. *Fix $\{X_1, \dots, X_n\}$ such that $D = D_{\between} \cap D_{\bigcirc}$ occurs with $\beta_{\bigcirc} \ge 128$. Let i, j be such that X_i and X_j fall in the same cell of the grid. If $c \ge \sqrt{(\beta_{\bigcirc}/2)\log n}$ then*

$$\mathbf{P}\left\{d(X_i, X_j) > 5 \mid X_1, \dots, X_n\right\} \le \frac{1}{n}(1 + o(1)) \ .$$

where $d(X_i, X_j)$ denotes the distance of i and j in the graph S_n.

Proof. Let $M_i \subset \{1, \ldots, n\}$ denote the set of all vertices k such that $d(X_i, X_k) \leq 2$ and X_k is within Euclidean distance $r/2$ of the center of the grid cell that contains X_i. The outline of the proof is the following: It suffices to show that M_i contains a large constant times $\log n$ vertices. Since the same is true for M_j and any two vertices in $M_i \cup M_j$ are within Euclidean distance r, with high probability there exists an edge between M_i and M_j, establishing a path of length 5 between i and j. Let N_i denote the set of c neighbors picked by i. Then each $h \in N_i$ chooses its c neighbors. Those that fall in the moon defined by X_h and the center of the cell belong to M_i, see Figure 3.

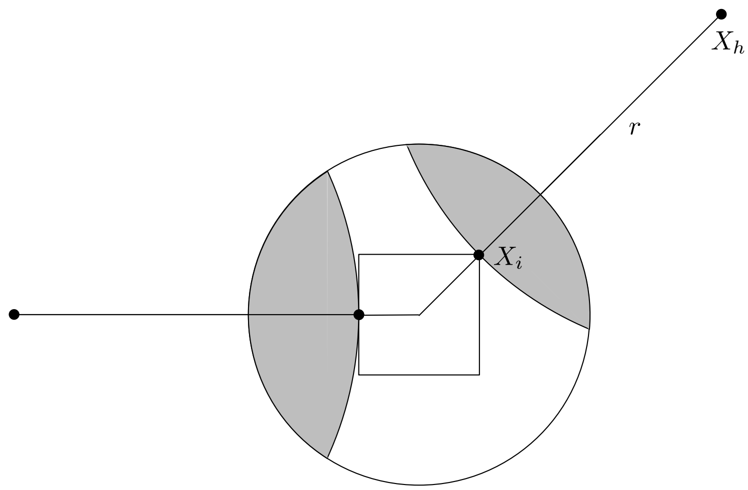

Figure 3: For any point X_i in a box of edge length ℓ, if $h \in N_i$ is a neighbor selected by i, then h may select neighbors within distance $r/2$ of the center of the square from the shaded regions–the so-called "moons". The volume of any moon is at least a constant times r^d. The figure shows two possible positions of X_i in a box with a corresponding neighbor and moon.

Next we establish the required lower bound for the cardinality of M_i. Clearly, $|M_i|$ is at least as large as the number of neighbors selected by the vertices in N_i that fall in R, the ball of radius $r/2$, centered at the mid-point of the cell into which X_i falls. Denote by $h_1, \ldots, h_c$ the c vertices belonging to N_i. Then

$$|M_i| \geq |N_{h_1} \cap R| + |N_{h_2} \setminus N_{h_1} \cap R| + \cdots + |N_{h_c} \setminus \left(N_{h_1} \cup \ldots \cup N_{h_{c-1}}\right) \cap R| \ .$$

h_1 picks its c neighbors among all vertices within distance r. The number of those neighbors falling in R has a hypergeometric distribution. Since we are on D, $|N_{h_1} \cap R|$ stochastically dominates H_1, a hypergeometric random variable with parameters $(c, \beta_{\bigcirc} \log n, (\beta_{\bigcirc} - \alpha_{\between}) \log n)$. To lower bound the second term on the right-hand side, and to gain independence, remove all c neighbors picked by h_1. Then $|N_{h_2} \setminus N_{h_1} \cap R|$ stochastically dominates H_2, a hypergeometric random variable with parameters $(c, \beta_{\bigcirc} \log n - c, (\beta_{\bigcirc} - \alpha_{\between}) \log n + c)$ (independent of H_1). Continuing this fashion, we obtain that $|M_i|$ is stochastically greater than $\sum_{i=1}^{c} H_i$ where the H_i are independent and H_i is hypergeometric with parameters $(c, \beta_{\bigcirc} \log n - (i-1)c, (\beta_{\bigcirc} - \alpha_{\between}) \log n + (i-1)c)$. Since $c \geq \sqrt{(\beta_{\bigcirc}/2) \log n}$, this may be bounded further as $\sum_{i=1}^{c} H_i$ is also stochastically greater than $\sum_{i=1}^{c} \widetilde{H}_i$ where the $\widetilde{H}_i$ are i.i.d. hypergeometric random variables with parameters $(c, (\beta_{\bigcirc}/2) \log n, (3\beta_{\bigcirc}/2) \log n)$.

Clearly, $\mathbf{E}\{\sum_{i=1}^{c} \widetilde{H}_i\} = c^2/4 \geq (\beta_{\bigcirc}/8) \log n$. We may bound the lower tail probabilities of $\sum_{i=1}^{c_n} \widetilde{H}_i$ by recalling an observation of Hoeffding [16] according to which the expected value of any convex function of a hypergeometric random variable is dominated by that of the corresponding binomial random variable. Therefore, any tail bound obtained by Chernoff bounding for the binomial distribution also applies for the hypergeometric distribution. In particular,

$$\mathbf{P}\left\{\sum_{i=1}^{c} \widetilde{H}_i \leq \frac{1}{2}\mathbf{E}\left\{\sum_{i=1}^{c} \widetilde{H}_i\right\}\right\} \leq \exp\left(\frac{-\mathbf{E}\left\{\sum_{i=1}^{c} \widetilde{H}_i\right\}}{8}\right) \leq \exp\left(-\frac{\beta_{\bigcirc}}{64}\log n\right) \leq n^{-2} \ .$$

Thus, by the union bound, we obtain that

$$\mathbf{P}\left\{\exists i \in \{1,\ldots,n\} : |M_i| \leq \frac{\beta_{\bigcirc}\log n}{16} \;\middle|\; X_1,\ldots,X_n,F\right\} \leq \frac{1}{n} \ .$$

Thus, we have proved that with high probability, for every vertex i, the number of second generation neighbors (i.e., the neighbors selected by the neighbors selected by i) that end up within distance $r/2$ of the center of the grid cell containing i is proportional to $\log n$. In particular, if i and j are two vertices in the same cell, then both M_i and M_j contain at least $(\beta_{\bigcirc}/16)\log n$ vertices. If two of these vertices coincide, there is a path of length 4 between i and j. Otherwise, with very high probability, at least one vertex in M_i selects a neighbor in M_j, creating a path of length five. Indeed, the probability that all neighbors selected by the vertices in M_i miss all vertices in M_j, given that $|M_i|$ and $|M_j|$ are both greater than $(\beta_{\bigcirc}/16)\log n$ and $M_i \cap M_j = \emptyset$ is at most

$$\prod_{h\in M_i}\prod_{k=0}^{c-1}\left(1-\frac{(\beta_{\bigcirc}/16)\log n - k}{\beta_{\emptyset}\log n}\right) \leq \left(1-\frac{\beta_{\bigcirc}}{32\beta_{\emptyset}}\right)^{c(\beta_{\bigcirc}/16)\log n}$$

which goes to zero faster than any polynomial function of n. (Here we used that fact that $c \leq (\beta_{\bigcirc}/32)\log n$ for a sufficiently large n.) Finally, we may use the union bound over all pairs of at most $\binom{n}{2}$ pairs of vertices i and j to complete the proof of the lemma. $\square$

Acknowledgments

The authors would like to thank the referees for their valuable comments that contributed significantly to improve the quality and clarity of this manuscript.

References

[1] T. Bohman and A. Frieze. Hamilton cycles in 3-out. *Random Structures and Algorithms*, 35: 393–417, 2009.

[2] B. Bollobás and I. Leader. Compression and isoperimetric inequalities. *Journal of Combinatorial Theory, Series A*, 56:47–62, 1991.

[3] B. Bollobás and O. Riordan. *Percolation.* Cambridge University Press, 2006.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transaction on Information Theory*, 52:2508–2530, 2006.

[5] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23:493–507, 1952.

[6] P. Crescenzi, C. Nocentini, A. Pietracaprina, and G. Pucci. On the connectivity of bluetooth-based ad hoc networks. *Concurrency and Computation: Practice and Experience*, 21:875–887, 2009.

[7] J.-D. Deuschel and A. Pisztora. Surface order large deviations for high-density percolation. *Probability Theory and Related Fields*, 104:467–482, 1996.

[8] M. Draief and L. Massoulié. *Epidemics and Rumours in Complex Networks*, volume 369 of *London Mathematical Society Lecture Notes*. Cambridge University Press, Cambridge, 2010.

[9] D. Dubhashi, C. Johansson, O. Häggström, A. Panconesi, and M. Sozio. Irrigating ad hoc networks in constant time. In *Proceedings of the seventeenth annual ACM symposium on Parallelism in algorithms and architectures*, SPAA '05, pages 106–115, New York, NY, USA, 2005. ACM.

[10] D. Dubhashi, O. Häggström, G. Mambrini, A. Panconesi, and C. Petrioli. Blue pleiades, a new solution for device discovery and scatternet formation in multi-hop bluetooth networks. *Wireless Networks*, 13:107–125, 2007.

[11] T. Fenner and A. Frieze. On the connectivity of random m-orientable graphs and digraphs. *Combinatorica*, 2(4):347–359, 1982.

[12] T. Fenner and A. Frieze. On the existence of hamiltonian cycles in a class of random graphs. *Discrete Mathematics*, 45(2-3):301–305, 1983.

[13] F. Ferraguto, G. Mambrini, A. Panconesi, and C. Petrioli. A new approach to device discovery and scatternet formation in bluetooth networks. In *Proceedings of the 18th Interntional Parallel and Distributed Processing Symposium*, volume 13, page 221b, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[14] G. Grimmett. *Percolation*, volume 321 of *A Series of Comprehensive Studies in Mathematics*. Springer-Verlag, New York, NY, USA, 1989.

[15] P. Gupta and P. Kumar. Critical power for asymptotic connectivity in wireless networks. In *Decision and Control, 1998. Proceedings of the 37th IEEE Conference on*, volume 1, pages 1106–1110, 1998.

[16] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[17] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley, New York, 2000.

[18] H. Kesten. The critical probability of bond percolation on the square lattice equals 1/2. *Communications in Mathematical Physics*, 74(1):41–59, 1980.

[19] A. Panconesi and J. Radhakrishnan. Expansion properties of (secure) wireless networks. In *Proceedings of the sixteenth annual ACM symposium on Parallelism in algorithms and architectures*, SPAA '04, pages 281–285, New York, NY, USA, 2004. ACM.

[20] M. Penrose. The longest edge of the random minimal spanning tree. *The annals of applied probability*, 7(2):340–361, 1997.

[21] M. Penrose. *Random Geometric Graphs*. Oxford Studies in Probability. Oxford University Press, 2003.

[22] A. Pettarin, A. Pietracaprina, and G. Pucci. On the expansion and diameter of bluetooth-like topologies. In A. Fiat and P. Sanders, editors, *Algorithms - ESA 2009*, volume 5757 of *Lecture Notes in Computer Science*, pages 528–539. Springer, Berlin / Heidelberg, 2009.

[23] R. Stanley. *Enumerative combinatorics, Volume 2.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[24] R. Whitaker, L. Hodge, and I. Chlamtac. Bluetooth scatternet formation: a survey. *Ad Hoc Networks*, 3:403–450, 2005.