DIMACS Technical Report 2003-39
November 2003

The Bin-Covering Technique for Thresholding Random Geometric Graph Properties

by

S. Muthukrishnan[1,2] Gopal Pandurangan[3]

[1]Permanent Member
[2]Department of Computer Science, Rutgers University, Piscataway, NJ 08854, and AT&T Labs, USA. E-mail: muthu@cs.rutgers.edu. Supported by NSF EIA 0087022, NSF ITR 0220280 and NSF EIA 02-05116.
[3]Department of Computer Science, Purdue University, West Lafayette, IN 47907-2066, USA. This research work was initiated during the visit of the author to DIMACS in the summer of 2002. E-mail: gopal@cs.purdue.edu.

ABSTRACT

We study the emerging phenomenon of ad hoc, sensor-based communication networks. The communication is modeled by the geometric random graph model $G(n, r, \ell)$ where n points randomly placed within $[0, \ell]^d$ form the nodes, and any two nodes that correspond to points at most distance r away from each other are connected. The more widely studied $G(n, r)$ model where ℓ is held fixed at unity and $n \to \infty$ applies for dense ad hoc networks, but the $G(n, r, \ell)$ model is more detailed, and more generally applicable to modeling communication networks, sparse or dense.

We study fundamental properties of $G(n, r, \ell)$ of interest: connectivity, coverage, and routing-stretch. Our main contribution is a simple analysis technique we call *bin-covering* that we apply uniformly to get *(asymptotically) tight* thresholds for each of these properties. Typically, in the past, geometric random graph analyses involved sophisticated methods from continuum percolation theory; on contrast, our bin-covering approach is discrete and very simple, yet it gives us tight threshold bounds. Our specific results should also prove interesting to the networking community that has seen a recent increase in the study of geometric random graphs motivated by engineering ad hoc networks.

1 Introduction and Motivation

We study properties of certain random graphs. There is the classical random graph $G(n, p)$ [7] that has been studied for over 50 years now. There are many other random graphs, some motivated by physical phenomenon. For example, there are random graph models for studying social networks [18] and those for studying web graphs [1, 19] where a physical phenomenon is modeled as an appropriate random graph and understanding the structural properties of random graphs provides insights into the behavior of the underlying physical phenomenon. In this paper, we focus on the physical phenomenon due to ad hoc sensor networks, a technology in the horizon that may transform our lives significantly. The structural properties we study are motivated by engineering such networks—connectivity, coverage, routing, etc – and we prove bounds on the thresholds for emergence of these key properties in suitable random graphs.

1.1 Random Geometric Graph Models

We study the geometric random graph model $G(n, r, \ell) = (V, E)$ where n vertices uniformly and randomly placed within $[0, \ell]^d$ form the nodes[1] in V, and $(u, v) \in E$ iff $D(u, v) \leq r$, for some $0 < r \leq \ell$. The distance between points $D([a, b], [x, y]) = \max\{|x - a|, |y - b|\}$.[2] We call this the *Bernoulli model*. We also study its *Poisson* version where the number of points in $[0, \ell]^d$ is given by the Poisson distribution with mean n. The main advantage of the Poisson model is that the distribution of nodes in disjoint regions are independent; this simplifies analyses in some cases. Of our interest in this paper would be $d = 1, 2$, that is, placement of nodes along a line or on the plane, although our results can be generalized to higher dimensions.

The more widely-studied geometric random graph model is $G(n, r)$ where n nodes are distributed in $[0, 1]^d$. This model applies to *dense* random graphs by studying $n \to \infty$ while $r = f(n) \to 0$. However, the $G(n, r, \ell)$ is more detailed and general: the node density n/ℓ^2 can converge to zero, or to a constant $c > 0$, or diverge as $\ell \to \infty$, depending on the relative values of r, n, and ℓ. Thus, this model can be applied both to dense, as well as sparse networks. While there is a lot of literature on $G(n, r)$ ([4, 5, 6, 11, 12, 17, 25, 26] and see e.g. the book by Penrose [24] or the course by Aldous [3] or a short overview at [9]), work on the more-detailed model $G(n, r, \ell)$ is far fewer and more recent [29, 27, 28].

1.2 Motivation

Given the history of random geometric graphs and the nice techniques from continuum percolation used to analyze them, we believe that the study of the more detailed model $G(n, r, \ell)$ is well-justified in its own right. However, we arrived at the problems we study because of the rising motivation in networking community to understand and use threshold properties in *ad hoc sensor networks* (ASNs) using these geometric random graph models.

[1] Henceforth, a node refers to one of n vertices of the graph and a point refers to a real-valued coordinate in the d-dimensional square.

[2] Our results will easily extend to Euclidean distance with small changes in the constants.

Nodes in the geometric random graph are ad hoc elements like sensors. They are assumed to be randomly distributed on the plane. They communicate by wireless means, and r is the upper bound on the distance between any two communicating nodes. In the router networks in the Internet, we use as few routers as needed, and carefully optimize the infrastructure that connects them which is typically by wired means, and manage the network of the routers in (often centralized) careful manner. In contrast, with ASNs, the expectation is to use a lot of them, "sprinkle them liberally" in areas of interest, and let them connect to each other using wireless means, and manage them typically in a distributed manner. Thus while existing wired networks are engineered carefully – in terms of setting up routing paths, covering the area being served with careful placement of routers, etc – ASNs have an opportunity to make use of the "statistical" aspects, and be more flexibly engineered, i.e., have probabilistic coverage of the region or use several paths simultaneously to route information etc. These statistical aspects emerge from the random graph implicit in sprinkling a region with sensors and their ability to communicate by wireless means. That is the motivation for the increasing study of geometric random graph models in networking community.

Thus far, bulk of the work in networking community on random graph models for ASNs has been in the $G(n, r)$ model [8, 13, 20, 21, 30].[3] Recently, the more-detailed $G(n, r, \ell)$ model has been proposed as being more suitable for study of ASNs. In particular, in [29, 15] authors point out that in practice, sensor networks cannot be too dense, due to the problem of spatial reuse: when a node is transmitting, all the nodes within its transmitting range must be silent, in order not to corrupt the transmission. Thus $G(n, r, \ell)$ random graph may be more suitable to model ASNs.

1.3 Problems of Interest

We study three fundamental classes of problems. At the lowest level, the question is, when does a random set of nodes have the ability to broadcast and "cover" the entire region under consideration? We study this problem for two different notions of coverage. At a higher level, the question is, when does a random set of paths have a communicating path between any two pair of nodes? We study this formally as the connectivity problem. Finally, at a even higher level, how good are the communicating paths between pairs of nodes? We study this as the quality of paths with respect to the shortest path on the plane. We now present formal definitions of our problems.

1. *Physical Coverage.* There are two distinct notions of a ASN covering a given space. In the first, $area(\cup_{i=1}^{n} disc(i))$ where $area$ is the area of the region and $disc(i)$ is the set of points in the square within distance r of the node i. We call this *area coverage*. We say there is *full area coverage* if the area coverage is at least $1 - o(1)$.

   In the second, a box is said to *enclose* a connected component if no other box of smaller area contains all the nodes in the connected component inside. We define *coverage* as the area of

[3] The classical random graph model is $G(n, p)$ where there are n nodes and $(u, v) \in E$ with probability p independent of all other pairs of vertices. $G(n, p)$ is not a suitable model for communication in ASNs since each sensor node can only transmit to a bounded radial area. Furthermore, probability of existence of edges (i, j), (j, k), (i, k) is independent in $G(n, p)$, but there is clearly lot of dependence in ASNs based on the geometric distribution of i, j, k and their relative positions. Technically, analyses in $G(n, r, \ell)$ model becomes harder because of this dependence.

the largest box enclosing a connected component in $G(n, r, \ell)$. We look for threshold for *full coverage*, i.e., coverage at least $1 - o(1)$.

2. *Graph Connectivity.* When does $G(n, r, \ell)$ become connected, that is, there is a communication path between any two nodes?

3. *Route Stretch.* Given two nodes u, v in $G(n, r, \ell)$, $stretch(u, v)$ is defined as $D_G(u, v)/D(u, v)$, i.e., the ratio of the shortest distance between u and v in the graph (assuming that there is a path between u and v in $G(n, r, \ell)$, otherwise the stretch is ∞) to the norm distance on the plane. The *stretch* of $G(n, r, \ell)$ is $\max_{(u,v)} stretch(u, v)$. What is the stretch of $G(n, r, \ell)$ assuming $G(n, r, \ell)$ is connected?

Each of these formal problems is well-motivated. The two notions of coverage are implicit and inherent in several papers on geometric random graphs and percolation theory (see e.g. [24, 22]), and is widely studied as part of "coverage processes" (see e.g. [16]). The $o(1)$ term is included to ignore boundary effects in $[0, \ell]^d$. However, we do not know of any results that directly apply to the $G(n, r, \ell)$ model. Connectivity in $G(n, r)$ is among the most extensively studied problems in geometric random graphs [4, 14]. In $G(n, r)$ model, sharp threshold bounds are known. In the $G(n, r, \ell)$ model, there are gaps [29] which we will close as described later. Stretch is a well-established notion in communication graphs and has been studied in general and Euclidean graphs (see for e.g. [23]) but we do not know of relevant results for the $G(n, r, \ell)$ model. All of these problems have been empirically studied in wireless networking community [8, 13, 20, 21, 30].

1.4 Our Technical Results

Our main technical results are as follows: (We say thresholds are *asymptotically tight* if the lower bound on the threshold for the presence of a property differs from the upper bound by at most a constant; we say they are *sharp* if the upper and lower bounds match.)

1. *(Coverage).* For $d = 1$, we prove that both full coverage and connectivity occur at the *sharp* threshold of $rn \approx \ell \ln \ell$. This settles the problem left open in [29] of closing the gap between the upper and lower bounds for the threshold that differed by a factor of 2. For full area coverage, the sharp threshold is at $rn \approx (1/2)\ell \ln \ell$.

   For $d = 2$, we prove that full area coverage occurs at the sharp threshold of $r^2 n \approx (1/2)\ell^2 \ln \ell$. For full coverage, threshold is asymptotically tight with full coverage almost surely with $r^2 n > 4.34\ell^2$ and no full coverage almost surely with $r^2 n < 0.25\ell^2$.

2. *(Connectivity).* For $d = 2$, we present asymptotically tight bounds for the threshold. If $r^2 n > 2\ell^2 \ln \ell$, graph is almost surely connected, and if $r^2 n < 0.25\ell^2 \ln \ell$, graph is almost surely not connected.

3. *(Stretch.)* For $d = 2$, we show that if $r^2 n > (2.5/\alpha)\ell^2 \ln \ell$, then the stretch is at most $1 + \alpha/2$ almost surely. In fact, we show a simple local algorithm for finding such high quality paths. This is asymptotically optimal since from previous result, if $r^2 n < 0.25\ell^2 \ln \ell$ then stretch is unbounded.

All the above results are the first known (asymptotically) tight threshold bounds for $G(n, r, \ell)$, and they hold both for the Bernoulli as well as the Poisson models. One way to interpret these technical results is that in the random graph $G(n, r, \ell)$ the "regime" of connectivity is when $r^2 n = c\ell^2 \ln \ell$ for some c. In the asymptotic vicinity of the connectivity regime, interesting things happen: not only does the graph get connected, but the paths between vertices are of high quality (that is they have small stretch), and the entire domain gets covered. This is a very useful parameterization for network designers to know.

1.5 Technical Overview and Our Contribution

All our results are obtained by the following method that we call *bin covering*. We cover the region with overlapping "bins" with some overlap parameter. We rewrite the desired property as random variables in terms of gaps between points, the number of empty bins etc and optimize the overlap parameter to get the best threshold bound for desired random variables. The whole approach is more detailed than merely bucketing the domain into disjoint regions; in fact, the overlap between bins is crucially used in getting our results. The bin-covering approach consists of a precise set of steps described in detail in Section 2). It is equally useful for both Bernoulli and Poisson models. In fact, the technique is applied in a similar fashion in both models; the independence in the Poisson model simplifies analysis only a little.

Previous works on $G(n, r)$ (see [24]) used sophisticated probabilistic results from continuum percolation theory to derive their bounds. In contrast, our bin-covering technique is discrete, quite simple, and it is nearly a cook-book technique to find thresholds for many geometric random graph properties. Taking the approach based on percolation theory might give a more refined analysis in some cases, but we feel that the simplicity and wide applicability of bin-covering approach and its ability to give asymptotically tight bounds (and sharp thresholds in some cases) makes it a significant contribution.

1.6 Map

In Section 2, we study the one dimensional case. This is mainly for introducing the bin-covering technique, and we get sharp thresholds in this section. In Sections 3 we study all three problems for the two dimensional case by applying the bin-covering technique in a variety of ways. Extensions and concluding remarks are in Section 4.

2 One-Dimensional Case

To illustrate our technique of bin-covering, we first focus on one dimensional $G(n, r, \ell)$ *model*: n nodes are placed randomly on a line of length ℓ and two nodes are connected if they are within a distance of r on the line. We denote the corresponding random graph induced to be $G_1(n, r, \ell)$. We now apply our technique of bin-covering to analyze *connectivity*, *full coverage*, and *full area coverage* of $G_1(n, r, \ell)$. Our technique yields sharp thresholds for all the three properties.

Note that in the $G(n, r, \ell)$ model, the asymptotic behavior is studied as $\ell \to \infty$. Since full coverage is defined in terms of the largest connected component (i.e., the length of the line enclosing the largest connected component is almost the entire line) it turns out that the threshold for connectivity holds for full coverage also. Although full coverage is related to connectivity it is actually a distinct property: for example, unlike connectivity the presence of isolated vertices does not imply the absence of full coverage. In fact, in the two dimensional case, the connectivity and full coverage thresholds are asymptotically different (see next Section).

We assume the Bernoulli model for the following theorem, although the theorem holds for the Poisson model as well, with almost no change in the proof technique.

Theorem 2.1 *Consider the $G_1(n, r, \ell)$ model and let $r = r(\ell) = \Theta(\ell^\epsilon f(\ell))$ for some $0 \leq \epsilon < 1$ and $f(\ell)$ a function that grows strictly slower than any function of type ℓ^γ where $\gamma > 0$. Let $n = \Omega(1)$. Given any two constants $c_1 > 1 - \epsilon > c_0$,*

- *$G_1(n, r, \ell)$ is connected and has full coverage a.a.s.*[4]*, if $rn \geq c_1 \ell \ln \ell$.*
- *$G_1(n, r, \ell)$ is disconnected and has no full coverage a.a.s. if $rn \leq c_0 \ell \ln \ell$.*

Also, $G_1(n, r, \ell)$ has full area coverage a.a.s. if $rn \geq \frac{c_1}{2} \ell \ln \ell$ and has no full area coverage a.a.s if $rn \leq \frac{c_0}{2} \ell \ln \ell$.

We observe that the conditions on the magnitude of r and n are not restrictive. In fact, if $r = \Omega(\ell)$, then every node can directly transmit to most other nodes and connectedness, full coverage, and full area coverage are ensured independently of n.

Proof: We use the following technique that we call bin-covering. Choose a spacing parameter $s < r$, whose value will be fixed appropriately later. For every $i = js, 0 \leq j \leq \lceil r/s \rceil$, starting from point i from the left end, split up the line into equally spaced bins of size r - denoted by the set B_i. (We can view the line as split up by strips of length s; we number the strips with index $j, 1 \leq j \leq \lfloor \ell/s \rfloor$, starting from the left end.) Thus, there are a total of $\lceil \ell/r \rceil \lceil r/s \rceil$ bins (each of size r) where each bin overlaps with $\lceil r/s \rceil - 1$ other such bins. Let r.v. X_i, $1 \leq i \leq \lceil \ell/s \rceil$ be an indicator for an empty bin numbered i (starting from the left end). Let r.v. $X = \sum_i X_i$ denote the total number of empty bins.

It is clear that there will be full coverage and connectivity if the following two conditions are true: if there are no empty bins and there is no gap greater than r between any two consecutive nodes in the line. By our bin-covering scheme the above two conditions are ensured if

1. There are no empty bins, and
2. For every node $i, 1 \leq i \leq n - 1$ (assume that nodes are numbered according to their value on the line with ties broken arbitrarily) the following configuration does not occur: i occurs in some strip j and $i + 1$ occurs in strip $j + \lceil r/s \rceil - 1$ at a distance greater than r from it.

[4] We say that an event E_ℓ describing a property of a random structure depending on a parameter ℓ holds *asymptotically almost surely* (a.a.s.) if $\Pr(E_\ell) \to 1$ as $\ell \to \infty$.

We will use r.v. $Y_i, 1 \leq i \leq n-1$ to denote if the configuration mentioned in condition (2) occurs between nodes i and $i+1$. Let $Y = \sum_{i=1}^{n-1} Y_i$. Let the r.v $Z = X + Y$. Thus we will upper bound $E[Z]$ and show that it tends to zero for an appropriate threshold function. Then by the first moment method [2] it will follow that $P[Z=0] \to 1$, i.e., there will be full coverage a.a.s.

We have $E[X] \leq \lceil \ell/r \rceil \lceil r/s \rceil (1-r/\ell)^n$.

We upper bound $\Pr(Y_i = 1)$ by the probability that some node (say p) appears in some strip $j + \lceil r/s \rceil - 1$ (assuming i appears in strip j, for *some* j) and there is a gap of length r somewhere between i and p, i.e.,

$$\Pr(Y_i = 1) \leq (1-r/\ell)^{n-2} ns/\ell$$

And we have $E[Y] \leq (1-r/\ell)^{n-2} n^2 s/\ell$. Thus,

$$E[Z] = E[X] + E[Y] \leq \lceil \ell/r \rceil \lceil r/s \rceil (1-r/\ell)^n + n^2 s/\ell (1-r/\ell)^{n-2} \leq \frac{\ell}{s} e^{-rn/\ell} + \frac{n^2 s}{\ell} e^{-r(n-2)/\ell}$$

Plugging $rn = k\ell \ln \ell$ we have,

$$E[Z] \leq \frac{1}{s\ell^{k-1}} + \frac{k^2 (\ln \ell)^2 s}{\ell^{k-1} r^2}$$

The value of s that minimizes the above is $s = \frac{r}{k \ln \ell}$. Let $r = \ell^\epsilon f(\ell)$ as in the theorem. Then, $E(Z)$ goes to zero for $k > 1-\epsilon$ as $\ell \to \infty$.

To show the lower bound, we use the second moment method [2]. Let $\delta = \delta(\ell) = \Theta(r/(k\ell \ln \ell))$. Divide the line between $(\delta\ell, (1-\delta)\ell)$ into equally spaced bins of size r. We show that there is an empty bin in this middle part of the line almost surely if $c < 1-\epsilon$, for any fixed constant c, implying that there can not be full coverage. Let $X = \sum_i X_i$ be the total number of empty bins,as defined earlier; we show that $\Pr(X=0) \to 0$, i.e., a.a.s there is an empty bin.

$$E[X] = \frac{l(1-2\delta)}{r}(1-\frac{r}{\ell})^n$$

$$E[X^2] = E[X] + 2\sum_{i \neq j} E[X_i X_j]$$

$$= \sum_{\text{bins } i \text{ and } j \text{ don't overlap}} E[X_i X_j] + \sum_{\text{bins } i \text{ and } j \text{ overlap}} E[X_i X_j]$$

$$= (1-2r/\ell)^n (\frac{l(1-2\delta)}{r})(\frac{l(1-2\delta-r)}{r}) + \frac{l(1-2\delta)}{r}(1-r/\ell)^n$$

Since $r = \Theta(\ell^\epsilon f(\ell))$,

$$\frac{E[X^2]}{E^2[X]} \sim 1 - \frac{r}{\ell(1-2\delta)} + \frac{r}{\ell(1-2\delta)} e^{k \ln \ell}$$

which $\to 1$ if $k < 1-\epsilon$. Since, $\Pr(X=0) \leq \frac{E[X^2]}{E^2[X]} - 1$, there is at least one non empty bin in $(\delta\ell, (1-\delta)\ell)$, the lower bound for full coverage follows. To see the same for connectivity, we note that there will be at least one node a.a.s on either side of the empty bin since $(1-2\delta)^n \sim e^{-2\delta k \ln \ell / r} \sim \Theta(1)$, by our choice of δ, we conclude that the graph will be also disconnected a.a.s. for $k < 1 - \epsilon$.

We will have full area coverage if there are no empty bins and there is no gap greater than $2r$ between any two consecutive nodes in the line. Then the analysis is similar as above. $\square$

We note that we could have tried to get threshold bounds by simply dividing the line into a bins of size, say $r/2$ (for upper bound) or r (for lower lower), but this would give us worse bounds. Intuitively, it seems better to have bins of size r, but all bins being non-empty does not imply connectivity. In our bin-covering technique, we allow overlap of bins of suitable size (say r) and we use a spacing parameter to control the amount of overlap. Note that we could also have considered bins of size $r - 2s$ (instead of r) and forcing that situation that there are no empty bins. This would have spared us from invoking the r.v. Y. The reader can verify that the first moment step yields the same upper bound, but finding the best value of s (which consequently determines the threshold) becomes a matter of trial and error and it is difficult to optimize.

Bin-Covering Technique The bin-covering approach consists of the following steps (works for both the Bernoulli and the Poisson model):

1. Cover the space using many overlapping bins, the amount of overlap being determined by the spacing parameter s.

2. Define a random variable (Z as in proof of Theorem 2.1) which represents the number of gaps of desired length. Upper bound Z by the number of empty bins (X) and the number of gaps between points (Y).

3. $E[X]$ will be inversely proportional to s, while $E[Y]$ will be directly proportional to s. Thus, there is a value of s which minimizes the function. Plug the above value of s in the above upper bound and guess the form of the function of r (for example, $rn = k\ell \ln \ell$ as in the proof of Theorem 2.1).

4. Determine the best value of the threshold function which makes $E[Z]$ to go to zero asymptotically.

Applying the same technique to the one-dimensional version of the $G(n, r)$ model yields the following thresholds for connectivity, coverage, and full area coverage. We note that the same connectivity threshold also follows from the work of [10], but our derivation here is significantly simpler.

Theorem 2.2 *Given any two constants $c_1 > 1 > c_0$, $G_1(n, r)$ has full coverage and connectivity a.a.s. if $r \geq \frac{c_1 \ln n}{n}$ and no full coverage and is disconnected a.a.s. if $r \leq \frac{c_0 \ln n}{n}$. Also, $G_1(n, r)$ has full area coverage a.a.s. if $r \geq \frac{c_1 \ln n}{2n}$ and has no full area coverage a.a.s if $r \leq \frac{c_0 \ln n}{2n}$.*

3 Two-Dimensional Case

We now focus on the 2-dimensional case, i.e., $G_2(n, r, \ell)$ model (or just $G(n, r, \ell)$ for convenience). We assume that the corners of the $\ell \times \ell$ square are $(0, 0), (\ell, 0), (\ell, \ell), (0, \ell)$ (listed counterclockwise). The left, right, top, bottom sides of the square have the obvious interpretation. We assume the Bernoulli model for the proofs unless otherwise stated; the theorems also hold for the Poisson model with little change in the proofs.

We need the following concept (a similar concept is used in continuum percolation [22]). A *left-to-right (L-R) crossing* in $G(n, r, \ell)$ is a path starting from a node within a distance of $o(\ell)$ from the left side of the unit square and ending in a node within a distance of $o(\ell)$ from the right side of the square. A *Top-to-Bottom (T-B)* crossing is defined similarly. It is easy to show the following Lemma:

Lemma 3.1 *The number of nodes in any L-R crossing or T-B crossing in $G(n, r, \ell)$ is at least $\ell(1 - o(1))/r$. There is full coverage in $G(n, r, \ell)$ if and only if there is a L-R crossing and a T-B crossing.*

We also need the following. Divide the square into blocks of size $r/2 \times r/2$. Let the blocks be numbered by their row and column numbers (columns are numbered left to right and rows from top to bottom starting from 0 to $\lfloor 2/r \rfloor$). Define the following (directed) graph H: vertices are the set of empty blocks (i.e., no nodes of G in the blocks) and there is an edge from a block (i, j) to blocks $(i, j + 1)$ (right), $(i + 1, j + 1)$ (right-down), $(i + 1, j - 1)$ (right-up), $(i, j + 1)$ (down) and $(i, j - 1)$ (up). We omit the proof of the following lemma which can be shown by appealing to planar graph duality.

Lemma 3.2 *If there is no simple path of length $\lfloor 2\ell/r \rfloor - 1$ in H then there exists a T-B crossing in $G(n, r, l)$.*

3.1 Coverage

Theorem 3.1 *There is no full coverage a.a.s. in $G(n, r, \ell)$ when $r^2 n < \ell^2/4$ and $r = o(\ell)$, and there is full coverage a.a.s. when $r^2 n > (4 \ln 3)\ell^2$ and $r = o(\ell)$.*

Proof: For the first part, we show that the probability of an L-R crossing in $G(n, r, \ell)$ tends to zero as ℓ tends to ∞. We do so by constructing an appropriate branching process as explained below.

Let p be an arbitrary node within a distance of $o(\ell)$ from the left side. We note that the number of neighbors of any node is stochastically dominated by the random variable $Z \sim B(n, 4r^2/\ell^2)$. Now consider the following sequence of random variables defined by:

$Y_0 = 1$, $Y_i = Y_{i-1} + Z_i - 1$, where $Z_i \sim Z$.

Thus, Y_i stochastically upper bounds the number of neighbors in step i (in step 0, we have just node p). The Y_i's and Z_i's can be naturally interpreted as a branching process (for example, see [2]). Let T be the least t for which $Y_t = 0$ ($T = \infty$ if no such t exists). From Lemma 3.1, a necessary condition for an L-R crossing to exist in $G(n, r, \ell)$ is that the $T > \ell(1 - o(1))/r$. Since $Z_1 + \cdots + Z_t$ has a binomial distribution with mean $nt(4r^2)/\ell^2 < t$ we have by Chernoff bound,

$$\Pr(T > t) \leq \Pr(Y_t > 0) = \Pr(Z_1 + \cdots + Z_t \geq t) \leq e^{-\delta t}$$

for some constant (independent of n,r, and ℓ) $\delta > 0$. Thus the probability that an L-R crossing exists from p is at most $e^{-\ell(1-o(1))\delta/r}$; and the probability that L-R crossing exists at all (from any starting point within a distance of $o(\ell)$ from the left) is at most $ne^{-\ell(1-o(1))\delta/r} \to 0$ as $\ell \to \infty$.

For the second part, Lemma 3.2 gives a sufficient condition for the existence of a T-B crossing. The number of paths of length $\lfloor 2\ell/r \rfloor - 1$ starting from the left is upper bounded by $\frac{2\ell}{r} 3^{2\ell/r}$. Thus the expected number of paths having all of its $\lfloor 2\ell/r \rfloor$ blocks empty is

$$\leq \frac{2\ell}{r} 3^{\frac{2\ell}{r}} (1 - \frac{r}{2\ell})^n \sim e^{\ln(\frac{2\ell}{r}) + \frac{2\ell}{r} ln 3 - \frac{rn}{2\ell}}$$

Substituting $n = \frac{k\ell^2}{r^2}$, we see that the above tends to zero for a constant $k > 4 \ln 3 \approx 4.3944$. A similar argument holds good for the existence of a L-R crossing. □

We now outline how bin covering helps tighten the sufficiency condition for the existence of a T-B crossing. In what follows, we use the Poisson model for convenience.

Divide the $\ell \times \ell$ square into blocks of size $r \times r$. Fix a spacing parameter s. For every $i = js$, $0 \leq j \leq \lceil r/s \rceil$, starting from point i from the left side of the square, split up the square into equally spaced vertical bins of area $r \times \ell$. Analogous to the 1-D case, the square can be viewed as being covered by vertical slabs of area $s \times \ell$ – a total of $\lceil \ell/s \rceil$. Similarly, for every $i = js$, $0 \leq j \leq \lceil r/s \rceil$, starting from point i from the bottom side of the square, split up the square into equally spaced horizontal slabs of area $r \times \ell$. The vertical and horizontal slabs induce a total of $\lceil \ell^2/s^2 \rceil$ bins of size $r \times r$. Each such bin is partitioned into thin vertical slabs of size $r \times s$.

Define a thin vertical slab to be *almost empty* if it is sandwiched between two points which are distance r apart and one of the points is in the slab at either the top or bottom region of size $s \times s$ in the slab. Define a directed graph H as before, whose vertices are empty slabs of size $r \times s$ and an edge exists from a slab w to all empty and almost empty adjoining slabs to the right including those which touch w at the top and bottom corners, $2\lceil r/s \rceil + 1$ in all.

The following lemma is analogous to 3.2:

Lemma 3.3 *If there is no simple path of length $\lfloor \ell/s \rfloor - 1$ in H then there exists a T-B crossing.*

Theorem 3.2 *There is full coverage a.a.s. in $G(n, r, \ell)$ when $r^2 n > 1.2608\ell^2$ and $r = o(\ell)$.*

Proof: The argument is similar to the proof of Theorem 3.1. The expected number of paths in G is upper bounded by:

$$\frac{\ell}{s} (\frac{2r}{s} + 1)^{\ell/s} \left(e^{-rsn/\ell^2} (1 + (1 - e^{-s^2 n/\ell^2})^2 \right)^{\ell/s},$$

where $e^{-rsn/\ell^2}(1 + (1 - e^{-s^2 n/\ell^2})^2)$ is the probability of an empty or an almost empty vertical slab. This is decreasing function of s. Choosing $s = r$, and $r^2 n = k\ell^2$, the above expression tends to zero if $k \approx 1.2608$. □

3.2 Area Coverage

The following theorem shows a threshold for full area coverage in two dimensions. Again, dividing the square into $r \times r$-size bins (totaling ℓ^2/r^2 approximately) and arguing that every bin should contain a node gives a weaker bound. On the other hand, if we use bins of size $2r \times 2r$, then even if every bin contains a node, this does not imply full coverage.

Theorem 3.3 *Consider the $G(n, r, \ell)$ model and let $r = r(\ell) = \Theta(\ell^\epsilon f(\ell))$, for some $0 \leq \epsilon < 1$, and $f(\ell)$ is a function which grows strictly slower than any function of type ℓ^γ where $\gamma > 0$. Let $n = \Omega(1)$. Given any two constants $c_1 > \frac{1}{2} - \frac{1}{2}\epsilon > c_0$,*

- *$G(n, r, \ell)$ has full coverage a.a.s. if $r^2 n \geq c_1 \ell^2 \ln \ell$.*
- *$G(n, r, \ell)$ does not have full coverage a.a.s. if $r^2 n \leq c_0 \ell^2 \ln \ell$.*

Proof: We show that we can tightly cover the area by squares of dimension $2r \times 2r$ in such a way there is no "hole" of dimension $2r \times 2r$. This will ensure that every point in the unit square is covered by some node.

Divide the $\ell \times \ell$ square into blocks of size $2r \times 2r$. Fix a spacing parameter s. For every $i = js$, $0 \leq j \leq \lceil r/s \rceil$, starting from point i from the left side of the square, split up the (entire) square into equally spaced vertical bins of size $\ell \times 2r$. Analogous to the one dimensional case, the square can be viewed as being covered by vertical slabs of size $\ell \times s$ – a total of $\lceil \ell/s \rceil$. Similarly, for every $i = js$, $0 \leq j \leq \lceil r/s \rceil$, starting from point i from the bottom side of the square, split up the square into equally spaced horizontal bins of area $2r \times \ell$. Similarly, the square can also be viewed as being covered by horizontal slabs of size $s \times \ell$ – a total of $\lceil \ell/s \rceil$. The vertical and horizontal slabs induce a total of $\lceil \ell^2/s^2 \rceil$ bins of size $2r \times 2r$.

There will be full coverage if (1) there are no empty bins among all the $\lceil \ell^2/s^2 \rceil$ bins (2) the following configuration does not occur: given nodes u, v, w, x such that u occurs in slab b_1, v occurs in slab b_2 such that $|b_1 - b_2| = \lceil 2r/s \rceil - 1$ and w occurs in (vertical) bin b_3 and x occurs in bin b_4, such that $|b_3 - b_4| = \lceil 2r/s \rceil - 1$ and there is an empty square bin of size $2r \times 2r$ between the points (i.e., the line joining u, v and w, x goes through the empty bin).

Let the random variable X denote the total number of empty bins. Let random variable Y_i, $1 \leq i \leq \binom{n}{2}$ be the indicator for the event of (2) between a pair of points. Let $Z = X + Y$. Then

$$E[X] \leq \frac{\ell^2}{s^2}(1 - \frac{4r^2}{\ell^2})^n$$

and

$$E[Y] \leq n(n-1)(n-2)(n-3)(8rs/\ell^2)(2rs/\ell^2)(4r/s)(2rs/\ell^2)(1 - 4r^2/\ell^2)^{n-4}$$

Choosing $s = \frac{\ell^2}{n(128)^{1/4} r}$ (found by minimizing $E[Z]$ as a function of s) and letting $r^2 n = k\ell^2 \ln \ell$, shows that the expectation tends to zero as $\ell \to \infty$, for any constant $k > \frac{1}{2} - \frac{1}{2}\epsilon$.

The lower bound is shown by applying the second moment method – we omit the details here – the approach is similar to Theorem 2.2. $\square$

We can derive the corresponding theorem for $G(n, r)$ similarly:

Theorem 3.4 *Given any two constants $c_1 > 1/4 > c_0$, $G(n, r)$ has full coverage a.a.s. if $r \geq \sqrt{\frac{c_1 \ln n}{n}}$ and no full coverage a.a.s. if $r \leq \sqrt{\frac{c_0 \ln n}{n}}$.*

3.3 Connectivity

The following theorem establishes bounds on the connectivity for $G(n, r, \ell)$.

Theorem 3.5 *Consider the $G(n, r, \ell)$ model and let $r = r(\ell) = \Theta(\ell^\epsilon f(\ell))$, for some $0 \leq \epsilon < 1$, and $f(\ell)$ is a function which grows strictly slower than any function of type ℓ^γ where $\gamma > 0$. Let $n = \Omega(1)$. Given any two constants $c_1 > 2 - 2\epsilon$ and $c_0 < \frac{1}{2} - \frac{1}{2}\epsilon$,*

- *$G(n, r, \ell)$ is connected a.a.s. if $r^2 n \geq c_1 \ell^2 \ln \ell$, and*
- *$G(n, r, \ell)$ is disconnected a.a.s. if $r^2 n \leq c_0 \ell^2 \ln \ell$.*

Proof: The graph is connected if we can tightly cover the whole area by $r \times r$ bins such that there is no "hole" of size $r \times r$. The analysis is similar to that of full area coverage, except that we cover by bins of size $r \times r$ instead of $2r \times 2r$. This yields the upper threshold for connectivity. A necessary condition for connectivity is the absence of isolated nodes in $G(n, r, \ell)$. Let r.v. X denote the number of isolated nodes. Applying the second moment method to enforce that $\Pr(X = 0) \to 1$ yields the lower threshold for connectivity. Details are omitted. □

In this case, even though bin covering gives asymptotically tight bounds, it does not yield sharp thresholds. Is the necessary condition , i.e., the absence of isolated vertices in $G(n, r, \ell)$, also a sufficient condition for connectivity of $G(n, r, \ell)$? We conjecture that it is true (similar relationship exists in the $G(n, r)$ model[4, 25, 26]). The above theorem, however, establishes the asymptotic connectivity regime.

3.4 Stretch

We study the stretch of $G(n, r, \ell)$ in the (asymptotic) connectivity regime, i.e., $r^2 n \geq k\ell^2 \ln \ell$ for some constant $k > \frac{1}{2} - \frac{1}{2}\epsilon$. This is reasonable since the stretch is unbounded if $G(n, r, \ell)$ is not connected.

First, we show the following theorem which is a consequence of the approach used in the connectivity threshold.

Theorem 3.6 *Consider the $G(n, r, \ell)$ model with parameters as defined in Theorem 3.5 and let $r^2 n \geq k\ell^2 \ln \ell$, where $k > 2 - 2\epsilon$. Then the stretch of $G(n, r, \ell)$ is at most 2.*

Proof: Here is a sketch. In the proof of Theorem 3.5, we show that we can tightly cover the unit square by squares of dimension $r \times r$ in such a way there is no "hole" of dimension $r \times r$. This will ensure that between any two nodes u, v we can construct a path which has distance at most two times $d(u, v)$, since we can construct a path from u to v which will advance a distance of r by traveling at most $2r$ through the nodes. □

The following theorem shows the existence of *high quality paths* (i.e. stretch is at most $1+\epsilon$) between any two nodes in the network for a value of k (as in $r^2n = k\ell^2 \ln \ell$) only slightly larger than that needed for connectivity. The theorem actually gives a tradeoff result between k and stretch.

Theorem 3.7 *In $G(n, r, \ell)$ and let $r^2n = k\ell^2 \ln \ell$, and $r = r(\ell) = \Theta(\ell^\epsilon f(\ell))$, for some $0 \le \epsilon < 1$, as before. Let $0 < \alpha \le 1$ be a fixed constant. Then for any constant $k > \frac{5}{2\alpha}$, the stretch is $1 + \alpha/2$ a.a.s. Further, if we consider only the subset F of nodes such that $D(u, v) = \omega(r)$ for all $u, v \in F$ then the stretch restricted to this subset is 1 a.a.s.*

Proof: We show that there is always a path of length less than $D(u, v) + \alpha/2$, for every pair of vertices u and v.

For a given pair of nodes u, v we apply the bin covering technique as follows. Consider the line joining u and v and let the angle made by this line to the horizontal be less than 45 degrees i.e., $D(u, v)$ is determined by the difference in the horizontal coordinates (the proof will be similar if this angle is greater than 45 degrees with the width and length of the boxes interchanged). We cover this line by equally spaced bins (boxes) of dimensions $\alpha r/2 \times r/2$ with spacing parameter s – a total of $\lceil D(u, v)/s \rceil$ bins. We call this a *strip* of width α.

Assume that there is no empty box and there is no "hole" of dimension $\alpha r/2 \times r/2$ in the strip, i.e., there is no configuration such that we can slide an empty box of dimension $\alpha r/2 \times r/2$ between consecutive points. We now show that there will be a path between u and v of length less than $D(u, v) + \alpha$. We construct such a path $P : u = x_0, x_1, \ldots, x_{k-1}, v = x_k$ as follows: x_i is the neighbor of x_{i-1} which is closest to v and is contained in a box. We show by induction that for $0 \le j < k - 1$, $D(x_j, x_{j+1}) \ge r/2$. Clearly $D(x_0, x_1) \ge r/2$. Assume $D(x_{j-1}, x_j) \ge r/2$ for $2 \le j \le i$. Then $D(x_i, x_{i+2}) \ge r$ (otherwise x_{i+2} will be chosen as the neighbor of x_i). If $D(x_i, x_{i+1}) < r/2$ then $D(x_{i+1}, x_{i+2}) > r/2$. Since there is no empty box (or a hole) of length $r/2$ between x_{i+1}, x_{i+2} (by our assumption), there must be a node (say y) of within a distance of $r/2$ of x_{i+1} implying that $D(x_i, y) < r$. This means that y is a neighbor of x_i which closer to v that x_{i+1}, contradicting the choice of x_{i+1}. Note that since the width of all the boxes is less than $r/2$, $D(x_i, x_{i+1})$, for $0 \le i < k-1$ is the horizontal distance covered by the path towards v. Thus the length of P is bounded by the $D(u, v) + D(x_{k-1}, v) \le D(u, v) + \alpha r/2$. Since, $D(u, v) \ge r$ (otherwise the stretch is 1) the stretch of P is bounded by $1 + \alpha/2$. If $D(u, v) = \omega(r)$ then stretch is $1 + o(1)$.

Repeating this argument for all $\binom{n}{2}$ pairs of nodes we bound the total number of empty bins and the number of "hole" configurations:

$$\binom{n}{2} \left(\frac{max_{(u,v)} D(u, v)}{s} (1 - \frac{\alpha r^2}{4\ell^2})^n + (1 - \frac{\alpha r^2}{4\ell^2})^n \frac{n^2 \alpha s r}{2\ell^2} \right)$$

As usual, we choose the best value of s and letting $r^2 = k\ell^2 \ln \ell$, the RHS tends to zero as $\ell \to \infty$ if $k > \frac{5}{2\alpha}$. □

A byproduct of the proof of the above theorem (i.e., in computing the path P) is that it gives an efficient and *local* routing algorithm for finding a path of low stretch between any two nodes in the sensor network. We note that the algorithm is optimal with respect to space and number of sensors visited and can be easily implemented in a local distributed fashion.

Corollary 3.1 *Let $G(n, r, \ell)$ have parameters as defined in Theorem 3.7. Then given two nodes u and v, there is a local algorithm that (a.a.s.) finds a path between u and v of stretch at most $1 + \alpha/2$ that takes $O(1)$ space (per vertex), and visits $2D(u, v)/r$ vertices altogether.*

4 Concluding Remarks

We introduced the bin-covering technique to analyze threshold bounds for fundamental properties like coverage, connectivity and stretch of geometric random graph $G(n, r, \ell)$. Even though the technique is simple, it gives first known (asymptotically) tight threshold bounds for all these properties. In contrast, most previous analyses of geometric random graphs like $G(n, r)$ used continuum percolation methods which are more involved. Our bin-covering technique may be used to analyze other properties of interest, e.g. the number of multiple paths of high quality and the threshold for dominating sets.

To what extent is $G(n, r, \ell)$ a faithful model for ASNs? One of the biggest disadvantages of $G(n, r, \ell)$ is that in practice, an ad hoc network need not be uniformly distributed; also, r may differ from node to node. Any model with suitable non-uniform spatial distribution of nodes would be of great interest. Nevertheless, $G(n, r)$ and more generally $G(n, r, \ell)$ are used in networking community to give rough parameters to understand ASNs. There has been a surge of results in the networking community on analyses and empirical analyses of geometric random graphs for connectivity, dominating sets, etc. and a spurt of distributed, local algorithms. We believe our results will be of great interest to them.

References

[1] W. Aiello, F. Chung-Graham and L. Lu. Random evolution of massive graphs, in *Handbook on Massive Data Sets*, (Eds. James Abello et al.), Kluwer Academic Publishers, (2002), 97-122. Extended abstract appeared in FOCS 2001, 510-519.

[2] N. Alon and J. Spencer. *The Probabilistic Method*, John Wiley, 2000.

[3] D. Aldous. http://stat-www.berkeley.edu/users/aldous/Networks/ Spring 2003.

[4] M. Appel and R. Russo. The connectivity of a graph on uniform points on $[0, 1]^d$, *Statistics and Probability Letters*, **60**, 351-357.

[5] M. Appel and R. Russo. The Maximum Vertex Degree of a Graph on Uniform Points in $[0, 1]^d$, *Adv. Appl. Prob.*, 567-581, 1997.

[6] M. Appel and R. Russo. The Minimum Vertex Degree of a Graph on Uniform Points in $[0, 1]^d$, *Adv. Appl. Prob.*, 582-594, 1997.

[7] B. Bollobas. *Random Graphs*, Academic Press, 1985.

[8] B. Deb, S. Bhatnagar and B. Nath. A topology discovery algorithm for sensor networks with applications to network management. *Tech Report DCS-TR 441*, Rutgers Univ., `athos.rutgers.edu/dataman/papers.html#top`, 2002.

[9] J. Diaz, J. Petit, and M. Serna. Random Geometric Problems on $[0, 1]^2$, *Randomization and Approximation Techniques in Computer Science*, **1518**, Lecture Notes in Computer Science, 294-306, 1998.

[10] L. Devroye. A Log Log Law for maximal uniform spacings. *Ann. Probab.*, **26**, 67-80.

[11] J. Diaz, M. Penrose, J. Petit, and M. Serna. Linear ordering of random geometric graphs, *Graph Theoretic Concepts in Computer Science*, **1665**, Lecture Notes in Computer Science, 1999.

[12] A. Flaxman, A. Frieze, and E. Upfal. Efficient Communication in an Ad-hoc Network. Preprint, 2003.

[13] D. Ganesan, R. Govindan, S. Shenker and D. Estrin. Highly resilient, Energy-efficient Multipath Routing in Wireless Sensor Networks. *Mobile Computing and Communication Review*, Vol 1, No. 2, 2002.

[14] P. Gupta and P. Kumar. Critical Power for Asympototic Connectivity in Wireless Networks, in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, W. M. McEneaney, G. Yin, and Q. Zhang (Eds.), Birkhuser, Boston, 1998.

[15] P. Gupta and P. R. Kumar, The capacity of wireless networks. *IEEE Transactions on Information Theory*, **46**(2), 2000, pp. 388-404.

[16] P. Hall. *Introduction to the Theory of Coverage Processes*, John Wiley, 1988.

[17] P. Jaillet. On Properties of Geometric Random Graph Problems in the Plane, *Annals of Operations Research*, **61**, 1-20, 1995.

[18] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. in *Proc. 32nd ACM Symposium on Theory of Computing*, 2000.

[19] J. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. The Web as a graph: Measurements, models and methods. Invited survey at the *5th Annual International Computing and Combinatorics conference (COCOON)*, 1999.

[20] B. Krishnamachari, S. Wicker, R. Bejar and M. Pearlman. Critical density thresholds in distributed wireless networks. in *Proc. of GLOBECOM*, 2001. Enlarged version to appear in *Communications, Information and Network Security*, eds. H. Bhargava, H.V. Poor, V. Tarokh, and S. Yoon, Kluwer Publishers, 2002.

[21] S. Meguerdichian, F. Koushanfar, M. Potkonjak and M. Srivastava. Coverage problems in wireless ad-hoc sensor networks. In *Proc. of INFOCOM* 2001, 1380-1387.

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

[23] G. Narasimhan and M. Smid. Approximating the Stretch Factor of Euclidean Graphs, *SIAM Journal of Computing*, **30**, 978-989, 2000.

[24] M.D. Penrose. *Random Geometric Graphs*, Oxford University Press, 2003.

[25] M.D. Penrose. The longest edge of the random minimal spanning tree, *Annals of Applied Probability*, **7**, 1997, 340-361.

[26] M. Penrose. On k-Connectivity for a Geometric Random Graph. *Random Structures and Algorithms*, **15**, 145-164, 1999.

[27] T. Phillips, S. Panwar and A. Tantawi. Connectivity properties of a packet radio network model. *IEEE Trans on Information Theory*, 35(5), 1044–1047, 1989.

[28] P. Piret. On the connectivity of radio networks. *IEEE Trans on Information Theory*, 37(5), 1490–1492, 1991.

[29] P. Santi and D. Blough. The Critical Transmitting Range for Connectivity in Sparse Wireless Ad Hoc Networks, *IEEE Transactions on Mobile Computing*, **2**(1), 2003, pp. 25-39.

[30] A. Tsirigos, Z. Haas, S. Tabrizi. Multipath routing in mobile adhoc networks or how to route in the presence of topological changes. *IEEE MILCOM*, 2001.