Sharp threshold for embedding balanced spanning trees in random geometric graphs

(Extended abstract)

Alberto Espuny Díaz[*]   Lyuben Lichev[†]   Dieter Mitsche[‡]
Alexandra Wesolek[§]

Abstract

Consider the random geometric graph $\mathcal{G}(n, r)$ obtained by independently assigning a uniformly random position in $[0, 1]^2$ to each of the n vertices of the graph and connecting two vertices by an edge whenever their Euclidean distance is at most r. We study the event that $\mathcal{G}(n, r)$ contains a spanning copy of a balanced tree T and obtain sharp thresholds for these events. Our methods provide a polynomial-time algorithm for finding a copy of such trees T above the threshold.

1   Introduction

The *random geometric graph* $\mathcal{G}(n, r)$ is a classic model of random graphs defined as follows. Let n be a positive integer, and let r be a positive real number. The vertices of the graph

[*]Institut für Mathematik, Technische Universität Ilmenau, 98684 Ilmenau, Germany.   E-mail: alberto.espuny-diaz@tu-ilmenau.de.   Supported by the Carl-Zeiss-Foundation and by DFG grant PE 2299/3-1.

[†]Institut Camille Jordan, University Jean Monnet, Saint-Etienne, France.   E-mail: lyuben.lichev@univ-st-etienne.fr.

[‡]IMC, Pont. Univ. Católica, Chile and Institut Camille Jordan, Univ. Jean Monnet, Saint-Etienne, France. E-mail: dieter.mitsche@mat.uc.cl. Supported by grant GrHyDy ANR-20-CE40-0002 and by Fondecyt grant 1220174.

[§]Université Lyon, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69622 Villeurbanne, France. E-mail: agwesole@sfu.ca.

are n points sampled uniformly at random and independently from $[0,1]^2$, and two vertices are connected by an edge if their Euclidean distance is at most r. Since their introduction by Gilbert [7] as a model for telecommunication networks, random geometric graphs have received a lot of attention from both applied and theoretical points of view [1, 6, 11, 13, 14].

1.1 Thresholds in random geometric graphs

Random geometric graphs are known to exhibit *threshold* behavior for many graph properties, meaning that there are some special values of the parameters of the model around which a drastic change in the behavior of the graph with respect to these properties takes place. Formally, a function $r^* = r^*(n)$ is a *threshold* for some monotone increasing property $\mathcal{P}$ in $\mathcal{G}(n,r)$ if

$$\lim_{n \to \infty} \mathbb{P}[\mathcal{G}(n,r) \in \mathcal{P}] = \begin{cases} 0 & \text{if } r = o(r^*), \\ 1 & \text{if } r = \omega(r^*). \end{cases}$$

Moreover, we say that a function $r^* = r^*(n)$ is a *sharp threshold* for $\mathcal{P}$ if, for every $\varepsilon \in (0,1)$,

$$\lim_{n \to \infty} \mathbb{P}[\mathcal{G}(n,r) \in \mathcal{P}] = \begin{cases} 0 & \text{if } r \leq (1-\varepsilon)r^*, \\ 1 & \text{if } r \geq (1+\varepsilon)r^*. \end{cases}$$

Goel, Rai and Krishnamachari [8] gave a general upper bound for the threshold width in $\mathcal{G}(n,r)$, and Bradonjić and Perkins [2] characterized vertex-monotone properties which exhibit a sharp threshold. While the results in [2, 8] serve to prove the existence of (sharp) thresholds, they give no indication of where these thresholds actually are. Determining the (sharp) thresholds for different properties of interest is one of the main problems in the area, and it has received much attention. In this extended abstract, our aim is to determine the (sharp) threshold for the appearance of different spanning trees.

There are some results about thresholds which are closely related to our problem. The sharp threshold for connectivity was determined independently by Gupta and Kumar [9] and Penrose [12], who proved that it is $\sqrt{\log n / \pi n}$. Díaz, Mitsche and Pérez-Giménez [4] obtained the sharp threshold for $\mathcal{G}(n,r)$ to contain a Hamilton cycle, which coincides with the sharp threshold for connectivity. In particular, this gives the sharp threshold for the containment of a *spanning path*.

The question of trying to find the (sharp) threshold for the appearance of different families of spanning trees pops up naturally. The results about the sharp threshold for Hamiltonicity can be used to deduce thresholds for path-like trees. Indeed, using the triangle inequality, for any fixed $k \geq 2$, one may show that the threshold for Hamiltonicity is also a threshold for the property that $\mathcal{G}(n,r)$ contains the k-th power of a Hamilton cycle. It immediately follows that every spanning tree which can be embedded into the k-th power of a Hamilton cycle has this same threshold. This is the case, for instance, of spanning caterpillars with constant maximum degree.

One may naturally wonder whether all spanning trees with bounded maximum degree have the same threshold. Incidentally, in the model of binomial random graphs $\mathcal{G}(n,p)$ where

each of the $\binom{n}{2}$ possible edges appears independently with probability p, Montgomery [10] proved that this is the case: the threshold for all bounded degree spanning trees is $\log n / n$. However, this turns out to be very far from the truth in random geometric graphs. Indeed, there are bounded degree trees T whose diameter is logarithmic in the number of vertices, and this diameter directly imposes a much higher lower bound on the threshold r^* for the property of containing a copy of T: since a spanning subgraph of $\mathcal{G}(n,r)$ cannot have smaller diameter than $\mathcal{G}(n,r)$ itself, the threshold for trees with diameter $O(\log n)$ must satisfy $r^* = \Omega(1/\log n)$, which is far larger than the connectivity threshold mentioned above. The results of Goel, Rai and Krishnamachari [8] imply that, for any such tree, there is a sharp threshold. Our goal is to determine the value of this threshold.

Out of all trees with logarithmic diameter, binary trees are especially interesting due to their many applications as data structures (see, e.g., [3]). Identifying the sharp threshold for embedding these trees into $\mathcal{G}(n,r)$ is thus an important particular case of our study.

1.2 Main results

An *s-ary tree* is a rooted tree where every node has at most s children. We say that an s-ary tree is *balanced* if there is an integer h such that all vertices at (graph) distance at most $h-1$ from the root have exactly s children, and all vertices at distance h from the root are leaves. Our first result determines the sharp threshold for $\mathcal{G}(n,r)$ to contain a spanning copy of the balanced s-ary tree for any fixed integer $s \geq 2$.

Theorem 1. *Fix an integer $s \geq 2$. Let h be a positive integer, and set $n := \sum_{i=0}^{h} s^i$. Let T_h be the balanced s-ary tree of height h (and on n vertices). Then, $r^* := 1/\sqrt{2}h$ is the sharp threshold for $\mathcal{G}(n,r)$ to contain a copy of T_h.*

In fact, Theorem 1 is a particular case of a similar result for a larger class of trees. We may think of the vertices of the rooted tree T as being partitioned into *layers* $V_0, V_1, \ldots$, where V_i contains all vertices at (graph) distance i from the root. For any vertex $v \in V(T)$, if $v \in V_i$, we refer to its neighbors in V_{i+1} as its *children*, and to all vertices which can be reached by a path from v without going through V_{i-1} as its *descendants*. The *height* of the rooted tree is the maximum (graph) distance between the root and another vertex of T. Given positive integers h and $(s_i)_{i=1}^{h}$, we say that a tree T is the *balanced tree over the sequence* $(s_i)_{i=1}^{h}$ if it has height h and, for each $i \in \{1, \ldots, h\}$, every vertex of T in V_{i-1} has exactly s_i children. In particular, such a tree T contains exactly $\sum_{i=0}^{h} \prod_{j=1}^{i} s_j$ vertices (where, by convention, the empty product equals 1). If $2 \leq s_i \leq M$ for all $i \in \{1, \ldots, h\}$ and for some positive integer M, we say that T is a *balanced M-tree*.

For a graph G, we denote by $|G|$ the size of the vertex set of G. As typical in random graphs literature, we focus on asymptotic statements. Our asymptotic notations will be taken with respect to the height of the trees, which then also yields asymptotic results with respect to n. Our next result extends Theorem 1 to all balanced M-trees as long as M is not too large compared to h.

Theorem 2. *Let $2 \leq M = M(h) = o(h/\log h)$. Let $(T_h)_{h \geq 1}$ be a sequence of trees where T_h is a balanced M-tree of height h. Then, $r^* := 1/\sqrt{2}h$ is the sharp threshold for $\mathcal{G}(|T_h|, r)$ to contain a copy of T_h.*

We first provide here the (much easier) proof of the lower bound of $1/\sqrt{2}h$ in Theorem 2. Note that a.a.s. (that is, with probability tending to 1 as $h \to \infty$) there are vertices $u, v \in V(\mathcal{G}(|T_h|, r))$ at Euclidean distance $(1 - o(1))\sqrt{2}$ from each other. Indeed, a.a.s. the squares $c_0, c_1 \subseteq [0, 1]^2$ of side length $1/h$ containing the corners $(0, 0)$ and $(1, 1)$, respectively, each contain at least a vertex. Conditioning on this event, assume that G admits T_h as a spanning tree. Since T_h has diameter $2h$, u and v must be at distance at most $2h$ in G. Thus, by the triangle inequality, the Euclidean distance between them must be at most $2hr$, so we must have $2hr \geq \text{dist}(c_0, c_1) = (1 - o(1))\sqrt{2}$.

The proof of the upper bound is much more involved. In order to simplify our exposition, in the next section we provide a sketch of the proof only for the case of balanced binary trees. This particular case already contains most ideas of the more general theorem. For all the details of the proof, we refer the reader to the full version of our paper [5].

2 Embedding algorithm for the upper bound

Fix $\varepsilon \in (0, 1)$. Let T be a balanced binary tree of height h, and set $n := 2^{h+1} - 1$. Let $G = \mathcal{G}(n, r)$. Suppose that $r \geq (1 + \varepsilon)r^*$. We embed the layers $V_0, V_1, \ldots, V_h$ of T into G one at a time, starting from the root. Throughout, we refer to the embedding of one of the layers of T as a *step* of the algorithm (for simplicity, we assume that V_0 is embedded at the 0-th step). For each $i \in \{0, \ldots, h\}$, at the end of the i-th step, we call a vertex of G into which a vertex on layer V_i has been embedded *active*. Moreover, we refer to the vertices of G into which no vertex of T has been embedded as *unseen*. For simplicity of notation, once a vertex of T has been embedded into G, we often interchangeably use the same notation to refer to either of the two vertices.

Let k be the smallest integer which satisfies that

$$2^{1/2-k} < \frac{\varepsilon r^*}{8}. \tag{1}$$

Let $\mathcal{S}$ be the tessellation of $[0, 1]^2$ into 2^{2k} congruent closed axis-parallel squares. We first focus on embedding the layers $V_0, \ldots, V_{m+2}$ into G, where m will be defined below, in such a way that the vertices in V_{m+2} are distributed "sufficiently uniformly" in $[0, 1]^2$. To be more precise, we ensure that each square of $\mathcal{S}$ contains 2^{m+2-2k} vertices from V_{m+2}. We then finish the embedding of the remaining layers with a suitable application of Hall's theorem.

For each $\ell \in \{0, \ldots, k\}$, let $\mathcal{S}_\ell$ be the tessellation of $[0, 1]^2$ into $2^{2\ell}$ congruent closed axis-parallel squares obtained by combining the squares of $\mathcal{S}$ into groups of size $2^{2(k-\ell)}$. In particular, $\mathcal{S}_0 = \{[0, 1]^2\}$ and $\mathcal{S}_k = \mathcal{S}$. For a square $q \subseteq [0, 1]^2$, we denote its center by $c(q)$. Moreover, for every $i, j \in \{0, \ldots, k\}$ with $i < j$ and any square $q \in \mathcal{S}_i$, let $\sigma_j(q)$ denote the set of four subsquares of q in $\mathcal{S}_j$ that form the axis-parallel square of side length 2^{1-j}

and center $c(q)$. In particular, $\sigma_{i+1}(q)$ is a tessellation of q into four subsquares in $\mathcal{S}_{i+1}$, and for each $j \in \{i+2, \dots, k\}$, the set $\sigma_j(q)$ is obtained from $\sigma_{j-1}(q)$ by homothety with center $c(q)$ and ratio $1/2$. Note that we sometimes abuse notation and identify $\sigma_j(q)$ with its geometric realization; in particular, we identify $\bigcup_{p \in \sigma_j(q)} p$ with $\sigma_j(q)$ itself.

To begin with, we embed the layers V_0 and V_1 of T into an arbitrary square $q_0 \in \sigma_k([0,1]^2)$, and then the four vertices in V_2 are evenly distributed among the four squares in $\sigma_k([0,1]^2)$. Our algorithm for embedding $V_3, \dots, V_h$ has two main parts, which we call *subroutines*. The first subroutine is used to embed layers $V_3, \dots, V_{m+2}$, while the second subroutine deals with the remaining layers.

2.1   The first subroutine

The m steps of this subroutine are grouped into $k-1$ different blocks. For each $\ell \in \{1, \dots, k-1\}$, we proceed iteratively as follows. Suppose that at the beginning of the ℓ-th block we have a configuration in which every square $q \in \mathcal{S}_{\ell-1}$ contains the same number of active vertices, and that these are equally distributed among all subsquares in $\sigma_k(q)$ (note that this is verified in the case $\ell = 1$). Then, for each $q \in \mathcal{S}_{\ell-1}$, we proceed as follows.

Iteration: We proceed to distributing the descendants of the currently active vertices in a way that we embed them at increasing distances from $\sigma_k(q)$ as follows.

Define $\phi_\ell \colon \sigma_k(q) \to \sigma_\ell(q)$ as the bijection obtained by homothety with center $c(q)$ and ratio $2^{k-\ell}$. To each square $p \in \sigma_k(q)$ we associate a sequence of squares $(p_1, \dots, p_t)$ in $\mathcal{S}_k$, for some appropriately chosen t which does not depend on p, which satisfies that $p_1 = p$, $p_t \in \sigma_k(\phi_\ell(p))$, and for all $i \in \{1, \dots, t-1\}$ we have $\|c(p_{i+1}) - c(p_i)\| \leq (1 + 7\varepsilon/8)\, r^*$. Note that the last condition together with the triangle inequality and (1) ensures that, for every $i \in \{1, \dots, t-1\}$ and every choice of points $x \in p_i$ and $y \in p_{i+1}$, we have

$$\|x-y\| \leq \|x-c(p_i)\| + \|c(p_i)-c(p_{i+1})\| + \|c(p_{i+1})-y\| \leq \frac{\varepsilon}{16}r^* + \left(1 + \frac{7\varepsilon}{8}\right)r^* + \frac{\varepsilon}{16}r^* = r. \quad (2)$$

Hence, if there is an active vertex v in p_i, it is possible to embed all children of v in p_{i+1}.

Let us prove that such sequences of squares can indeed be constructed. Fix $p \in \sigma_k(q)$ as above. We begin by setting $p_1 := p$. For each $i \geq 1$, while $\|c(p_i) - c(\phi_\ell(p))\| > (1 + 7\varepsilon/8)r^*$, we choose a square $p_{i+1} \in \mathcal{S}_k$ such that

$$\|c(p_{i+1})-c(\phi_\ell(p))\| \leq \|c(p_i)-c(\phi_\ell(p))\| - \left(1 + \frac{5\varepsilon}{8}\right)r^* \text{ and } \|c(p_{i+1})-c(p_i)\| \leq \left(1 + \frac{7\varepsilon}{8}\right)r^*. \quad (3)$$

Finally, when we reach a point where $\|c(p_i) - c(\phi_\ell(p))\| \leq (1 + 7\varepsilon/8)r^*$, we set $t := i+1$ and choose p_t as a square in $\sigma_k(\phi_\ell(p))$ at smallest distance to $c(p_i)$. Note that (2) holds for $i = t-1$ and this choice of p_t. It remains to prove that a choice as prescribed by (3) can indeed be made.

Assume we have already defined p_i so that it satisfies $\|c(p_i) - c(\phi_\ell(p))\| > (1 + 7\varepsilon/8)r^*$. Then, choose $p_{i+1} \in \mathcal{S}_k$ to be a square containing the point w on the segment $c(p_i)c(\phi_\ell(p))$

at distance exactly $(1 + 3\varepsilon/4)r^*$ from $c(p_i)$ (if more than one such square exists, we choose one arbitrarily). Then, the triangle inequality and (1) imply that

$$
\begin{aligned}
\|c(p_{i+1}) - c(\phi_\ell(p))\| &\leq \|w - c(\phi_\ell(p))\| + \|c(p_{i+1}) - w\| \\
&= \|c(p_i) - c(\phi_\ell(p))\| - \|c(p_i) - w\| + \|c(p_{i+1}) - w\| \\
&\leq \|c(p_i) - c(\phi_\ell(p))\| - \left(1 + \frac{5\varepsilon}{8}\right)r^*
\end{aligned}
$$

and

$$
\|c(p_i) - c(p_{i+1})\| \leq \|c(p_i) - w\| + \|w - c(p_{i+1})\| \leq \left(1 + \frac{3\varepsilon}{4}\right)r^* + \frac{\varepsilon}{16}r^* \leq \left(1 + \frac{7\varepsilon}{8}\right)r^*,
$$

so (3) is verified.

Now, once the sequence $(p_1, \ldots, p_t)$ is constructed, we can describe the distribution of the vertices of the tree in each step of the ℓ-th block of the algorithm. Recall that, at the beginning of the ℓ-th block of the algorithm, p_1 contains some active vertices. Then, for the next $t - 1$ steps, we simply embed the children of all currently active vertices in p_i into p_{i+1} arbitrarily (which is possible thanks to (2)). Lastly, in one more step of the algorithm, we distribute the children of all currently active vertices in p_t equally among the four subsquares of $\sigma_k(\phi_\ell(p))$. Finally, if $\ell = k - 1$ we terminate the subroutine, and otherwise we increment the value of ℓ by 1 and proceed to the next block of the first subroutine.

2.2 The second subroutine

We say that two squares of $\mathcal{S}$ are *adjacent* if they share at least one corner vertex. Our next goal is to embed the vertices in the remaining layers. In particular, if $v \in V_{m+2}$ lies in $q \in \mathcal{S}$, then we embed all descendants of v into q or a square adjacent to q. Note that, for every such v, this is possible by our choice of k since the vertices in any square of side length 2^{1-k} form a clique. To show that this can be done simultaneously for all $v \in V_{m+2}$, we use Hall's theorem. The application of Hall's theorem is standard, hence we omit it, but it can be found in the full version of our paper [5].

3 Proof sketch for the feasibility of the embedding algorithm

We show that our embedding algorithm succeeds a.a.s. We make use of the following distribution property of the vertices of G, which follows directly from Chernoff bounds.

Claim 1. *A.a.s., for every square $q \in \mathcal{S}$, the number of vertices of G in q is in $[2^{-2k}n - n^{2/3}, 2^{-2k}n + n^{2/3}]$.*

Following the description of the first subroutine, we must prove that it reaches a desired configuration in a suitable number of steps. A bound on the number of steps is given by the following claim.

Claim 2. *The first subroutine runs for at most $(1 - \varepsilon/4)h$ steps.*

Proof. For every $\ell \in \{1, \ldots, k-1\}$, we have that

$$\|c(p) - c(\phi_\ell(p))\| \leq \|c(q) - c(\phi_\ell(p))\| = \frac{(2^{1-\ell} - 2^{-\ell})}{\sqrt{2}},$$

where the inequality holds since $c(p)$ belongs to the segment $c(q)c(\phi_\ell(p))$ (recall that $\phi_\ell(p)$ is obtained from p by homothety with center $c(q)$ and ratio $2^{k-\ell} > 1$). Hence, by (3), the total number of steps performed by the first subroutine is at most

$$\sum_{\ell=1}^{k-1} \left(\frac{(2^{1-\ell} - 2^{-\ell})/\sqrt{2}}{(1 + 5\varepsilon/8)r^*} + 1 \right) \leq k + \frac{1/\sqrt{2}}{(1 + 5\varepsilon/8)r^*} \leq \left(1 - \frac{\varepsilon}{4}\right)h. \qquad \blacktriangleleft$$

It follows that the total number of vertices embedded during the first subroutine is

$$\sum_{i=0}^{m+2} 2^i \leq 2^{-(h-m-2)}n \leq 2^{-\varepsilon h/5}n = o(2^{-2k}n). \tag{4}$$

Therefore, by Claim 1, there are sufficiently many vertices in each $q \in \mathcal{S}_k$ to guarantee that the choices we made in the first subroutine can indeed be carried out.

Claims 1 and 2 guarantee that a.a.s. the algorithm succeeds. That is, a.a.s. we have an embedding of the layers $V_0, \ldots, V_{m+2}$ of T into G and, moreover, each $q \in \mathcal{S}_k$ contains the same number of vertices of V_{m+2}.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, Wireless sensor networks: a survey. *Comput. Netw.* 38.4 (2002), 393—422, DOI: 10.1016/S1389-1286(01)00302-4.

[2] M. Bradonjić and W. Perkins, On sharp thresholds in random geometric graphs. *Approximation, randomization, and combinatorial optimization. Algorithms and techniques. Proceedings of the 17th international workshop on approximation algorithms for combinatorial optimization problems (APPROX 2014) and the 18th international workshop on randomization and computation (RANDOM 2014), Universitat Politècnica de Catalunya, Barcelona, Spain, September 4–6, 2014*, 500–514, Wadern: Schloss Dagstuhl – Leibniz Zentrum für Informatik, ISBN 978-3-939897-74-3 (2014), DOI: 10.4230/LIPIcs.APPROX-RANDOM.2014.500.

[3] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms.* MIT Press, Cambridge, MA, 4th ed. (2022), ISBN 978-0-262-04630-5.

[4] J. Díaz, D. Mitsche and X. Pérez, Sharp Threshold for Hamiltonicity of Random Geometric Graphs. *SIAM J. Discrete Math.* 21 (2007), 57–65, DOI: 10.1137/060665300.

[5] A. Espuny Díaz, L. Lichev, D. Mitsche and A. Wesolek, Sharp threshold for embedding balanced spanning trees in random geometric graphs. *arXiv e-prints* (2023). arXiv: 2303.14229.

[6] D. Funke, S. Lamm, U. Meyer, M. Penschuck, P. Sanders, C. Schulz, D. Strash and M. von Looz, Communication-free massively distributed graph generation. *J. Parallel Distrib. Comput.* 131 (2019), 200–217, DOI: 10.1016/j.jpdc.2019.03.011.

[7] E. N. Gilbert, Random Plane Networks. *J. Soc. Ind. Appl. Math.* 9 (1961), 533–543, DOI: 10.1137/0109045.

[8] A. Goel, S. Rai and B. Krishnamachari, Monotone properties of random geometric graphs have sharp thresholds. *Ann. Appl. Probab.* 15.4 (2005), 2535–2552, DOI: 10.1214/105051605000000575.

[9] P. Gupta and P. R. Kumar, Critical power for asymptotic connectivity in wireless networks. W. M. McEneaney, G. G. Yin and Q. Zhang (eds.), *Stochastic analysis, control, optimization and applications. A volume in honor of Wendell H. Fleming, on the occasion of his 70th birthday*, 547–566, Birkhäuser, Boston, ISBN 0-8176-4078-9 (1999), DOI: 10.1007/978-1-4612-1784-8_33.

[10] R. Montgomery, Spanning trees in random graphs. *Adv. Math.* 356 (2019), 92, DOI: 10.1016/j.aim.2019.106793. Id/No 106793.

[11] M. Nekovee, Worm epidemics in wireless ad-hoc networks. *New J. Phys.* 9.6 (2007), paper n. 189, DOI: 10.1088/1367-2630/9/6/189.

[12] M. Penrose, The longest edge of the random minimal spanning tree. *Ann. Appl. Probab.* 7.2 (1997), 340–361, DOI: 10.1214/aoap/1034625335.

[13] ———, *Random geometric graphs*, *Oxf. Stud. Probab.*, vol. 5. Oxford: Oxford University Press (2003).

[14] ———, Lectures on random geometric graphs. *Random graphs, geometry and asymptotic structure*, 67–101, Cambridge: Cambridge University Press (2016), DOI: 10.1017/CBO9781316479988.004.