Fast Low Degree Connectivity of Ad-Hoc Networks via Percolation

Emilio De Santis[1], Fabrizio Grandoni[2], and Alessandro Panconesi[2]

[1] Dipartimento di Matematica, Sapienza Università di Roma, P.le Aldo Moro 2, 00185 Roma, Italy, desantis@mat.uniroma1.it
[2] Dipartimento di Informatica, Sapienza Università di Roma, Via Salaria 113, 00198 Roma, Italy, {grandoni,ale}@di.uniroma1.it

Abstract. Consider the following classical problem in ad-hoc networks: n devices are distributed uniformly at random in a given region. Each device is allowed to choose its own transmission radius, and two devices can communicate if and only if they are within the transmission radius of each other. The aim is to (quickly) establish a connected network of low average and maximum degree.
In this paper we present the first efficient distributed protocols that, in poly-logarithmically many rounds and with high probability, set up a connected network with $O(1)$ average degree and $O(\log n)$ maximum degree. This is asymptotically the best possible.
Our algorithms are based on the following result, which is a non-trivial consequence of classical percolation theory: suppose that all devices set up their transmission radius in order to reach the K closest devices. There exists a universal constant K (independent of n) such that, with high probability, there will be a unique giant component, i.e. a connected component of size $\Theta(n)$. Furthermore, all remaining components will be of size $O(\log^2 n)$. This leads to an efficient distributed probabilistic test for membership in the giant component, which can be used in a second phase to achieve full connectivity.
Preliminary experiments suggest that our approach might very well lead to efficient protocols in real wireless applications.

1 Introduction

In this paper we study a geometric random graph model that has interesting applications to wireless networking. We are given n points distributed uniformly at random within the unit square. Each point v is connected via a directed arc to the *closest* $k(v)$ points, according to the Euclidean distance, where $k(v)$ is a positive integer value. Given this directed graph we define an undirected graph G with the same vertex set as follows: $vw \in E(G)$ if and only if there is a directed arc from v to w and viceversa. Henceforth, we will refer to the the points also as *nodes* or *devices*.

The question that we study in this paper is how to determine the value of the $k(v)$'s in order to meet two conflicting goals: G should be connected, but its average degree should be as small as possible. Moreover, the maximum degree

should also be small. In this paper we give two efficient (i.e. poly-logarithmic) distributed algorithms that set up a connected network G in such a way that (a) the expected degree of a node is constant and (b) the maximum degree is $O(\log n)$ (resp. $O(\log^2 n)$) with high probability. The number of communication rounds needed is $O(\log^3 n)$ (resp. $O(\log^2 n)$).

These results appear to be relevant to wireless networking. Our model of connectivity for G is the most realistic from the point of view of wireless applications since the communication primitives of standards such as IEEE 802.11 and Bluetooth rely on ack messages, and therefore a communication link really exists only when both nodes are within transmission radius of each other. Our algorithms are very simple and we give experimental evidence that they might very well admit efficient wireless implementations. Limiting the degree of nodes can be beneficial in many ways. For instance, in security applications, nodes exchange keys and run cryptographic protocols with their neighbors (see, for for instance, [13]). Limiting the degree reduces the amount of traffic and computation. Moreover, the transmission radius of v is set in order to reach its $k(v)$ closest neighbors. Hence, the larger $k(v)$, the larger the power v needs. Limiting the $k(v)$'s thus reduces the overall transmission power and translates in longer network lifetimes. In particular, we can show that while (a) the optimal power consumption is, with high probability, proportional to the area of the region within which the nodes are randomly distributed, (b) the expected power consumption to sustain the network with our approach is order of the area and hence it is in some sense optimal. (This is a consequence of our main results and its proof is omitted from this extended abstract). Probably the most important benefit is that, by bounding $k(v)$ and by setting the transmission power accordingly, interference is kept under control: The lower a node's degree, the lower the number of neighbors affected by a transmission and, consequently, the lower the number of possible packet collisions and corresponding retransmission (see, for instance, [1]). Note that our high probability bound on the maximum degree ensures that not only things are good on average, but also that no node will be penalized too much.

Let us now describe our algorithms. Probably the simplest distributed algorithm one can think of is the following: set beforehand $k(v) = K$, for all nodes v and for a suitable constant K (see [5, 6, 10, 11] for experimental results). Unfortunately, there is no constant K which guarantees connectivity with probability going to 1 as n grows. To reach that goal, K must grow like $\log n$ [12].

If points can communicate, the situation changes. Indeed, Kucera [7] gives a protocol to decide $k(v)$, for all v, that sets up a connected network of expected constant degree and maximum degree $O(\log n)$. The result however is existential in flavor: the protocol requires nodes to explore linear-size components of the network, linearly many times, making it completely impractical. Our faster protocols are based on the following insight.

Theorem 1. *There is a universal constant K, independent of n, such that, if all the devices set $k(v) = K$, with probability going to 1 as n grows, the network has*

the following special structure: (a) there is a unique giant component containing $\Theta(n)$ nodes; (b) all other components have size $O(\log^2 n)$.

This theorem says that it is possible to set up a giant component in a very simple way, a useful fact by itself (e.g. for coverage applications). It also says that there is an efficient distributed test for membership to the giant component: a node belongs to the unique giant component if and only if it belongs to a component with more than (order of) $\log^2 n$ nodes.

Given this, the following strategy is very natural. Devices that discover to be trapped inside small components increase their transmitting power in order to reach a device that belongs to the giant component. A node in the giant component that is contacted in this way will respond, setting its power in order to reach the calling node. We shall refer to this as Algorithm A.

Theorem 2. *Algorithm A sets up a network in which the expected number of neighbors of each device is constant. Furthermore, with probability going to 1 as n grows, the network is connected and its maximum degree is $O(\log^2 n)$. The number of communication rounds required is $O(\log^2 n)$.*

This result gives an exponential speed up with respect to [7]. We can improve the bound on the maximum degree at the expense of an increased communication cost. Suppose that each device v belonging to a small component increases its transmitting power a bit at a time, each time checking if it has reached a node in the giant component. Nodes closer to the giant component will join it first. Nodes farther away might be able to connect to such closer nodes, rather than expanding their radius all the way to the closest node in the original giant component. In the next section we will give a precise description of this, referred to as Algorithm B.

Theorem 3. *Algorithm B sets up a network in such a way that the expected number of neighbors of each device is constant. Furthermore, with probability going to 1 as n grows, the network is connected and its maximum degree is $O(\log n)$. The number of communication rounds required is $O(\log^3 n)$.*

Preliminary experiments show that our approach might very well lead to efficient protocols in real wireless applications.

2 The algorithms

The input to the algorithms consists of n devices that are spread uniformly at random within the unit box. The value of n is known to the devices. We assume that the network is synchronous and that in one communication round each device is able to send messages to all neighbors and to receive messages from all of them. The running time of the protocols is given by the number of such communication rounds. Each device is initially marked as *lacustrine*. Algorithm A has two constant parameters K and φ, and works as follows.

Phase 1: Every device v sets its own transmission radius in order to reach the closest $k(v) := K$ neighbors (all the devices if $n < K$).

Phase 2: Every device v explores its own connected component, denoted as $C(v)$. If $|C(v)| > C = \varphi \log^2 n$, v marks itself as *continental*. Every lacustrine device v increases $k(v)$ in order to reach the closest continental device, denoted as $s(v)$. Device $s(v)$ responds by increasing its transmission radius in order to reach v (if this is not already the case).

Algorithm B, has a third constant parameter $\mu > 0$, and works as follows:

Phase 1: As in Algorithm A.
Phase 2: Repeat $\mu \log n$ many times: Let v be lacustrine. If $|C(v)| > C = \varphi \log^2 n$ then v marks itself as *continental*. Otherwise, v increases $k(v)$ by one, in order to reach the next closest device $s(v)$. If $s(v)$ is continental, it responds by increasing its transmission radius in order to reach v.

The mapping $s(v)$ is, for all practical purposes, well-defined since almost surely all pairwise distances are different. The constants K and φ (independent of n) ensure that, with high probability, at the end of the first phase there is a unique giant component of $\Theta(n)$ points, while all the other components contain $\varphi \log^2 n$ points. Observe that Phase 1 does not require any global information, such as the value of n. We shall refer to a round of Phase 2 of Algorithm B as an *expansion* round. The constant μ ensures that, with high probability, within $\mu \log n$ expansion rounds all the nodes become continental. As a consequence, Algorithms A and B achieve connectivity with high probability. Moreover they require $O(\log^2 n)$ and $O(\log^3 n)$ communication rounds, respectively.

3 Overview

Since the proof of Theorems 1, 2, and 3 is rather involved we first give an overview. The basic idea is to reduce our connectivity problem to site percolation in a finite box (for an introduction to percolation, see, e.g., [3, 9]). It is known that in the supercritical phase, with high probability there is a unique giant cluster in the box and that its complement consists of small regions each containing $O(\log^2 n)$ sites (see, among others, [2–4]). In the following we shall refer to the maximal regions in the complement of the giant cluster as *lakes*. The reduction will ensure that the unique giant cluster in the box will correspond to a unique giant component of points, and that the remaining components of points are trapped inside lakes, each containing $O(\log^2 n)$ points. This is the situation at the end of Phase 1 (with high probability).

The reduction to site percolation is achieved via several intermediate steps. The first is to replace the uniform distribution of points with a Poisson distribution, to exploit the strong independence properties of the latter. In particular, unlike the uniform distribution, the Poisson distribution ensures that the configuration of points in one region does not affect the distribution of points of any other disjoint region. There are some standard and rather general ways to connect the two settings, but here we will make use of a coupling construction that gives stronger bounds than these general tools. The configurations of points given by the mentioned Poisson processes is referred to as scenario $\mathcal{A}$.

We introduce next a first percolation problem, scenario $\mathcal{B}$, by subdividing the unit square into a grid of non-overlapping square cells. The area of each cell is such that the expected number of points inside it is a constant parameter α. This parameter is crucial for the whole construction. A cell is *good* if the number of points that it contains is in $[\frac{\alpha}{2}, 2\alpha]$.

Scenario $\mathcal{B}$ is a Bernoulli field but unfortunately clusters of good cells do not translate necessarily into connected components of points. Therefore another percolation problem, scenario $\mathcal{C}$, is introduced by defining a cell i *open* if it is good and moreover all cells within distance D from i are good. The value of D is a constant, independent of n. The definition is such that the points belonging to a cluster of open cells form themselves a component of points.

The problem with scenario $\mathcal{C}$ is that it is not a Bernoulli field– knowing that a cell i is open or closed alters the distribution of neighboring cells. However the mentioned dependence only involves cells at distance at most $h = 2D$ from i, that is the field is *h-dependent* (see [3]). Therefore a new scenario $\mathcal{D}$ is introduced. Scenario $\mathcal{D}$ is given by a general construction of [8]. This construction translates scenario $\mathcal{C}$ into a Bernoulli field (i.e. $\mathcal{D}$) that is stochastically dominated by $\mathcal{C}$: if a cell is open in scenario $\mathcal{D}$ then it is also open in scenario $\mathcal{C}$. It follows that if a giant cluster of open cells exists with probability p in scenario $\mathcal{D}$, the same cluster exists in scenario $\mathcal{C}$ with probability at least p. Essentially, scenario $\mathcal{C}$ ensures that the unique giant component of cells that, with high probability, exists in it translates into a connected component of points in scenario $\mathcal{A}$, and that all other components are small. While scenario $\mathcal{D}$ is used to compute the probability that these events take place.

The probability that sites are on or off in the various scenarios depends on the value of the constant K of the protocol. We will fix K in such a way that a unique giant cluster of open cells exists in scenario $\mathcal{D}$ with high probability. By construction, this translates into a giant component of points in scenario $\mathcal{A}$, henceforth denoted as $\mathcal{G}$.

To ensure that $\mathcal{G}$ is unique in scenario $\mathcal{A}$ we make use of the definition of open cells of scenario $\mathcal{C}$ which ensures that points trapped inside lakes cannot connect to points in other lakes, bypassing $\mathcal{G}$.

Remark 1. By setting the radius of each point to $\sim n^{-1/2}$ we would obtain a simpler reduction to site percolation to show the emergence of a giant component. Our reduction however is independent of n, showing that a giant component can be created with no global information at all. This might be of independent interest.

3.1 Preliminaries

As mentioned, in scenarios $\mathcal{B}, \mathcal{C}$, and $\mathcal{D}$, we consider a partition of the unit square into a grid of non-overlapping square cells of the same size. The number of cells is $m = k^2$, where $k := \lfloor \sqrt{\frac{n}{\alpha}} \rfloor$, and α is a constant. This partition naturally induces a mesh, where the nodes are the cells and each cell has (at most) four neighbors:

the cells on the left, right, top and bottom. Let $i_{x,y}$ be the cell in position (x, y) in the grid. The *distance* between i_{x_1,y_1} and i_{x_2,y_2} is $\max\{|x_1 - x_2|, |y_1 - y_2|\}$. The *star-neighbors* of cell i are the cells at distance one from i. We call *cluster* a connected component of cells, and *star-cluster* a connected component of cells with respect to star-neighborhood. We will use this distance in the mesh, while we will use the Euclidean distance when talking about points in the unit square.

A *giant* cluster is a cluster of open cells which contains at least $\delta\, m$ cells, for a given constant $\delta \in (0, 1]$. Assuming a unique giant cluster (an event that we will show happening with high probability), a *lake* is a maximal star-cluster in the complement of the giant cluster. A *giant component* is a connected component of points of linear (in n) size in the network set-up by the protocol. With $|X|$ we denote either the number of cells of X or the number of points of X, depending on whether X is a cluster or a component, respectively.

4 Emergence of a giant component

In this section we show that after Phase 1 of the algorithm $|\mathcal{G}| = \Theta(n)$ with high probability.

As outlined previously we consider four different scenarios. In scenario $\mathcal{A}$ points are placed in the unit box by means of a Poisson process. More precisely, we consider two Poisson processes P_0 and P_t. Process P_0 has parameter $\mu_0 := n - \epsilon\, n$, where ϵ is a small positive constant, say $\epsilon = \frac{1}{4}$. Process P_t is built on top of P_0 by adding to it a new independent Poisson process $\Delta\, P$ with parameter $2\,\epsilon\, n$. It is well-known that P_t is a Poisson process with parameter $\mu_t := \mu_0 + 2\,\epsilon\, n = n + \epsilon\, n$. We then define a sequence of point processes $\{Q_i\}$ sandwiched between P_0 and P_t. Starting from $Q_0 := P_0$, Q_{i+1} is given by Q_i by adding one point chosen uniformly at random in $P_t - Q_i$.

Our reduction to site percolation will apply simultaneously to all Q_i's, showing the existence of a unique giant component in scenario $\mathcal{A}$ for each Q_i with high probability. Each Q_i generates points uniformly in the box (conditioned on the given number of points). The next lemma shows that, with high probability, one of the Q_i will generate exactly n points. As a consequence, if something holds for all Q_i's of scenario $\mathcal{A}$ simultaneously, it also holds for the original n-points problem.

Lemma 1. *Let N_0 and N_t be the Poisson variables relative to P_0 and P_t, respectively. There is a positive constant γ (independent of n) such that*

$$\Pr\left(\overline{\{N_0 \le n \le N_t\}}\right) \le e^{-\gamma n}.$$

Proof. (Sketch) Apply the large deviation principle to $\{N_0 > n\}$ and to $\{n < N_t\}$.

We now define scenario $\mathcal{B}$. Let us subdivide the unit square into a grid of $m = k^2$ non-overlapping square cells, where $k := \lfloor\sqrt{\frac{n}{\alpha}}\rfloor$, and α is a constant. Note that $m = \Theta(n)$ and the expected number of points in a cell is (roughly) α.

The parameter α plays a crucial role in the whole proof. This parameter should be thought of as a large constant. Its value will be fixed later.

Definition 1. *A cell is* good *if the number of points in the cell given by both P_0 and P_t is in $[\frac{\alpha}{2}, 2\alpha]$. The cell is* bad *otherwise.*

In scenario $\mathcal{B}$ we define a site percolation problem with a Bernoulli field, where the good cells will be the on sites in the finite box. Note that if a cell is good then its number of points is in $[\frac{\alpha}{2}, 2\alpha]$ for all Q_i's of scenario $\mathcal{A}$. By construction, cells are good independently of each other and the probability of being good is the same for all cells.

Lemma 2. *Let p_α be the probability that a cell is good. Then* $\lim_{\alpha\to\infty} p_\alpha = 1$.

Proof. Apply the standard large deviation principle to the Poisson random variables corresponding to the number of points given by P_0 and P_t in the cell considered.

We would like to show that large connected clusters of good cells in scenario $\mathcal{B}$ give raise to large connected components of points in scenario $\mathcal{A}$. This however is not true. This motivates the next scenario $\mathcal{C}$. In it we consider another site percolation problem which is not, however, independent. Let $D \geq 3$ be a constant to be fixed later.

Definition 2. *A cell i is* open *if i and all the cells at distance at most D from i are good. The cell is* closed *otherwise.*

We now choose K and D in order to enforce the following two properties: First, if we have a giant cluster of open cells in scenario $\mathcal{C}$, then the points inside these cells belong to a giant component $\mathcal{G}$ of scenario $\mathcal{A}$ for each Q_i. Second, components other than $\mathcal{G}$ will be trapped inside lakes (delimited by closed cells), i.e. points inside distinct lakes cannot establish links among them directly, bypassing $\mathcal{G}$. The first property is guaranteed by choosing

$$K := 7^2(2\alpha) = 98\,\alpha$$

(the maximum number of points in the cells at distance at most 3 from a given open cell). This choice of K ensures that the transmission radius of every point in an open cell will reach all points in neighboring cells. Thus, if two neighboring cells are open the points in them will form a clique. Observe that, for similar reasons, any point inside a cell at distance at most $D-3$ from an open cell i belongs to the same connected component to which the points in i belong. This implies that, by choosing D such that $2(D-3)\alpha/2 > K$, say $D = 102$, also the second property is ensured (see the proof of Lemma 5). We have not tried to optimize the values of D and K.

In scenario $\mathcal{C}$ we have the desired translation of connectivity– if we have a cluster of open cells then all points belonging to these cells form a connected component of points in scenario $\mathcal{A}$ (for all Q_i's). Moreover, the probability q_α that a cell is open satisfies $q_\alpha \geq p_\alpha^{(2D+1)^2}$ and thus $\lim_{\alpha\to\infty} q_\alpha = 1$. Unfortunately

scenario $\mathcal{C}$ is not a Bernoulli field. Definition 2 however ensures that it is h-dependent with $h = 2D$ (the probability that a cell is open is independent from what happens in cells at distance $2D + 1$ or larger).

Therefore we introduce a fourth scenario $\mathcal{D}$ that is a Bernoulli field. The connection between scenarios $\mathcal{C}$ and $\mathcal{D}$ is given by a very general theorem of [8]. The theorem states that there is a coupling between scenario $\mathcal{C}$ and a Bernoulli field, referred to as scenario $\mathcal{D}$, with site probability r_α such that: (a) If q_α goes to 1 so does r_α, and hence $\lim_{\alpha\to\infty} r_\alpha = 1$; and, (b) If a cell is open in scenario $\mathcal{D}$ the same cell is open in scenario $\mathcal{C}$. Therefore, if we have a giant cluster in scenario $\mathcal{D}$, the same cells form a cluster also in scenario $\mathcal{C}$. In turn, all points inside these cells will be connected in scenario $\mathcal{A}$.

Scenario $\mathcal{D}$ allows us to estimate the probability of relevant events. A result of Deuschel and Pisztora [2] ensures that, for every constant $\delta \in (0, 1)$ there is a value of $r_\alpha < 1$ such that, with probability at least $1 - e^{-\gamma\sqrt{m}}$ there is a unique giant cluster of at least δm open cells in scenario $\mathcal{D}$, for some constant $\gamma > 0$. The following theorem and corollary summarize the discussion above.

Theorem 4. *Let $\mathcal{G}$ denote a maximum cardinality component of points at the end of Phase 1 of the algorithm. For every $c \in (0, \frac{1}{2})$ there is a choice of $\alpha > 0$, and so a corresponding choice of K, such that*

$$\Pr(|\mathcal{G}| \le cn) \le 2\, e^{-\xi\sqrt{n}}$$

where $\xi > 0$ is a constant independent of n.

Proof. Let $m \simeq \frac{n}{\alpha}$ be the number of cells in scenario $\mathcal{D}$, and let C be a maximum size cluster in scenario $\mathcal{D}$. By [2], for any given $\delta \in (0, 1)$ there is a value of α such that $\Pr(|C| \le (1-\delta)m) \le e^{-\gamma\sqrt{m}}$. The same set of cells is open in scenario $\mathcal{C}$. By definition, an open cell contains at least $\frac{\alpha}{2}$ points. Thus, C corresponds to a giant component of at least $(1 - \delta)m\frac{\alpha}{2}$ points in scenario $\mathcal{A}$ (for all Q_i's). Recalling Lemma 1 and by choosing $\delta < 1 - 2c$, we have

$$\Pr(|\mathcal{G}| \le cn) \le \Pr(|C| \le 2cn/\alpha) + \Pr(\overline{\{N_0 \le n \le N_t\}}) \le 2\, e^{-\xi\sqrt{n}},$$

for some constant $\xi > 0$ and n large enough. This follows from the fact that if the condition $\{N_0 \le n \le N_t\}$ does not hold we give up, while we pursue the construction of the 4 scenarios only if it holds.

Remark 2. By choosing ϵ appropriately in the definition of the two Poisson processes P_0 and P_t of scenario $\mathcal{A}$, and by defining a cell to be good if its number of points in both P_0 and P_t is in the interval $[(1 - \epsilon')\alpha, (1 + \epsilon')\alpha]$, the size of $\mathcal{G}$ can be made arbitrarily close to 1, for a proper choice of ϵ'.

Corollary 1. *For every $c \in (0, 1)$ there exist constants $K > 0$ and $\gamma > 0$ such that, if every point v sets $k(v) = K$, then $\Pr[|\mathcal{G}| < cn] \le e^{-\gamma\sqrt{n}}$.*

Proof. It follows from the proof of Theorem 4 and Remark 2.

5 Uniqueness of the giant component

In this section we show that at the end of Phase 1 of the algorithm, with probability $1 - o(1)$, there is a unique giant component $\mathcal{G}$ with linearly many points, and that all other components contain $O(\log^2 n)$ points .

The next lemma bounds the number of cells of a lake in scenario $\mathcal{C}$.

Lemma 3. *For any lake L of scenario $\mathcal{C}$, $\Pr(|L| > k) \leq e^{-\gamma\sqrt{k}}$, where $\gamma > 0$ is a constant and $|L|$ is the number of cells of L.*

Proof. It is well-known that for a Bernoulli field, in the super-critical phase for percolation, if we take any star-cluster S in the complement of the giant cluster then $\Pr(|S| > k) \leq e^{-\gamma\sqrt{k}}$, for some constant $\gamma > 0$ [3]. Therefore the same holds for any lake of scenario $\mathcal{D}$. Now, by the monotonicity implied by the coupling construction of [8] that relates scenario $\mathcal{C}$ and scenario $\mathcal{D}$, the same bound holds for L in scenario $\mathcal{C}$ (lakes can only be smaller).

The next, crucial lemma bounds the number of points of each lake in scenario $\mathcal{A}$. The difficulty is to analyze the dependencies carefully– knowing that a cell is closed/open affects not only the distribution of points inside this cell, but also that of neighboring cells.

Lemma 4. *Let Z_i be the number of points in cell i, and let L be a lake in scenario $\mathcal{C}$ with n points. Then, for large enough number of points n, there is a constant $\gamma > 0$ such that*

$$\Pr(\sum_{i \in L} Z_i > h) \leq e^{-\gamma\sqrt{h}}.$$

Proof. Let $B := (B_1, \ldots, B_m)$ be the random vector denoting which cells are good or bad, and $b = (b_1, \ldots, b_m)$ any particular such configuration. Then

$$\begin{aligned}
\Pr(\sum_{i \in L} Z_i > h) &= \sum_k \Pr\left(\sum_{i \in L} Z_i > h \mid |L| = k\right) \Pr(|L| = k) \\
&= \sum_k \sum_b \Pr\left(\sum_{i \in L} Z_i > h \mid |L| = k, B = b\right) \Pr(B = b \mid |L| = k) \Pr(|L| = k) \\
&= \sum_k \sum_b \Pr(\sum_{i \in L} Z_i > h \mid B = b) \Pr(B = b \mid |L| = k) \Pr(|L| = k).
\end{aligned}$$

The last equality follows, since if we know B we also know the size of L. We now focus on the term $\Pr(\sum_{i \in L} Z_i > h \mid B = b)$. We will show that we can replace the variables $(Z_i \mid B = b)$ with a set of i.i.d. variables that stochastically dominate them and that obey the large deviation principle.

The Poisson process can be realized as the product of m independent Poisson processes, each operating inside a cell. This implies that if we have a set of events

E_i where each event depends only on what happens in cell i, then $\Pr(\cap_i E_i) = \prod_i \Pr(E_i)$. Thus, we have

$$\begin{aligned}\Pr(\cap_i\{Z_i = h_i\}|B = b) &= \frac{\Pr(\cap_i\{Z_i = h_i, B_i = b_i\})}{\Pr(\cap_i\{B_i = b_i\})} = \frac{\prod_i \Pr(Z_i = h_i, B_i = b_i)}{\prod_i \Pr(B_i = b_i)} \\ &= \prod_i \Pr(Z_i = h_i|B_i = b_i).\end{aligned}$$

If we define $X_i = (Z_i|B_i = \text{good})$ and $Y_i = (Z_i|B_i = \text{bad})$, it follows that $\sum_i (Z_i|B)$ has the same law of the sum of independent variables each of which is X_i or Y_i depending on whether cell i is good or bad. Let us define a collection of i.i.d. random variables W_i's each of which has the distribution of $(Z_i|Z_i > 2\alpha)$. Each W_i stochastically dominates both X_i and Y_i so that

$$\Pr(\sum_{i\in L} Z_i > h \mid B = b) \le \Pr(\sum_{i\in L} W_i > h),$$

for each configuration b. Moreover the W_i obey the large deviation principle, i.e. the probability of large deviations from the mean is exponentially small. We thus have, for $\beta < 1/E[W_1]$,

$$\begin{aligned}\Pr(\sum_{i\in L} Z_i > h) &= \sum_k \sum_b \Pr(\sum_{i\in L} Z_i > h \mid B = b) \Pr(B = b \mid |L| = k) \Pr(|L| = k) \\ &\le \sum_k \sum_b \Pr(\sum_{i\le k} W_i > h) \Pr(B = b \mid |L| = k) \Pr(|L| = k) \\ &= \sum_k \Pr(\sum_{i\le k} W_i > h) \Pr(|L| = k) \\ &= \sum_{k\le\beta h} \Pr(\sum_{i\le k} W_i > h) \Pr(|L| = k) + \sum_{k>\beta h} \Pr(\sum_{i\le k} W_i > h) \Pr(|L| = k) \\ &\le \sum_{k\le\beta h} \Pr(\sum_{i\le\beta h} W_i > h) \Pr(|L| = k) + \sum_{k>\beta h} \Pr(\sum_{i\le k} W_i > h) \Pr(|L| = k) \\ &\le \sum_{k\le\beta h} \Pr(\sum_{i\le\beta h} W_i > h) + \sum_{k>\beta h} \Pr(|L| = k) \\ &= \beta h \Pr(\sum_{i\le\beta h} W_i > h) + \sum_{k>\beta h} \Pr(|L| = k) \\ &\le \beta h\, e^{-\gamma_1 h} + \sum_{k>\beta h} e^{-\gamma_2\sqrt{k}} \le e^{-\gamma\sqrt{h}}.\end{aligned}$$

The next lemma shows that, for any given Q_i of scenario $\mathcal{A}$ components of points inside distinct lakes cannot hook up together, by-passing $\mathcal{G}$.

Lemma 5. *Let u and v be points contained in two distinct lakes of scenario $\mathcal{C}$. Unless they both belong to $\mathcal{G}$ they are disconnected.*

Proof. Let us assume by contradiction that u and v are connected in scenario $\mathcal{A}$ without being connected to $\mathcal{G}$. Since u and v belong to different lakes, they must be separated by a portion of the giant cluster. Let i be an open cell and

let v_i be a point inside it. By definition of open, if v_j is a point inside a cell j within distance $D-3$ from i, then v_i and v_j belong to the same connected component. In particular, if v_i belongs to the giant component, so does v_j. Thus, u and v must be separated by at least $2(D-3)$ good cells. But each good cell contains at least $\alpha/2$ points, and $2(D-3)\alpha/2 = 102\alpha - 3\alpha > 98\alpha = K$, which is a contradiction.

The following lemma immediately follows from Lemmas 1, 4, and 5, and concludes the proof of Theorem 1.

Lemma 6. *Consider the following event $\mathcal{E}$: at the end of Phase 1 there is a unique giant component containing at least $c\,n$ points while the remaining components are trapped inside lakes, with each lake containing at most $\varphi \log^2 n$ points. For every constant $c \in (0,1)$, and for n sufficiently large, $\Pr[\overline{\mathcal{E}}] \leq 2/n^d$ where $d = \gamma\sqrt{\varphi} - 1$ and $\gamma > 0$ is a constant.*

Proof. By Corollary 1, for every $c \in (0,1)$, there exist constants γ_1 and $\hat{K}$, such that, when the algorithm is run with parameter $K \geq \hat{K}$, the probability that, at the end of phase 1, there is no component with at least $c\,n$ points, is at most $2e^{-\gamma_1\sqrt{n}}$. By Lemma 4 and the union bound, the probability that there exists a lake with more than $\varphi \log^2 n$ points is at most $ne^{-\gamma_2\sqrt{\varphi \log^2 n}} = n^{-\gamma_2\sqrt{\varphi}+1}$. Thus $\Pr[\overline{\mathcal{E}}] \leq 2e^{-\gamma_1\sqrt{n}} + n^{-\gamma_2\sqrt{\varphi}+1} \leq 2/n^d$ for n large enough.

Therefore, by choosing φ large enough, we can bound $\Pr[\overline{\mathcal{E}}]$ with any inverse polynomial.

6 Expected and maximum degree

Let us now consider the degree of the nodes at the end of the algorithms. We first analyze the maximum degree.

Lemma 7. *With probability $1-o(1)$ the final maximum degree is $O(\log^2 n)$ with Algorithm A.*

Proof. (Sketch) For a lacustrine node the bound follows from Lemma 6, observing that before reaching the closest continental node, lacustrine nodes that increase their radius will cover at most their own lake, which is of size $O(\log^2 n)$ with high probability. We omit the proof of the bound for continental nodes, which is analogous.

For lack of space we do not give the proof of the following lemma.

Lemma 8. *With probability $1-o(1)$ the final maximum degree is $O(\log n)$ with Algorithm B.*

Lemma 9. *With both Algorithms A and B, the expected final degree of a point is bounded by a constant.*

Proof. (Sketch) We bound the degree for Algorithm A. Basically the same proof holds for Algorithm B as well. Consider first the expected degree of any lacustrine point v. Let L be the lake containing v at the end of the first Phase and let w be the point of the initial giant component $\mathcal{G}$ closest to v. By Lemma 5 the value of $k(v)$ is bounded by $1 + \sum_{i \in L} Z_i$ since in the worst case, v will capture the points in L plus w. By Lemma 4,

$$E[\sum_{i \in L} Z_i] \leq \sum_h h \Pr(\sum_{i \in L} Z_i \geq h) \leq \sum_h h e^{-\gamma\sqrt{h}} < \infty.$$

The growth of the degree of continental nodes can be bounded in a similar way, and thus we omit the proof in this extended abstract.

This ends the proof of Theorems 2 and 3. We remark that the probability that the protocols fail can be made as small as n^{-d}, for any constant $d > 0$, by a suitable choice of constants that appear in the analysis.

References

1. D. Blough and M. Leoncini and G. Resta and P. Santi The k-Neighbors Approach to Interference Bounded and Symmetric Topology Control in Ad Hoc Networks, IEEE Trans. on Mobile Computing (to appear).
2. J.-D. Deuschel and A. Pisztora. Surface order large deviations for high-density percolation. *Probability Theory and Related Fields*, 104(4):467–482, 1996.
3. G. Grimmet. *Percolation*. Springer-Verlag, 1989.
4. O. Häggström and R. Meester. Nearest neighbor and hard sphere models in continuum percolation. *Random Structures and Algorithms*, 9:295–315, 1996.
5. T. Hou and V. Li. Transmission range control in multihop packet radio networks. *IEEE Transactions on Communications*, COM-34:38–44, 1986.
6. L. Kleinrock and J. Silvester. Optimum transmission radii for packet radio networks or why six is a magic number. In *IEEE National Telecommunication Conference*, pages 431–435, 1978.
7. L. Kucera. Low degree connectivity in ad-hoc networks. In *European Symposium on Algorithms (ESA)*, 203–214, 2005.
8. T. M. Liggett, R. H. Schonmann, and A. M. Stacey. Domination by product measures. *Annals of Probability*, 25(1):71–95, 1997.
9. R. Meester and R. Roy. *Continuum Percolation*. Cambridge University Press, Cambridge, UK, 1996.
10. J. Ni and S. A. G. Chandler, Connectivity Properties of a Random Radio Network, IEE Prec. Commun. 141 (1994) 289-296.
11. H. Takagi and L. Kleinrock. Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Transactions on Communications*, COM-32(3):246–257, 1984.
12. F. Xue and P. Kumar. The number of neighbors needed for connectivity of wireless networks. *Wireless Networks*, 10(2):169–181, 2004.
13. Sencun Zhu, Sanjeev Setia, and Sushil Jajodia. LEAP: Efficient Security Mechanisms for Large-Scale Distributed Sensor Networks. In Proc. of the 10th ACM Conference on Computer and Communications Security (CCS '03), Washington D.C., October, 2003.