On the Giant Component of Geometric Inhomogeneous Random Graphs

Thomas Bläsius
Karlsruhe Institute of Technology, Germany

Tobias Friedrich
Hasso Plattner Institute, University of Potsdam, Germany

Maximilian Katzmann
Karlsruhe Institute of Technology, Germany

Janosch Ruff
Hasso Plattner Institute, University of Potsdam, Germany

Ziena Zeif
Hasso Plattner Institute, University of Potsdam, Germany

Abstract

In this paper we study the threshold model of *geometric inhomogeneous random graphs* (GIRGs); a generative random graph model that is closely related to *hyperbolic random graphs* (HRGs). These models have been observed to capture complex real-world networks well with respect to the structural and algorithmic properties. Following comprehensive studies regarding their *connectivity*, i.e., which parts of the graphs are connected, we have a good understanding under which circumstances a *giant* component (containing a constant fraction of the graph) emerges.

While previous results are rather technical and challenging to work with, the goal of this paper is to provide more accessible proofs. At the same time we significantly improve the previously known probabilistic guarantees, showing that GIRGs contain a giant component with probability $1 - \exp(-\Omega(n^{(3-\tau)/2}))$ for graph size n and a degree distribution with power-law exponent $\tau \in (2, 3)$. Based on that we additionally derive insights about the connectivity of certain induced subgraphs of GIRGs.

2012 ACM Subject Classification Theory of computation → Random network models

Keywords and phrases geometric inhomogeneous random graphs, connectivity, giant component

Digital Object Identifier 10.4230/LIPIcs.ESA.2023.20

1 Introduction

Geometric inhomogeneous random graphs (GIRGs) are a generative graph model where vertices are weighted and placed in a geometric ground space and the probability for two of them to be adjacent depends on the product of their weights, as well as their distance [19]. In a sense the model combines the strengths of *inhomogeneous random graphs* [28] and *random geometric graphs* [26]. Introduced as a simplified and more general version of *hyperbolic random graphs* (HRGs) [23], GIRGs share crucial properties with complex real-world networks. Such networks are typically characterized by a *heterogeneous degree distribution* (with few high-degree vertices, while the majority of vertices has small degree), *high clustering* (vertices with common neighbors are likely adjacent themselves), and a *small diameter* (longest shortest path), and it has been shown that GIRGs and HRGs capture these properties well [17, 19, 25].

Beyond these structural properties, GIRGs have also been observed to be a good model for real-wold networks when it comes to the performance of graph algorithms [3]. This makes the GIRG framework relevant for algorithmic purposes in multiple ways. On the one hand, they are a useful tool in the context of average-case analysis, where they yield more

realistic instances than, e.g., the Erdős–Rényi model, while it is still sufficiently simple to be mathematically accessible [3, 7]. On the other hand, we can use GIRGs to generate an abundance of benchmark instances with varying properties, allowing us to perform thorough evaluations of algorithms even when real-world data is scarce [4, 5].

One of the most basic graph properties, which is also relevant from an algorithmic point of view, is *connectivity*, i.e., the question about what parts of a graph are connected via paths. For random graphs, the first question that typically arises in the context of connectivity revolves around the emergence of a so-called *giant component*, which is a connected component whose size is linear in the size of the graph. The existence of a giant has been researched on many related graph models like *Erdős–Rényi random graphs* [12, 13], *random geometric graphs* [2, 11, 26, 18], as well as on *Chung-Lu random graphs* that also capture inhomogeneous random graphs [1, 9, 10].

Unsurprisingly, being such a fundamental feature, connectivity has also been studied on GIRGs, and since HRGs are so closely related to them, we consider the corresponding results to be relevant here as well. For HRGs we know how the emergence of a giant depends on certain model properties that control the degrees of the resulting graph [6, 14]. We note that some analyses there are based on a coupling from HRGs to a continuum percolation model that exhibits a strong resemblance to GIRGs (see [14, Section 2] and [19, Part I, Section 3.5]). Beyond the giant we also have bounds on the size of the second largest component of HRGs [20]. For GIRGs it is known that a giant exists *asymptotically almost surely*, i.e., with probability $1 - o(1)$ [21], with another proof giving a certainty of $1 - n^{-\omega(1)}$ where n denotes the number of vertices in the graph [19, Theorem 4.2]. We note that the specific function in the exponent has not been determined before.

In this paper, we answer this question, by showing that threshold GIRGs have a giant component with probability at least $1 - \exp(-\Omega(n^{(3-\tau)/2}))$. This improves the previous results in two ways. First, our proof is simpler and shorter than the technical existing proofs for HRGs [6, 14]. Secondly, our probability bound is substantially stronger compared to previous bounds obtained for GIRGs. Moreover, we note that our improved bound does not only hold for the full graph but also translates to subgraphs located in restricted regions of the ground space. The argument for this is inspired by a technique used for HRGs [14, Section 4] (though it is much simpler in our case).

Besides providing more accessible insights in the connectivity of GIRGs, we believe that our results, in particular those on subgraphs in restricted regions, can be helpful for algorithmic applications. For example in problems like *balanced connected partitioning* [8], one is interested in partitioning a graph into connected components of (roughly) equal size and in *component order connectivity* [16] the goal is to find a small separator that divides the graph into components of bounded size. There it is important, that the graph cannot only be separated into smaller pieces but that these pieces remain actually connected.

In the following, we give a brief overview of the basic concepts used in the paper (Section 2) before presenting our proofs regarding the emergence of a giant in GIRGs (Section 3).

2 Preliminaries

Geometric Inhomogeneous Random Graphs. Let $\mathbb{B}^d = [0,1]^d$ be the d-dimensional hypercube ($\mathbb{B}$ for "box") and let dist be the L_∞ metric, i.e., for $x = (x_1, \dots, x_d) \in \mathbb{B}^d$ and $y = (y_1, \dots, y_d) \in \mathbb{B}^d$ we have $\text{dist}(x, y) = \max_{i \in [d]} |x_i - y_j|$.

A *geometric inhomogeneous random graph (GIRG)* $G = (V, E)$ with *ground space* $\mathbb{B}^d$ is obtained in three steps. The first step consists of a homogeneous Poisson point process on $\mathbb{B}^d$, with an intensity that yields n points in expectation. Each point is then considered to

be a vertex in the graph. In the second step, each vertex v is assigned a *weight* $w_v > 1$ that is sampled according to a Pareto distribution with exponent $\tau \in (2,3)$, i.e., $\Pr[w_v \leq w] = 1 - w^{-(\tau-1)}$. In the third step, any two vertices u and v are connected by an edge with a probability that depends on their distance and their weights. More precisely, there are two variants. In a *threshold GIRG*, u and v are adjacent if and only if

$$\operatorname{dist}(u,v) \leq \left(\frac{\lambda w_u w_v}{n}\right)^{1/d},$$

where the constant $\lambda > 0$ controls the expected average degree of the graph. We note that the relation between λ and the corresponding average degree is not trivial and refer to [5, Section 4.3] for details. In the *temperate* variant we have an additional temperature parameter $T \in (0,1)$ and the probability for u and v to be adjacent is given by

$$\Pr[\{u,v\} \in E] = \min\left\{1, \left(\frac{\lambda w_u w_v}{n \cdot (\operatorname{dist}(u,v))^d}\right)^{1/T}\right\}.$$

The threshold variant is the limit of the temperature variant for $T \to 0$. We denote the resulting probability distribution of graphs with $\mathcal{G}(n, \mathbb{B}^d, \tau, \lambda, T)$ for general GIRGs (allowing temperatures in $T \in [0,1)$). When we just refer to the threshold case, we use $\mathcal{G}(n, \mathbb{B}^d, \tau, \lambda)$. We assume the parameters d, τ, λ, and T to be constant, i.e., independent of n.

GIRG Variants. In the literature, several variants of the GIRG model have been studied and we want to briefly discuss the choice we made here. Usually, GIRGs are considered with a torus $\mathbb{T}^d$ as ground space, i.e., the distance in the ith dimension, between x and y is $\min\{|x_i - y_i|, 1 - |x_i - y_i|\}$ instead of just $|x_i - y_i|$. The torus usually makes arguments easier as it eliminates the special case close to the boundary of $\mathbb{B}^d$. However, in our case, this is not relevant. Moreover, as distances in $\mathbb{T}^d$ are only smaller than in $\mathbb{B}^d$, all our results concerning the largest connected component directly translate to the case where $\mathbb{T}^d$ is the ground space.

Moreover, instead of sampling n points uniformly at random in the ground space, we use a Poisson point process. This is a technique often used in geometric random graphs as it makes the number of vertices appearing in disjoint regions stochastically independent. This is a similar difference as the one between the Erdős–Rényi model $G(n,m)$ with a fixed number of edges m and the Gilbert model $G(n,p)$ with a fixed probability p for each individual edge to exist. While we generally advocate for using the Poisson variant of the GIRG model, we note that our result carries over to the uniform model.

Poisson Point Process. Let $R \subseteq \mathbb{B}^d$ be a region of the ground space with volume a. Then, the size of the vertex set $V(R)$, i.e., the number of vertices that are sampled in R is a random variable following a Poisson distribution with expectation $\mu = an$. This in particular means that the probability for R to contain no vertex is $\exp(-\mu)$.

We note that the Poisson point process we consider is a *marked* process, where each point sampled from $\mathbb{B}^d$ obtains a weight sampled from a weight space $\mathcal{W}$ as a mark. Due to the marking theorem, this is equivalent to considering an (inhomogeneous) Poisson point process of the product space $\mathbb{B}^d \times \mathcal{W}$, i.e., colloquially speaking, each point pops up with a position and a weight instead of initially only having a position and drawing the weight as an afterthought. This is also equivalent to just sampling the number of points N following a Poisson distribution and viewing the positions and the weights as marks that are sampled subsequently for each of the N points. Throughout the paper, we switch between these different perspectives without making this explicit.

Lowest Weights Dominate. We regularly consider weight ranges $[w_1, w_2]$ with $w_2 \geq c \cdot w_1$ for a constant $c > 1$. The probability for a v to have weight in $[w_1, w_2]$ is dominated by w_1:

$$\Pr[w_v \in [w_1, w_2]] = w_1^{-(\tau-1)} - w_2^{-(\tau-1)} \geq w_1^{-(\tau-1)} \cdot \left(1 - c^{-(\tau-1)}\right) \in \Theta(\Pr[w_v \geq w_1]).$$

3 Existence of a Giant Component

We want to show that a threshold GIRG is highly likely to contain a connected component of linear size. Our argument goes roughly as follows. We first note that vertices with weight at least $\sqrt{n/\lambda}$ form a clique, which we call the *core* of the graph. For each non-core vertex, we can show that the probability that it has a path into the core is non-vanishing, i.e., it is lower bounded by a non-zero constant. This already shows that we get a connected graph of linear size in expectation.

To show concentration, i.e., that we get a large connected component with the claimed probability, we essentially need to show that the events for different low-weight vertices to connect to the core are sufficiently independent of each other. To this end, we subdivide the ground space into a grid of regular *cells* of side length Δ. We call a cell *nice* if a linear number of its vertices connect to the core via paths not leaving the cell and then show that a cell is nice with non-vanishing probability. As this only considers paths within the cell, the different cells are independent. Thus, we get a series of independent coin flips, one for each cell. If a constant fraction of these coin flips succeeds, we have a connected component of linear size. Hence, if the number of cells is sufficiently large, we get concentration via a Chernoff bound. It follows that we essentially want to choose the cell width Δ to be as small as possible such that cells are still nice with non-vanishing probability.

In Section 3.1, we first show that every vertex has constant probability to have a path to the core. In fact, we show something slightly stronger, by considering not just any paths but so-called layer paths. Afterwards, we use this result in Section 3.2, to bound the probability for a cell to be nice. This then also informs us on how to choose the cell width Δ and thus on how many cells we obtain. With this, we can wrap up the argument in Section 3.3 by applying a Chernoff bound. Besides our main results, we there also mention immediate implications.

3.1 Layer Paths

We want to show that, for any individual vertex, the probability that it has a path to a vertex of the core is non-vanishing. For this, we define the ℓ-th *layer* V_ℓ to be the set of vertices with weight in $[e^{\ell/2}, e^{(\ell+1)/2})$. Note that the upper and lower bounds are a constant factor apart and thus (as mentioned in Section 2) the probability for a vertex to have layer ℓ is asymptotically dominated by the lower bound, i.e., $\Pr[v \in V_\ell] \in \Theta(\Pr[w_v \geq e^{\ell/2}]) = \Theta(e^{-\ell(\tau-1)/2})$.

A path $(v_0, \dots, v_k)$ is a *layer path* if it goes from one layer to the next in every step, i.e., $v_i \in V_\ell$ implies $v_{i-1} \in V_{\ell-1}$ for every $i \in [k]$. Note that vertices in layer $\lceil \log(n/\lambda) \rceil$ have weight at least $\sqrt{n/\lambda}$ and thus belong to the core. Thus, the following lemma shows that each vertex has a layer path to the core with non-vanishing probability.

▶ **Lemma 1.** *Let $G \sim \mathcal{G}(n, \mathbb{B}^d, \tau, \lambda)$ be a threshold GIRG and let v be a non-core vertex. The probability that there is a layer path from v to layer $\lceil \log(n/\lambda) \rceil$ is non-vanishing.*

Proof. We bound the probability that such a layer path exists in three steps. First, we bound the probability that a vertex u on layer ℓ has a neighbor in layer $\ell + 1$. In the second step, we consider the intersection of the events where this happens on all considered layers. Finally, we show that the resulting probability is non-vanishing.

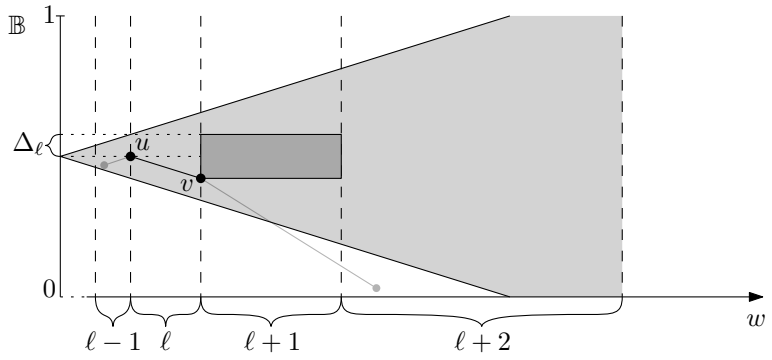

Figure 1 Excerpt of a one-dimensional GIRG with the weights on the x-axis and the ground space $\mathbb{B}$ on the y-axis. A layer path spans from layer $\ell-1$ to $\ell+2$. The gray region is the neighborhood of vertex u. The dark-gray region contains all vertices in layer $\ell+1$ that have distance at most Δ_ℓ to u.

For the first step, consider two vertices $u \in V_\ell$ and $v \in V_{\ell+1}$ in consecutive layers, as shown in Figure 1. Both their weights are at least $w = e^{\ell/2}$. Thus, they are definitely adjacent if their distance $\text{dist}(u,v)$ satisfies

$$\text{dist}(u,v) \leq \left(\frac{\lambda w^2}{n}\right)^{1/d} = \lambda^{1/d}\left(\frac{e^\ell}{n}\right)^{1/d} =: \Delta_\ell.$$

If vertex $u \in V_\ell$ is the current vertex from which we want to make the next step in a layer path, we are thus interested in the probability that there is a vertex v that lies in layer $\ell+1$ with $\text{dist}(u,v) \leq \Delta_\ell$. Since these two events (being in layer $\ell+1$ and having sufficiently low distance) are independent, the probability that both happen is $\Pr[v \in V_{\ell+1}] \cdot \Pr[\text{dist}(u,v) \leq \Delta_\ell]$. As mentioned above, we have $\Pr[v \in V_{\ell+1}] \in \Theta(e^{-\ell(\tau-1)/2})$. Moreover, $\Pr[\text{dist}(u,v) \leq \Delta_\ell] \in \Theta(\Delta_\ell^d) = \Theta(e^\ell/n)$. Hence, we obtain

$$\Pr[v \in V_{\ell+1}] \cdot \Pr[\text{dist}(u,v) \leq \Delta_\ell] \in \Theta\left(e^{-\ell(\tau-1)/2} \cdot e^\ell/n\right) = \Theta\left(e^{\ell(3-\tau)/2}/n\right).$$

To conclude the first step of the proof, let X_ℓ be the number of vertices in layer $\ell+1$ with distance at most Δ_ℓ to $u \in V_\ell$. By the above probability, we have $\mathbb{E}[X_\ell] = \Theta(e^{\ell(3-\tau)/2})$. We consider the event $X_\ell > 0$ and call it A_ℓ. Note that A_ℓ implies that u has at least one neighbor in the next layer. As X_ℓ follows a Poisson distribution, we get

$$\Pr[A_\ell] = 1 - \Pr[X_\ell = 0] = 1 - \exp(-\mathbb{E}[X_\ell]) = 1 - \exp\left(-\Theta\left(e^{\ell(3-\tau)/2}\right)\right).$$

In the second step of the proof, we now consider the intersection of all the independent events $A_0, A_1, \dots, A_{\lceil \log(n/\lambda)\rceil}$, which is sufficient for a layer path starting in layer 0 to exist. Note that a lower bound for the probability of this intersection also gives a lower bound for the existence of a layer path starting in any other layer. To show that this intersection occurs with non-vanishing probability, we utilize the fact that $\Pr[A_\ell]$ approaches 1 very quickly for increasing ℓ. More precisely, we show that for a constant c, all subsequent events A_ℓ with $\ell \geq c$ are sufficiently likely, that we can simply take the union bound over their complements. Thus, we obtain

$$\Pr\left[\bigcap_{\ell=0}^{\lceil \log(n/\lambda)\rceil} A_\ell\right] = \Pr\left[\bigcap_{\ell=0}^{c-1} A_\ell\right] \cdot \Pr\left[\bigcap_{\ell=c}^{\lceil \log(n/\lambda)\rceil} A_\ell\right].$$

Clearly, the first factor is non-vanishing as it is the product of constantly many non-zero constants. For the second factor, we consider the complementary events and apply the union bound to obtain

$$\begin{aligned}\Pr\left[\bigcap_{\ell=c}^{\lceil\log(n/\lambda)\rceil} A_\ell\right] &= 1 - \Pr\left[\bigcup_{\ell=c}^{\lceil\log(n/\lambda)\rceil} A_\ell^C\right] \\ &\geq 1 - \sum_{\ell=c}^{\lceil\log(n/\lambda)\rceil} (1 - \Pr[A_\ell]) \\ &= 1 - \sum_{\ell=c}^{\lceil\log(n/\lambda)\rceil} \exp\left(-\Theta(e^{\ell(3-\tau)/2})\right).\end{aligned}$$

Since the sum converges, we can choose c to be a sufficiently large constant such that the sum is bounded by any constant $\varepsilon > 0$. The above expression is thus at least $1 - \varepsilon$, which is non-vanishing. ◀

Observe that Lemma 1 already shows that the expected number of vertices with a layer path to the core is linear. Thus, the expected size of the connected component including the core vertices is linear. To show concentration, we separate the ground space into cells that are then considered independently.

3.2 A Coin Flip for Each Cell

We subdivide the ground space into a grid of regular cells of side length Δ. We first show that the high-weight vertices of each cell are likely to induce a connected graph. This is useful as we can afterwards focus on vertices of lower weight. As edges between low-weight vertices are short, layer paths on these vertices can cover only a small distance and thus only few of them leave their cell, which makes different cells (mostly) independent.

▶ **Lemma 2.** *Let $G \sim \mathcal{G}(n, \mathbb{B}^d, \tau, \lambda)$ be threshold GIRG, let C be a cell of side length Δ, and let w be a weight. Then, the graph induced by vertices in C of weight at least w is connected with probability at least*

$$1 - \frac{(2\Delta)^d}{\lambda w^2} \cdot \exp\left(-\frac{\lambda w^{3-\tau}}{2^d}\right) \cdot n.$$

Proof. We discretize the cell C into sub-cells, such that vertices in adjacent sub-cells are adjacent themselves, as shown in Figure 2. Note that two vertices u, v with weights $w_u, w_v \geq w$ are adjacent if their distance is bounded by

$$\operatorname{dist}(u, v) \leq \left(\frac{\lambda w^2}{n}\right)^{1/d}.$$

Thus, all vertices in adjacent sub-cells are guaranteed to be adjacent, if the side length of a sub-cell is

$$\Delta_w = \frac{1}{2}\left(\frac{\lambda w^2}{n}\right)^{1/d}.$$

Note that for very large w, we get $\Delta_w \geq \Delta$, in which case all vertices in C are pairwise adjacent with probability 1. In the following, we therefore assume that w is smaller. For a given sub-cell C', we compute the probability for a given vertex v to lie in C' as

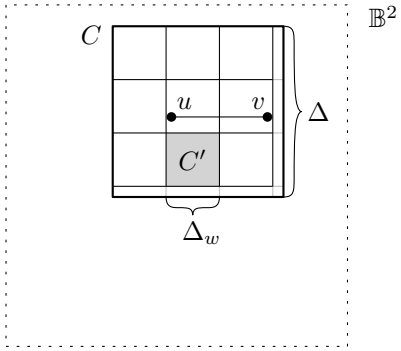

Figure 2 The cell C of width Δ is divided into sub-cells of width Δ_w. The sub-cell C' is completely contained in C. The vertices u and v are in adjacent sub-cells and are therefore adjacent themselves.

$$\Pr[v \in V(C')] = (\Delta_w)^d = \left(\frac{1}{2}\left(\frac{\lambda w^2}{n}\right)^{1/d}\right)^d = \frac{\lambda w^2}{2^d n}.$$

Additionally, the probability for v to have weight at least $w_v \geq w$, is given by

$$\Pr[w_v \geq w] = 1 - \Pr[w_v \leq w] = w^{-(\tau-1)}.$$

Together, we obtain

$$\Pr[v \in V(C') \wedge w_v \geq w] = \Pr[v \in V(C')] \cdot \Pr[w_v \geq w] = \frac{\lambda w^2}{2^d n} \cdot w^{-(\tau-1)} = \frac{\lambda w^{3-\tau}}{2^d n}.$$

Consequently, the expected number of vertices of weight at least w in C' is

$$\mathbb{E}\left[|\{v \in V(C') \mid w_v \geq w\}|\right] = \frac{\lambda w^{3-\tau}}{2^d}.$$

Since the vertices are distributed according to a Poisson distribution, the probability for C' to not contain any of these vertices is given by

$$\Pr[\{v \in V(C') \mid w_v \geq w\} = \emptyset] = \exp\left(-\frac{\lambda w^{3-\tau}}{2^d}\right).$$

Finally, we lower-bound the probability for the vertices of weight at least w in our initial cell C to induce a connected graph, by considering the probability that none of its sub-cells is empty. Note that we have

$$k = \left(\left\lfloor \frac{\Delta}{\Delta_w} \right\rfloor\right)^d$$

sub-cells $C'_1, \dots, C'_k$ that are *completely* contained in the cell C. Clearly, whether the remaining sub-cells (intersecting the boundary of C) are empty or not has no impact on the connectedness of the considered subgraph. The probability for all of the sub-cells $C'_1, \dots, C'_k$ to be non-empty can be simplified by applying union bound, which yields

$$\begin{aligned}\Pr[\forall C' \in \{C'_1, \dots, C'_k\} \colon V(C') \neq \emptyset] &= (1 - \Pr[V(C') = \emptyset])^k \\ &\geq 1 - k \cdot \Pr[V(C') = \emptyset] \\ &= 1 - \left(\left\lfloor \frac{\Delta}{\Delta_w} \right\rfloor\right)^d \cdot \exp\left(-\frac{\lambda w^{3-\tau}}{2^d}\right) \\ &\geq 1 - \Delta^d \cdot \frac{2^d n}{\lambda w^2} \cdot \exp\left(-\frac{\lambda w^{3-\tau}}{2^d}\right) \\ &= 1 - \frac{(2\Delta)^d}{\lambda w^2} \cdot \exp\left(-\frac{\lambda w^{3-\tau}}{2^d}\right) \cdot n. \end{aligned}$$

◀

The following lemma shows that we basically get an independent coin-flip with non-vanishing success probability for each cell to be nice. We want to point out three technical details of the lemma statement here. First, the lemma specifically considers the connected component containing a vertex of weight at least $\hat{w}$. We will later choose $\hat{w} = \sqrt{n/\lambda}$, i.e., this vertex is part of the core. As all core vertices form a clique, this makes sure that the components we get for the individual cells actually connect to one large component in the whole graph. Secondly, the lower bound on μ, which is the expected number of vertices in the cell, given by $\mu = \Delta^d n$, requires that the cells are sufficiently large to contain a vertex of weight $\hat{w}$ with non-vanishing probability. Thirdly, the lower bound on $\hat{w}$ ensures that vertices with higher weight are likely connected by Lemma 2.

▶ **Lemma 3.** *Let $G \sim \mathcal{G}(n, \mathbb{B}^d, \tau, \lambda)$ be a threshold GIRG, let $\hat{w}$ be a weight, and let C be a cell of side length Δ containing μ vertices in expectation. Then, with non-vanishing probability, the graph induced by the vertices in C contains a vertex of weight at least $\hat{w}$ whose connected component has size $\Theta(\mu)$, if $\mu \geq \hat{w}^{\tau-1}$, $\mu \in \omega((\log n)^{2/(3-\tau)} \log\log(n)^d)$, and $\hat{w} \in \omega((\log n)^{1/(3-\tau)})$.*

Proof. The overall argument goes as follows. First, the lower bound on μ ensures that C contains a vertex of weight $\hat{w}$ with non-vanishing probability. For a smaller weight $\overline{w} \leq \hat{w}$, we then apply Lemma 2 to get that all vertices of weight at least $\overline{w}$ form a connected component asymptotically almost surely. Afterwards, it remains to show that enough vertices of lower weight connect to a vertex of weight at least $\overline{w}$ via paths not leaving C. For the existence of these paths, we use Lemma 1. To show that most of them do not leave C, we use that the considered vertices have weight at most $\overline{w}$ and thus cannot deviate too much from the starting position.

Recall that the weight of a vertex is at least $\hat{w}$ with probability $\hat{w}^{-(\tau-1)}$. Thus, the expected number of vertices in cell C with weight at least $\hat{w}$ is $\mu\hat{w}^{-(\tau-1)}$. Plugging in the bound $\mu \geq \hat{w}^{\tau-1}$, everything cancels and we obtain an expected value of 1. As the number of vertices in C with weight above $\hat{w}$ follows a Poisson distribution, we get at least one such vertex with non-vanishing probability.

We set $\overline{w} = ((2^d/\lambda)\log n)^{1/(3-\tau)}$. Note that by the condition on $\hat{w}$ in the lemma statement, we have $\overline{w} \leq \hat{w}$ for sufficiently large n. Note further that $\overline{w}$ is chosen such that the exponent in the bound of Lemma 2 simplifies to $-\log n$. Thus by Lemma 2, the graph induced by the vertices of weight at least $\overline{w}$ in C is connected with probability at least $1 - (2\Delta)^d/(\lambda\overline{w}^2)$. As $\Delta \leq 1$ and $\overline{w}$ is increasing with n, this goes to 1 for $n \to \infty$.

Consider a vertex of weight below $\overline{w}$. Then, by Lemma 1, it has a layer path to a vertex with weight at least $\overline{w}$ with non-vanishing probability. In the following, with *layer path* we always refer to a layer path that ends in the layer belonging to $\overline{w}$. Note that a layer path

has length at most $O(\log\log n)$. Also note that the largest weight we encounter is in $O(\overline{w})$ as the path stops in the layer belonging to weight $\overline{w}$ and the weights increase only by a constant factor between layers. It follows that, in each dimension, the distance between two consecutive vertices on a layer path is in $O((\overline{w}^2/n)^{1/d})$, as the vertices would not be connected otherwise. Thus, the overall deviation of a layer path from the starting point is upper bonded by $O((\overline{w}^2/n)^{1/d}\log\log n) = O\left(((\log n)^{2/(3-\tau)}\log\log(n)^d/n)^{1/d}\right)$. By the second lower bound on μ, this is asymptotically less than Δ. Thus, shrinking C accordingly from all directions yields a subregion C' that contains $\Theta(\mu)$ vertices in expectation such that any layer path that starts in C' stays in C.

Instead of counting all vertices in C' that have layer paths, we only count vertices in the first layer. This has the advantage, that the event that an individual vertex in the first layer has a layer path is independent of the number of vertices in the first layer (while it depends on the number of vertices in higher layers). First note that the number of vertices in the first level of C' is a random variable following a Poisson distribution with expected value in $\Theta(\mu)$. Thus, there are $\Theta(\mu)$ such vertices with non-vanishing probability.

Now let $X \in [0,1]$ be the random variable that describes the fraction of vertices in the first layer that fail to have a layer path. By Lemma 1, the probability for an individual vertex to not have a layer path is a upper bounded constant $p < 1$ (i.e., the layer path exists with non-vanishing probability at least $1-p$). Thus, we get $\mathbb{E}[X] \leq p$. Markov's inequality then gives us $\Pr[X \geq c] \leq p/c$ and thus $\Pr[X < c] \geq 1 - p/c$. We can choose c to be a constant with $p < c < 1$, which gives us a non-vanishing probability that a fraction of at least $1 - c > 0$ vertices have the desired layer path. Note that this holds independently of the number of vertices actually sampled in the first layer of C'.

To wrap up, consider the three events that there exists a vertex of weight at least $\hat{w}$, that there are $\Theta(\mu)$ vertices in the first layer of C', and that a constant fraction of them have layer paths. Note that the three events are independent and each holds with non-vanishing probability. Thus, their intersection, which we denote with A, also holds with non-vanishing probability. Finally, the event B that all vertices of weight at least $\overline{w}$ induce a connected graph holds asymptotically almost surely. Though A and B are not independent, we can apply the union bound to their complements to obtain that A and B together hold with non-vanishing probability. ◀

3.3 Large Components are Likely to Exist

To obtain the following theorem, it remains to apply a Chernoff bound to the coin flips obtained for each cell by Lemma 3.

▶ **Theorem 4.** *Let $G \sim \mathcal{G}(n, \mathbb{B}^d, \tau, \lambda)$ be a threshold GIRG. Then G has a connected component of size $\Theta(n)$ with probability $1 - \exp\left(-\Omega(n^{(3-\tau)/2})\right)$.*

Proof. First note that the probability to have $\omega(n)$ vertices is exponentially small and thus we only have to show the lower bound on the size of the largest connected component. To apply Lemma 3, we choose the cell width Δ such that $\Delta^d n = \mu = \hat{w}^{\tau-1}$ where we set $\hat{w} = \sqrt{n/\lambda}$. With this, we obtain that the number of cells k is

$$k \in \Theta\left(\frac{1}{\Delta^d}\right) = \Theta\left(\frac{n}{\hat{w}^{\tau-1}}\right) = \Theta\left(\frac{n}{\left(\sqrt{n/\lambda}\right)^{\tau-1}}\right) = \Theta\left(n^{\frac{3-\tau}{2}}\right).$$

Note that this bound is valid even if Δ does not divide the ground space evenly. Further note that the chosen Δ satisfies the conditions of Lemma 3, the graph induced by each cell contains a vertex from the core whose connected component has size $\Theta(\mu)$ with non-vanishing probability. If this holds for a constant fraction of cells, we get a giant component, as all vertices of weight at least $\hat{w}$ form a clique in G. Thus, we have k independent coin flips, each succeeding with a probability of $p > 0$, and we are interested in the number of successes X. To show that $X \in \Theta(k)$ is highly likely, we can simply apply a Chernoff bound (see [24, Theorem 4.4]). For $\delta \geq 0$, we get

$$\Pr\left[X \leq (1-\delta)\mathbb{E}[X]\right] \leq \exp\left(-\frac{\delta^2}{2}\mathbb{E}[X]\right).$$

As $\mathbb{E}[X] \in \Theta(k)$, this implies $\Pr\left[X \in o(k)\right] \leq \exp\left(-\Omega(k)\right)$. Inserting k yields the claim. ◀

As already mentioned in Section 2, this directly implies the following corollary.

▶ **Corollary 5.** *Theorem 4 also holds with the torus $\mathbb{T}$ as ground space.*

The following lemma states a well known property of GIRGs, see e.g. [19, Lemma 3.12] and [22, Definition 2.8]. For the sake of transparency, we give a simple proof based on the notation established throughout the paper.

▶ **Lemma 6** (folklore)**.** *Let $H \sim \mathcal{G}(n, \mathbb{B}^d, \tau, \lambda, T)$ be a GIRG and let G be the subgraph induced by the vertices within a cell of side length $\Delta = (f(n)/n)^{1/d}$. Then $G \sim \mathcal{G}(f(n), \mathbb{B}^d, \tau, \lambda, T)$.*

Proof. Note that we basically consider two ways to generate a graph and claim that they give the same probability distribution over graphs. Intuitively, this can be seen by generating points with weights in the cell $[0, \Delta]^d$, scaling it to the full ground space $[0, 1]^d$, and making three observations. First, for the vertex positions, this is equivalent to directly sampling points in $[0, 1]^d$. Secondly, the weight distribution is independent of the number of vertices. Thirdly, the connection probabilities between vertices are the same in the scaled variant as they are in the cell. To make this more formal, draw G as a subgraph of H as stated in the lemma and draw $G' \sim \mathcal{G}(f(n), \mathbb{B}^d, \tau, \lambda, T)$. We show that G and G' follow the same distribution.

Recall that we consider the Poisson variant of the GIRG model, i.e., the vertices are the result of a Poisson point process in the product space $\mathbb{B}^d \times \mathcal{W}$. Thus, the vertex set of G can be generated by first determining the number of points n_G with positions in $[0, \Delta]^d$, which is a random variable following a Poisson distribution with expectation $n \cdot \Delta^d = f(n)$. Then, independently for each of the n_G vertices, a position is drawn uniformly at random from $[0, \Delta]^d$ and a weight is drawn from $(1, \infty)$ with probability density function $(\tau - 1) \cdot w^{-\tau}$.

To generate G', we can also first determine the number of points $n_{G'}$, which is also Poisson distributed with expectation $f(n)$. Thus, we can couple n_G and $n_{G'}$ to have the same value and we assume a one-to-one correspondence between the vertices in G and G' in the following. For each vertex, the weight is again a random variable with density $(\tau - 1) \cdot w^{-\tau}$, which only depends on τ. Thus, for each vertex, we can couple its weight in G with its weight in G' to assume them to be equal. The position in G' is drawn uniformly from $[0, 1]^d$. Thus, we can couple the random variables for the positions in G' with those in G such that a vertex with position $x \in [0, \Delta]^d$ in G has position x/Δ in G'. Note that this has the effect that all distances between vertices in G' are scaled by a factor of $1/\Delta$ compared to the corresponding distance in G.

It remains to show that for every vertex pair u, v the connection probability in G is the same as in G'. Let w_u and w_v be the weight of u and v (which is the same for G and G' due to the coupling). Also, let $\text{dist}(u, v)$ be the distance between u and v in G and let $\text{dist}'(u, v) = \text{dist}(u, v)/\Delta$ be their distance in G'. Then (for $T > 0$) the connection probability of u and v in G is

$$\Pr[\{u, v\} \in E] = \min\left\{\left(\frac{\lambda w_u w_v}{n \,\text{dist}(u, v)^d}\right)^{1/T}, 1\right\}.$$

The two things that change for G' is that n is replaced by $f(n)$ and $\text{dist}(u, v)^d$ is replaced by $\text{dist}'(u, v)^d = (\text{dist}(u, v)/\Delta)^d = n/f(n) \cdot \text{dist}(u, v)^d$. The $f(n)$ cancels out, yielding the same connection probability for G and G'. For $T = 0$, the argument works analogously. ◀

Together with Theorem 4 this yields the following corollary. We note that this also yields large connected components within cells that are too small to contain a core vertex. For such cells, we know that we get a large connected component but we do not know whether it connects to the giant of the whole graph. Clearly, the same statement holds with the torus $\mathbb{T}^d$ as ground space.

▶ **Corollary 7.** *Let $H \sim \mathcal{G}(n, \mathbb{B}^d, \tau, \lambda)$ be a threshold GIRG and let G be the subgraph induced by the vertices within a cell of side length $\Delta = (f(n)/n)^{1/d}$. Then G has a connected component of size $\Theta(f(n))$ with probability $1 - \exp\left(-\Omega(f(n)^{(3-\tau)/2})\right)$.*

4 Conclusion

Our proof for the emergence of a giant component in geometric inhomogeneous random graphs builds on three simple arguments. First, GIRGs are likely to contain a clique of high-weight vertices. Second, the remaining vertices are sufficiently likely to connect to this core via layer-paths, whose vertices have exponentially increasing weight. And, third, most of these paths exist sufficiently independently from each other.

We note that the same argumentation also works for the closely related hyperbolic random graph model, where the discretization into weight layers translates to a natural discretization of the underlying geometric space that was previously used to bound the diameter of HRGs [15].

Our resulting strong probability bound can be combined with a simple coupling argument to identify connected subgraphs of arbitrary size in certain subregions of the geometric ground space. In particular, when these subregions are the cells of a regular grid (as used several times throughout the paper), we obtain connected subgraphs of roughly equal size. We believe that this property can be utilized in the context of problems with connectivity constraints. For example, in the previously mentioned *balanced connected partitioning* problem [8, 27], the goal is to partition the vertices of a graph into a given number of sets of approximately equal size, such that their induced subgraphs are connected. Moreover, in *component order connectivity* [16] the aim is to find a minimum number of vertices such that after their removal each connected component has bounded size. We conjecture that our structural insights in Corollary 7 may prove useful in obtaining efficient algorithms for these problems on GIRGs and the networks they represent well.

References

1 William Aiello, Fan Chung, and Linyuan Lu. A random graph model for power law graphs. *Experimental Mathematics*, 10(1):53–66, 2001. doi:10.1080/10586458.2001.10504428.

2 Martin J.B. Appel and Ralph P. Russo. The connectivity of a graph on uniform points on $[0,1]^d$. *Statistics & Probability Letters*, 60(4):351–357, 2002. doi:10.1016/S0167-7152(02)00233-X.

3 Thomas Bläsius and Philipp Fischbeck. On the External Validity of Average-Case Analyses of Graph Algorithms. In *European Symposium on Algorithms (ESA 2022)*, pages 21:1–21:14, 2022. doi:10.4230/LIPIcs.ESA.2022.21.

4 Thomas Bläsius, Tobias Friedrich, and Christopher Weyand. Efficiently Computing Maximum Flows in Scale-Free Networks. In *29th Annual European Symposium on Algorithms (ESA 2021)*, volume 204, pages 21:1–21:14, 2021. doi:10.4230/LIPIcs.ESA.2021.21.

5 Thomas Bläsius, Tobias Friedrich, Maximilian Katzmann, Ulrich Meyer, Manuel Penschuck, and Christopher Weyand. Efficiently generating geometric inhomogeneous and hyperbolic random graphs. *Network Science*, 10(4):361–380, 2022. doi:10.1017/nws.2022.32.

6 Michel Bode, N. Fountoulakis, and Tobias Müller. On the largest component of a hyperbolic model of complex networks. *Electronic Journal of Combinatorics*, 22:1–46, August 2015. doi:10.1214/17-AAP1314.

7 Karl Bringmann, Ralph Keusch, Johannes Lengler, Yannic Maus, and Anisur R. Molla. Greedy routing and the algorithmic small-world phenomenon. *Journal of Computer and System Sciences*, 125:59–105, 2022. doi:10.1016/j.jcss.2021.11.003.

8 Yong Chen, Zhi-Zhong Chen, Guohui Lin, Yao Xu, and An Zhang. Approximation Algorithms for Maximally Balanced Connected Graph Partition. *Algorithmica*, 83:3715–3740, 2021. doi:10.1007/s00453-021-00870-3.

9 Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002. doi:10.1073/pnas.252631999.

10 Fan Chung and Linyuan Lu. Connected Components in Random Graphs with Given Expected Degree Sequences. *Annals of Combinatorics*, 6(2):125–145, 2002. doi:10.1007/PL00012580.

11 Josep Díaz, Dieter Mitsche, and Xavier Pérez-Giménez. On the Connectivity of Dynamic Random Geometric Graphs. In *Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 601–610, 2008. URL: https://dl.acm.org/doi/abs/10.5555/1347082.1347149.

12 Paul Erdős, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

13 Paul Erdős and Alfréd Rényi. On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

14 Nikolaos Fountoulakis and Tobias Müller. Law of large numbers for the largest component in a hyperbolic model of complex networks. *The Annals of Applied Probability*, 28(1):607–650, 2018. URL: https://www.jstor.org/stable/26542317.

15 Tobias Friedrich and Anton Krohmer. On the Diameter of Hyperbolic Random Graphs. *SIAM Journal on Discrete Mathematics*, 32(2):1314–1334, 2018. doi:10.1137/17M1123961.

16 Daniel Gross, Monika Heinig, L. Iswara, L.W. Kazmierczak, Kristi Luttrell, John Saccoman, and C. Suffel. A Survey of Component Order Connectivity Models of Graph Theoretic Networks. *WSEAS Transactions on Mathematics*, 12:895–910, 2013.

17 Luca Gugelmann, Konstantinos Panagiotou, and Ueli Peter. Random Hyperbolic Graphs: Degree Sequence and Clustering. In *39th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 573–585, 2012. doi:10.1007/978-3-642-31585-5_51.

18 P. Gupta and P.R. Kumar. Critical power for asymptotic connectivity. In *37th IEEE Conference on Decision and Control*, volume 1, pages 1106–1110, 1998. doi:10.1109/CDC.1998.760846.

19 Ralph Keusch. *Geometric Inhomogeneous Random Graphs and Graph Coloring Games*. PhD thesis, ETH Zurich, 2018. doi:10.3929/ethz-b-000269658.

20 Marcos Kiwi and Dieter Mitsche. On the second largest component of random hyperbolic graphs. *SIAM Journal on Discrete Mathematics*, 33(4):2200–2217, 2019. doi:10.1137/18M121201X.

21 Júlia Komjáthy, John Lapinskas, and Johannes Lengler. Penalising transmission to hubs in scale-free spatial random graphs. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(4):1968–2016, 2021. doi:10.1214/21-AIHP1149.

22 Júlia Komjáthy and Bas Lodewijks. Explosion in weighted hyperbolic random graphs and geometric inhomogeneous random graphs. *Stochastic Processes and their Applications*, 130(3):1309–1367, 2020. doi:10.1016/j.spa.2019.04.014.

23 Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82:036106, 2010. doi:10.1103/PhysRevE.82.036106.

24 Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. doi:10.1017/CBO9780511813603.

25 Tobias Müller and Merlijn Staps. The Diameter of KPKVB Random Graphs. *Advances in Applied Probability*, 51(2):358–377, 2019. doi:10.1017/apr.2019.23.

26 Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

27 Stephan Schwartz. An overview of graph covering and partitioning. *Discrete Mathematics*, 345(8):112884, 2022. doi:10.1016/j.disc.2022.112884.

28 Bo Söderberg. General formalism for inhomogeneous random graphs. *Phys. Rev. E*, 66:066121, 2002. doi:10.1103/PhysRevE.66.066121.