Cliques in High-Dimensional Geometric Inhomogeneous Random Graphs

Tobias Friedrich
Hasso Plattner Institute, University of Potsdam, Germany

Andreas Göbel
Hasso Plattner Institute, University of Potsdam, Germany

Maximilian Katzmann
Karlsruhe Institute of Technology, Karlsruhe, Germany

Leon Schiller
Hasso Plattner Institute, University of Potsdam, Germany

Abstract

A recent trend in the context of graph theory is to bring theoretical analyses closer to empirical observations, by focusing the studies on random graph models that are used to represent practical instances. There, it was observed that geometric inhomogeneous random graphs (GIRGs) yield good representations of complex real-world networks, by expressing edge probabilities as a function that depends on (heterogeneous) vertex weights and distances in some underlying geometric space that the vertices are distributed in. While most of the parameters of the model are understood well, it was unclear how the dimensionality of the ground space affects the structure of the graphs.

In this paper, we complement existing research into the dimension of geometric random graph models and the ongoing study of determining the dimensionality of real-world networks, by studying how the structure of GIRGs changes as the number of dimensions increases. We prove that, in the limit, GIRGs approach non-geometric inhomogeneous random graphs and present insights on how quickly the decay of the geometry impacts important graph structures. In particular, we study the expected number of cliques of a given size as well as the clique number and characterize phase transitions at which their behavior changes fundamentally. Finally, our insights help in better understanding previous results about the impact of the dimensionality on geometric random graphs.

2012 ACM Subject Classification Mathematics of computing → Random graphs; Theory of computation → Computational geometry

Keywords and phrases random graphs, geometry, dimensionality, cliques, clique number, scale-free networks

Digital Object Identifier 10.4230/LIPIcs.ICALP.2023.22

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://arxiv.org/abs/2302.04113` [22]

Funding *Andreas Göbel*: is funded by the project PAGES (project No. 467516565) of the German Research Foundation (DFG).

1 Introduction

Networks are a powerful tool to model all kinds of processes that we interact with in our day-to-day lives. From connections between people in social networks, to the exchange of information on the internet, and on to how our brains are wired, networks are everywhere. Consequently, they have been in the focus of computer science for decades. There, one of the most fundamental techniques used to model and study networks are *random graph models*. Such a model defines a probability distribution over graphs, which is typically done by specifying a random experiment on how to construct the graph. By analyzing the rules of the experiment, we can then derive structural and algorithmic properties of the resulting graphs. If the results match what we observe on real-world networks, i.e., if the model represents the graphs we encounter in practice well, then we can use it to make further predictions that help us understand real graphs and utilize them more efficiently.

The quest of finding a good model starts several decades ago, with the famous Erdős-Rényi (ER) random graphs [19, 24]. There, all edges in the graph exist independently with the same probability. Due to its simplicity, this model has been studied extensively. However, because the degree distribution of the resulting graphs is rather homogeneous and they lack clustering (due to the independence of the edges), the model is not considered to yield good representations of real graphs. In fact, many networks we encounter in practice feature a degree distribution that resembles a power-law [3, 38, 39] and the clustering coefficient (the probability for two neighbors of a vertex to be adjacent) is rather high [35, 40]. To overcome these drawbacks, the initial random graph model has been adjusted in several ways.

In *inhomogeneous random graphs (IRGs)*, often referred to as *Chung-Lu random graphs*, each vertex is assigned a weight and the probability for two vertices to be connected by an edge is proportional to the product of the weights [1, 11, 12]. As a result, the expected degrees of the vertices in the resulting graphs match their weight. While assigning weights that follow a power-law distribution yields graphs that are closer to the complex real-world networks, the edges are still drawn independently, leading to vanishing clustering coefficients.

A very natural approach to facilitate clustering in a graph model is to introduce an underlying geometry. This was done first in *random geometric graphs (RGGs)*, where vertices are distributed uniformly at random in the Euclidean unit square and any two are connected by an edge if their distance lies below a certain threshold, i.e., the neighborhood of a vertex lives in a disk centered at that vertex [36]. Intuitively, two vertices that connect to a common neighbor cannot be too far away from each other, increasing the probability that they are connected by an edge themselves. In fact, random geometric graphs feature a non-vanishing clustering coefficient [13]. However, since all neighborhood disks have the same size, they all have roughly the same expected degree, again, leading to a homogeneous degree distribution.

To get a random graph model that features a heterogeneous degree distribution *and* clustering, the two mentioned adjustments were recently combined to obtain *geometric inhomogeneous random graphs (GIRGs)* [28]. There, vertices are assigned a weight *and* a position in some underlying geometric space and the probability for two vertices to connected increases with the product of the weights but decreases with increasing geometric distance between them. As a result, the generated graphs have a non-vanishing clustering coefficient and, with the appropriate choice of the weight sequence, they feature a power-law degree distribution. Additionally, recent empirical observations indicate that GIRGs represent real-world networks well with respect to certain structural and algorithmic properties [5].

We note that GIRGs are not the first model that exhibits a heterogeneous degree distribution and clustering. In fact, *hyperbolic random graphs (HRGs)* [30] feature these

properties as well and have been studied extensively before (see, e.g., [7, 20, 21, 23, 26]). However, in the pursuit of finding good models to represent real-world networks, GIRGs introduce a parameter that sets them apart from prior models: the choice of the underlying geometric space and, more importantly, the dimensionality of that space.

Unfortunately, this additional parameter that sets GIRGs apart from previous models, has not gained much attention at all. In fact, it comes as a surprise that, while the underlying dimensionality of real-world networks is actively researched [2, 8, 15, 25, 31] and there is a large body of research examining the impact of the dimensionality on different homogeneous graph models [13, 17, 18] with some advancements being made on hyperbolic random graphs [41], the effects of the dimension on the structure of GIRGs have only been studied sparsely. For example, while it is known that GIRGs exhibit a clustering coefficient of $\Theta(1)$ for any fixed dimension [28], it is not known how the hidden constants scale with the dimension.

In this paper, we initiate the study of the impact of the dimensionality on GIRGs. In particular, we investigate the influence of the underlying geometry as the dimensionality increases, proving that GIRGs converge to their non-geometric counterpart (IRGs) in the limit. With our results we are able to explain seemingly disagreeing insights from prior research on the impact of dimensionality on geometric graph models. Moreover, by studying the clique structure of GIRGs and its dependence on the dimension d, we are able to quantify how quickly the underlying geometry vanishes. In the following, we discuss our results in greater detail. We note that, while we give general proof sketches for our results, the complete proofs are deferred to the full version [22].

2 (Geometric) Inhomogeneous Random Graphs

Before stating our results in greater detail, let us recall the definitions of the two graph models we mainly work with throughout the paper.

Inhomogeneous Random Graphs (IRGs). The model of inhomogeneous random graphs was introduced by Chung and Lu [1, 11, 12] and is a natural generalization of the Erdős-Rényi model. Starting with a vertex set V of n vertices, each $v \in V$ is assigned a weight w_v. Each edge $\{u, v\} \in \binom{V}{2}$ is then independently present with probability

$$\Pr[u \sim v] = \min\left\{1, \frac{\lambda w_u w_v}{n}\right\},$$

for some constant $\lambda > 0$ controlling the average degree of the resulting graph. Note that assigning the same weight to all vertices yields the same connection probability as in Erdős-Rényi random graphs. For the sake of simplicity, we define $\kappa_{uv} = \min\{\lambda w_u w_v, n\}$ such that $\Pr[u \sim v] = \kappa_{uv}/n$. Additionally, for a set of vertices $U_k = \{v_1, \dots, v_k\}$ with weights $w_1, \dots, w_k$, we introduce the shorthand notation $\kappa_{ij} = \kappa_{v_i v_j}$ and write $\{\kappa\}^{(k)} = \{\kappa_{ij} \mid 1 \leq i < j \leq k\}$.

Throughout the paper, we mainly focus on inhomogeneous random graphs that feature a power-law degree distribution in expectation, which is obtained by sampling the weights accordingly. More precisely, for each $v \in V$, we sample a weight w_v from the Pareto distribution $\mathcal{P}$ with parameters $1 - \beta, w_0$ and distribution function

$$\Pr[w_v \leq x] = 1 - \left(\frac{x}{w_0}\right)^{1-\beta}.$$

Then the density of w_v is $\rho_{w_v}(x) = \frac{\beta-1}{w_0^{1-\beta}} x^{-\beta}$. Here, $w_0 > 0$ is a constant that represents a lower bound on the weights in the graph and β denotes the power-law exponent of the resulting

degree distribution. Throughout the paper, we assume $\beta > 2$ such that a single weight has finite expectation (and thus the average degree in the graph is constant), but possibly infinite variance. We denote a graph obtained by utilizing the above weight distribution and connection probabilities with $\mathrm{IRG}(n, \beta, w_0)$. For a fixed weight sequence $\{w\}_1^n$, we denote the corresponding graph by $\mathrm{IRG}(\{w\}_1^n)$.

Geometric Inhomogeneous Random Graphs (GIRGs). Geometric inhomogeneous random graphs are an extension of IRGs, where in addition to the weight, each vertex v is also equipped with a position $\mathbf{x}_v$ in some geometric space and the probability for edges to form depends on their weights and the distance in the underlying space [28]. While, in its raw form, the GIRG framework is rather general, we align our paper with existing analysis on GIRGs [6, 29, 34] and consider the d-dimensional torus $\mathbb{T}^d$ equipped with L_∞-norm as the geometric ground space. More precisely, in what we call the *standard* GIRG model, the positions $\mathbf{x}$ of the vertices are drawn independently and uniformly at random from $\mathbb{T}^d$, according to the standard Lebesgue measure. We denote the i-th component of $\mathbf{x}_v$ by $\mathbf{x}_{vi}$. Additionally, the geometric distance between two points $\mathbf{x}_u$ and $\mathbf{x}_v$, is given by

$$d(\mathbf{x}_u, \mathbf{x}_v) = \|\mathbf{x}_u - \mathbf{x}_v\|_\infty = \max_{1 \le i \le d} \{|\mathbf{x}_{ui} - \mathbf{x}_{vi}|_C\},$$

where $|\cdot|_C$ denotes the distance on the circle, i.e,

$$|\mathbf{x}_{ui} - \mathbf{x}_{vi}|_C = \min\{|\mathbf{x}_{ui} - \mathbf{x}_{vi}|, 1 - |\mathbf{x}_{ui} - \mathbf{x}_{vi}|\}.$$

In a standard GIRG, two vertices $u \neq v$ are adjacent if and only if their distance $d(\mathbf{x}_u, \mathbf{x}_v)$ in the torus is less than or equal to a *connection threshold* t_{uv}, which is given by

$$t_{uv} = \frac{1}{2}\left(\frac{\lambda w_u w_v}{n}\right)^{1/d} = \left(\frac{w_u w_v}{\tau n}\right)^{1/d},$$

where $\tau = 2^d/\lambda$. Using L_∞ is motivated by the fact that it is the most widely used metric in the literature because it is arguably the most natural metric on the torus. In particular, it has the "nice" property that the ball of radius r is a cube and "fits" entirely into $\mathbb{T}^d$ for all $0 \le r \le 1$.

Note that, as a consequence of the above choice, the marginal connection probability $\Pr[u \sim v]$ is the same as in the IRG model, i.e., $\Pr[u \sim v] = \kappa_{uv}/n$. However, while the probability that any given edge is present is the same as in the IRG model, the edges in the GIRG model are *not* drawn independently. We denote a graph obtained by the procedure described above with $\mathrm{GIRG}(n, \beta, w_0, d)$. As for IRGs, we write $\mathrm{GIRG}(\{w\}_1^n, d)$ when considering standard GIRGs with a fixed weight sequence $\{w\}_1^n$.

As mentioned above, the standard GIRG model is a commonly used instance of the more general GIRG framework [28]. There, different geometries and distance functions may be used. For example, instead of L_∞-norm, any L_p-norm for $1 \le p < \infty$ may be used. Then, the distance between two vertices u, v is measured as

$$\|\mathbf{x}_u - \mathbf{x}_v\|_p := \begin{cases} \left(\sum_{i=1}^d |\mathbf{x}_{ui} - \mathbf{x}_{vi}|^p\right)^{1/p} & \text{if } p < \infty \\ \max_{1 \le i \le d}\{|\mathbf{x}_{ui} - \mathbf{x}_{vi}|\} & \text{otherwise.} \end{cases}$$

With this choice, the volume (Lebesgue measure) of the ball $B_p(r)$ of radius r under L_p-norm is equal to the probability that a vertex u falls within distance at most r of v (if $r = o(1)$).

We denote this volume by $\nu(r)$. We call the corresponding graphs *standard GIRGs with any L_p-norm* and note that some of our results extend to this more general model. Finally, whenever our insights consider an even broader variant of the model (e.g., variable ground spaces, distances functions, weight distributions), we say that they hold for *any GIRG* and mention the constraints explicitly.

3 Asymptotic Equivalence

Our first main observations is that large values of d diminish the influence of the underlying geometry until, at some point, our model becomes strongly equivalent to its non-geometric counterpart, where edges are sampled independently of each other. We prove that the *total variation distance* between the distribution over all graphs of the two models tends to zero as n is kept fixed and $d \to \infty$. We define the total variation distance of two probability measures P and Q on the measurable space $(\Omega, \mathcal{F})$ as

$$\|P, Q\|_{\mathrm{TV}} = \sup_{A \in \mathcal{F}} |P(A) - P(B)| = \frac{1}{2} \sum_{\omega \in \Omega} |P(\omega) - Q(\omega)|,$$

where the second equality holds if Ω is countable. In our case, Ω is the set $\mathcal{G}(n)$ of all possible graphs on n vertices, and P, Q are distributions over these graphs. If G_1, G_2 are two random variables mapping to Ω, we refer to $\|G_1, G_2\|_{\mathrm{TV}}$ as the total variation distance of the induced probability measures by G_1 and G_2, respectively. Informally, this measures the maximum difference in the probability that any graph G is sampled by G_1 and G_2.

▶ **Theorem 1.** *Let $\mathcal{G}(n)$ be the set of all graphs with n vertices, let $\{w\}_1^n$ be a weight sequence, and consider $G_{IRG} = \mathrm{IRG}(\{w\}_1^n) \in \mathcal{G}(n)$ and a standard GIRG $G_{GIRG} = \mathrm{GIRG}(\{w\}_1^n, d) \in \mathcal{G}(n)$ with any L_p-norm. Then,*

$$\lim_{d \to \infty} \|G_{GIRG}, G_{IRG}\|_{TV} = 0.$$

We note that this theorem holds for arbitrary weight sequences that do not necessarily follow a power law and for arbitrary L_p-norms used to define distances in the ground space. For $p \in [1, \infty)$, the proof is based on the application of a multivariate central limit theorem [37], in a similar way as used to prove a related statement for *spherical random geometric graphs (SRGGs)*, i.e., random geometric graphs with a hypersphere as ground space [17]. Our proof generalizes this argument to arbitrary L_p-norms and arbitrary weight sequences. For the case of L_∞-norm, we present a proof based on the inclusion-exclusion principle and the bounds we develop in the full version [22, Section 4].

Remarkably, while a similar behavior was previously established for SRGGs, there exist works indicating that RGGs on the hypercube do not converge to their non-geometric counterpart [13, 18] as $d \to \infty$. We show that this apparent disagreement is due to the fact that the torus is a homogeneous space while the hypercube is not. In fact, our proof shows that GIRGs on the hypercube *do* converge to a non-geometric model in which edges are, however, not sampled independently. This lack of independence is because, on the hypercube, there is a positive correlation between the distances from two vertices to a given vertex, leading to a higher tendency to form clusters, as was observed experimentally [18]. Due to the homogeneous nature of the torus, the same is not true for GIRGs and the model converges to the plain IRG model with independent edges.

Table 1 Asymptotic behavior of the expected number of k-cliques. The behavior in the first column is the same as in hyperbolic random graphs [7], and the behavior in the third column is the same as in the IRG model [14]. Results marked with * were previously known for constant k [34].

	$\mathbb{E}[K_k]$ for $k \geq 4$		
	$d = \Theta(1)$	$d = o(\log(n))$	$d = \omega(\log(n))$
$2 < \beta < 3, k > \frac{2}{3-\beta}$	$n^{\frac{k}{2}(3-\beta)}\Theta(k)^{-k}*$	$n^{\frac{k}{2}(3-\beta)}\Theta(k)^{-k}$	$n^{\frac{k}{2}(3-\beta)}\Theta(k)^{-k}$
$2 < \beta < 3, k < \frac{2}{3-\beta}$	$n\Theta(k)^{-k}*$	$ne^{-\Theta(1)dk}\Theta(k)^{-k}$	$n^{\frac{k}{2}(3-\beta)}\Theta(k)^{-k}$
$\beta > 3$	$n\Theta(k)^{-k}$	$ne^{-\Theta(1)dk}\Theta(k)^{-k}$	$o(1)$

4 Clique Structure

To quantify for which dimensions d the graphs in the GIRG model start to behave similar to IRGs, we investigate the number and size of cliques. Previous results on SRGGs indicate that the dimension of the underlying space heavily influences the clique structure of the model [4, 17]. However, it was not known how the size and the number of cliques depends on d if we use the torus as our ground space, and how the clique structure in high-dimensions behaves for inhomogeneous weights.

We give explicit bounds on the expected number of cliques of a given size k, which we afterwards turn into bounds on the *clique number* $\omega(G)$, i.e., the size of the largest clique in the graph G. While the expected number of cliques in the GIRG model was previously studied by Michielan and Stegehuis [34] when the power-law exponent of the degree distribution satisfies $\beta \in (2,3)$, to the best of our knowledge, the clique number of GIRGs remains unstudied even in the case of constant (but arbitrary) dimensionality. We close this gap, reproduce the existing results, and extend them to the case $\beta \geq 3$ and the case where d can grow as a function of the number of vertices n in the graph. Furthermore, our bounds for the case $\beta \in (2,3)$ are more explicit and complement the work of Michielan and Stegehuis, who expressed the (rescaled) asymptotic number of cliques as converging to a non-analytically solvable integral. Furthermore, we show that the clique structure in our model eventually behaves asymptotically like that of an IRG if the dimension is sufficiently large. In summary, our main contributions are outlined in Tables 1, 2, and Table 3.

We observe that the structure of the cliques undergoes three phase transitions in the size of the cliques k, the dimension d, and the power-law exponent β.

Transition in k. When $\beta \in (2,3)$ and $d \in o(\log(n))$, the first transition is at $k = \frac{2}{3-\beta}$, as was previously observed for hyperbolic random graphs [7] and for GIRGs of constant dimensionality [34]. The latter work explains this behavior by showing that for $k < \frac{2}{3-\beta}$, the number of cliques is strongly dominated by "geometric" cliques forming among vertices whose distance is of order $n^{-1/d}$ regardless of their weight. For $k > \frac{2}{3-\beta}$, on the other hand, the number of cliques is dominated by "non-geometric" cliques forming among vertices with weights in the order of $\sqrt{n}$. This behavior is in contrast to the behavior of cliques in the IRG model, where this phase transition does not exist and where the expected number of k cliques is $\Theta\left(n^{\frac{k}{2}(3-\beta)}\right)$ for all $k \geq 3$ (if $\beta \in (2,3)$) [14].

Transition in d. Still assuming $\beta \in (2,3)$, the second phase transition occurs as d becomes superlogarithmic. More precisely, we show that in the high-dimensional regime, where $d = \omega(\log(n))$, the phase transition in k vanishes, as the expected number of cliques of size

Table 2 Asymptotic behavior of the expected number of triangles. The case $\beta = \infty$ refers to the case of constant weights. While in the case $\beta < 3$, the number of triangles already behaves like that of the IRG model if $d = \omega(\log(n))$, in the case $\beta > 3$, the number of triangles remains superconstant as long as $d = o\left(\log^{3/2}(n)\right)$.

Expected number of triangles $\mathbb{E}[K_3]$			
	$d = o(\log(n))$	$d = \omega(\log(n))$	$d = \omega(\log^2(n))$
$2 < \beta < \frac{7}{3}$	$n^{\frac{3}{2}(3-\beta)}\Theta(1)$	$n^{\frac{3}{2}(3-\beta)}\Theta(1)$	$n^{\frac{3}{2}(3-\beta)}\Theta(1)$
$\frac{7}{3} < \beta < 3$	$ne^{-\Theta(1)d}\Theta(1)$	$n^{\frac{3}{2}(3-\beta)}\Theta(1)$	$n^{\frac{3}{2}(3-\beta)}\Theta(1)$
$\beta > 3$	$ne^{-\Theta(1)d}\Theta(1)$	$\Omega\left(\exp\left(\frac{\ln^3(n)}{d^2}\right)\right)$	$\Theta(1)$
$\beta = \infty$	$ne^{-\Theta(1)d}\Theta(1)$	$\Theta\left(\exp\left(\frac{\ln^3(n)}{d^2}\right)\right)$	$\Theta(1)$

$k \geq 4$ behaves asymptotically like its counterpart in the IRG model. Nevertheless, we can still differentiate the two models as long as $d = o(\log^{3/2}(n))$, by counting triangles among low degree vertices as can be seen in Table 2.

The reason for this behavior is that the expected number of cliques in the case $d = \omega(\log(n))$ is already dominated by cliques forming among vertices of weight close to $\sqrt{n}$. For those, the probability that a clique is formed already behaves like in an IRG although, for vertices of small weight, said probability it is still larger.

Regarding the clique number, in the case $\beta > 3$, we observe a similar phase transition in d. For constant d, the clique number of a GIRG is $\Theta(\log(n)/\log\log(n)) = \omega(1)$. We find that this asymptotic behavior remains unchanged if $d = \mathcal{O}(\log\log(n))$. However, if $d = \omega(\log\log(n))$ but $d = o(\log(n))$, the clique number scales as $\Theta(\log(n)/d)$, which is still superconstant. Additionally if $d = \omega(\log(n))$, we see that, again, GIRGs show the same behavior as IRGs. That is, there are asymptotically no cliques of size larger than 3.

Transition in β. The third phase transition occurs at $\beta = 3$ in the high-dimensional case, which is in line with the fact that networks with a power-law exponent $\beta \in (2, 3)$ contain with high probability (w.h.p., meaning with probability $1 - O(1/n)$) a densely connected "heavy core" of $\Theta\left(n^{\frac{1}{2}(3-\beta)}\right)$ vertices with weight $\sqrt{n}$ or above, which vanishes if β is larger than 3. This heavy core strongly dominates the number of cliques of sufficient size and explains why the clique number is $\Theta\left(n^{\frac{1}{2}(3-\beta)}\right)$ regardless of d if $\beta \in (2, 3)$. As β grows beyond 3, the core disappears and leaves only very small cliques. Accordingly for $\beta > 3$ IRGs contain asymptotically almost surely (a.a.s., meaning with probability $1 - o(1)$) no cliques of size greater than 3. In contrast to that, for GIRGs of dimension $d = o(\log(n))$ (and HRGs), the clique number remains superconstant and so does the number of k-cliques for any constant $k \geq 3$. If $d = \omega(\log(n))$, there are no cliques of size greater than 3 like in an IRG. However, as noted before, GIRGs feature many more triangles than IRGs as long as $d = o(\log^{3/2}(n))$.

Beyond the three mentioned phase transitions, we conclude that, for constant d, the main difference between GIRGs and IRGs is that the former contain a significant number of cliques that form among vertices of low weight, whereas, in the latter model only high-weight vertices contribute significantly to the total number of cliques. In fact, here, the expected number of k cliques in the heavy core is already of the same order as the total expectation of K_k in the whole graph. Similarly, in the GIRG model, the expected number of cliques forming in the low-weight subgraph $G_{\leq w}$ for some constant w, is already of the same order as

the total number of cliques if $k < \frac{2}{3-\beta}$ or $\beta \geq 3$ (otherwise, this number is, again, dominated by cliques from the heavy core).

The proofs of our results (i.e., the ones in the above tables) are mainly based on bounds on the probability that a set of k randomly chosen vertices forms a clique. To obtain concentration bounds on the number of cliques as needed for deriving bounds on the clique number, we use the second moment method and Chernoff bounds.

For the case of $d = \omega(\log(n))$, many of our results are derived from the following general insight. We show that for and all $\beta > 2$, the probability that a set of vertices forms a clique already behaves similar as in the IRG model if the weights of the involved nodes are sufficiently large. For $d = \omega(\log(n)^2)$, this holds in the entire graph, that is, regardless of the weights of the involved vertices. In fact our statement holds even more generally. That is, the described behavior not only applies to the probability that a clique is formed but also to the probability that any set of edges (or a superset thereof) is created.

▶ **Theorem 2.** *Let G be a standard GIRG and let $k \geq 3$ be a constant. Furthermore, let $U_k = \{v_1, \dots, v_k\}$ be a set of vertices chosen uniformly at random and let $\{\kappa\}^{(k)} = \{\kappa_{ij} \mid 1 \leq i, j \leq k\}$ describe the pairwise product of weights of the vertices in U_k. Let $E(U_k)$ denote the (random) set of edges formed among the vertices in U_k. Then, for $d = \omega\left(\log^2(n)\right)$ and any set of edges $\mathcal{A} \subseteq \binom{U_k}{2}$,*

$$\Pr\left[E(U_k) \supseteq \mathcal{A} \mid \{\kappa\}^{(k)}\right] = (1 \pm o(1)) \prod_{\{i,j\} \in \mathcal{A}} \frac{\kappa_{ij}}{n}.$$

If $d = \omega\left(\log(n)\right)$,

$$\Pr\left[E(U_k) \supseteq \mathcal{A} \mid \{\kappa\}^{(k)}\right] = (1 \pm o(1)) \prod_{\{i,j\} \in \mathcal{A}} \left(\frac{\kappa_{ij}}{n}\right)^{1 \mp \mathcal{O}\left(\frac{\log(n)}{d}\right)}.$$

For the proof we derive elementary bounds on the probability of the described events and use series expansions to investigate their asymptotic behavior. Remarkably, in contrast to our bounds for the case $d = o(\log(n))$, the high-dimensional case requires us to pay closer attention to the topology of the torus.

We leverage the above theorem to prove that GIRGs eventually become equivalent to IRGs with respect to the total variation distance. Theorem 2 already implies that the expected number of cliques in a GIRG is asymptotically the same as in an IRG for all $k \geq 3$ and all $\beta > 2$ if $d = \omega(\log^2(n))$. However, we are able to show that the expected number of cliques for $\beta \in (2, 3)$ actually already behaves like that of an IRG if $d = \omega(\log(n))$. The reason for this is that the clique probability among high-weight vertices starts to behave like that of an IRG earlier than it is the case for low-weight vertices and cliques forming among these high-weight vertices already dominate the number of cliques. Moreover, the clique number behaves like that of an IRG if $d = \omega(\log(n))$ for all $\beta > 2$. However, the number of triangles among vertices of constant weight asymptotically exceeds that of an IRG as long as $d = o(\log^{3/2}(n))$, which we prove by deriving even sharper bounds on the expected number of triangles. Accordingly, convergence with respect to the total variation distance cannot occur before this point (this holds for all $\beta > 2$).

In contrast to this, for the low-dimensional case (where $d = o(\log(n))$), the underlying geometry still induces strongly notable effects regarding the number of sufficiently small cliques for all $\beta > 2$. However, even here, the expected number of such cliques decays exponentially in dk. The main difficulty in showing this is that we have to handle the case of

Table 3 Asymptotic behavior of the clique number of G for different values of d in the GIRG model. The behavior of the first column is the same as in hyperbolic random graphs established in [7], and the behavior in the third column is the same as that of IRG graphs established in [27]. All results hold a.a.s. and under L_∞-norm.

	$\omega(G)$		
	$d = \mathcal{O}(\log\log(n))$	$d = o(\log(n))$	$d = \omega(\log(n))$
$\beta < 3$	$\Theta\left(n^{(3-\beta)/2}\right)$	$\Theta\left(n^{(3-\beta)/2}\right)$	$\Theta\left(n^{(3-\beta)/2}\right)$
$\beta = 3$	$\Theta\left(\frac{\log(n)}{\log\log(n)}\right)$	$\Omega\left(\frac{\log(n)}{d}\right)$	$\mathcal{O}(1)$
$\beta > 3$	$\Theta\left(\frac{\log(n)}{\log\log(n)}\right)$	$\Theta\left(\frac{\log(n)}{d}\right)$	≤ 3
	equivalent to HRGs [7]		equivalent to IRGs [27]

inhomogeneous weights, which significantly influence the probability that a set of k vertices chosen uniformly at random forms a clique. To this end, we prove the following theorem that bounds the probability that a clique among k vertices is formed if the ratio of the maximal and minimal weight is at most c^d. Note that the vertices forming a star is necessary for a clique to form. For this reason we consider the event $\mathbf{E}^c_{\text{star}}$ of the vertices forming a star centered at the lowest weight vertex. The theorem generalizes a result of Decreusefond et al. [16].

▶ **Theorem 3.** *Let G be a standard GIRG and consider $k \geq 3$. Furthermore, let $U_k = \{v_1, v_2, \dots, v_k\}$ be a set of vertices chosen uniformly at random and assume without loss of generality that $w_1 \leq \dots \leq w_k$. Let $\mathbf{E}^c_{star}$ be the event that v_1 connects to all vertices in $U_k \setminus \{v_1\}$ and that $w_k \leq c^d w_1$ for some constant $c \geq 1$ with $c^2 \left(w_1^2/(\tau n)\right)^{1/d} \leq 1/4$. Then, the probability that U_k is a clique conditioned on $\mathbf{E}^c_{star}$ fulfills*

$$\left(\frac{1}{2}\right)^{d(k-1)} k^d \leq \Pr\left[U_k \text{ is clique} \mid \mathbf{E}^c_{star}\right] \leq c^{d(k-2)} \left(\frac{1}{2}\right)^{d(k-1)} k^d.$$

Building on the variant by Decreusefond et al. [16], we provide an alternative proof of the original statement, showing that the clique probability conditioned on the event $\mathbf{E}^c_{\text{star}}$ is monotonous in the weight of all other vertices. Remarkably, this only holds if we condition on the event that the center of our star is of minimal weight among the vertices in U_k.

We apply Theorem 3 to bound the clique probability in the whole graph (where the ratio of the maximum and minimum weight of vertices in U_k is not necessarily bounded). Afterwards, we additionally use Chernoff bounds and the second moment method to bound the clique number.

5 Relation to Previous Analyses

In the following, we discuss how our results compare to insights obtained on similar graph models that (apart from not considering weighted vertices) mainly differ in the considered ground space. We not that, in the following, we consider GIRGs with uniform weights in order to obtain a valid comparison.

Random Geometric Graphs on the Sphere. Our results indicate that the GIRG model on the torus behaves similarly to the model of Spherical Random Geometric Graphs (SRGGs)

in the high-dimensional case. In this model, vertices are distributed on the surface of a $d-1$ dimensional sphere and an edge is present whenever the Euclidean distance between two points (measured by their inner product) falls below a given threshold. Analogous to the behavior of GIRGs, when keeping n fixed and considering increasing $d \to \infty$, this model converges to its non-geometric counterpart, which in their case is the Erdős–Rényi model [17]. It is further shown that the clique number converges to that of an Erdős–Rényi graph (up to a factor of $1 + o(1)$) if $d = \omega(\log^3(n))$.

Although the overall behavior of SRGGs is similar to that of GIRGs, the magnitude of d in comparison to n at which non-geometric features become dominant seems to differ. In fact, it is shown in [10, proof of Theorem 3] that the expected number of triangles in sparse SRGGs still grows with n as long as $d = o(\log^3(n))$, whereas its expectation is constant in the non-geometric, sparse case (as for Erdős–Rényi graphs). On the other hand, in the GIRG model, we show that the expected number of triangles in the sparse case converges to the same (constant) value as that of the non-geometric model if only $d = \omega(\log^{3/2}(n))$. This indicates that, in the high-dimensional regime, differences in the nature of the underlying geometry result in notably different behavior, whereas in the case of constant dimensionality, the models are often assumed to behave very similarly.

Random Geometric Graphs on the Hypercube. The work of Dall and Christensen [13] and the recent work of Erba et al. [18] show that RGGs on the hypercube do *not* converge to Erdős–Rényi graphs as n is fixed and $d \to \infty$. However, our results imply that this is the case for RGGs on the torus. These apparent disagreements are despite the fact that Erba et al. use a similar central limit theorem for conducting their calculations and simulations [18].

The tools established in our paper yield an explanation for this behavior. Our proof of Theorem 1 relies on the fact that, for independent zero-mean variables $Z_1, \dots, Z_d$, the covariance matrix of the random vector $Z = \sum_{i=1}^{d} Z_i$ is the identity matrix. This, in turn, is based on the fact that the torus is a *homogeneous space*, which implies that the probability measure of a ball of radius r (proportional to its Lebesgue measure or volume, respectively) is the same, regardless of where this ball is centered. It follows that the random variables $Z_{(u,v)}$ and $Z_{(u,s)}$, denoting the normalized distances from u to s and v, respectively, are independent. As a result their covariance is 0 although both "depend" on the position of u.

For the hypercube, this is not the case. Although one may analogously define the distance of two vertices as a sum of independent, zero-mean random vectors over all dimensions just like we do in this paper, the random variables $Z_{(u,v)}$ and $Z_{(u,s)}$ do *not* have a covariance of 0.

6 Conjectures & Future Work

While making the first steps towards understanding GIRGs and sparse RGGs on the torus in high dimensions, we encountered several questions whose investigation does not fit into the scope of this paper. In the following, we give a brief overview of our conjectures and possible starting points for future work.

In addition to investigating how the number and size of cliques depends on d, it remains to analyze among which vertices k-cliques form dominantly. For constant d and $\beta \in (2, 3)$ this was previously done by Michielan and Stegehuis who noted that cliques of size $k > \frac{2}{3-\beta}$ are dominantly formed among vertices of weight in the order of $\sqrt{n}$ like in the IRG model, whereas cliques of size $k < \frac{2}{3-\beta}$ dominantly appear among vertices within distance in the order of $n^{-1/d}$ [34]. This characterizes the geometric and non-geometric nature of cliques of size larger and smaller than $\frac{2}{3-\beta}$, respectively. As our work indicates that this phase

transition vanishes as $d = \omega(\log(n)))$, we conjecture that in this regime cliques of all sizes are dominantly found among vertices of weight in the order $\sqrt{n}$. For the case $\beta \geq 3$ it remains to analyze the position of cliques of all sizes. It would further be interesting to find out where cliques of superconstant size are dominantly formed as previous work in this regard only holds for constant k.

Additionally, it would be interesting to extend our results to a noisy variant of GIRGs. While the focus in this paper lies on the standard GIRGs, where vertices are connected by an edge if their distance is below a given threshold, there is a *temperate* version of the model, where the threshold is softened using a *temperature* parameter. That is, while the probability for an edge to exist still decreases with increasing distance, we can now have longer edges and shorter non-edges with certain probabilities. The motivation of this variant of GIRGs is based on the fact that real data is often noisy as well, leading to an even better representation of real-world graphs.

We note that we expect our insights to carry over to the temperate model, as long as we have constant temperature. Beyond that, we note that both temperature and dimensionality affect the influence of the underlying geometry. Therefore, it would be interesting to see whether a sufficiently high temperature has an impact on how quickly GIRGs converge to the IRGs.

Furthermore, it remains to investigate the dense case of our model, where the marginal connection probability of any pair of vertices is constant and does not decrease with n. For dense SRGGs, an analysis of the high-dimensional case has shown that the underlying geometry remains detectable as long as $d = o(n^3)$. As mentioned above, GIRGs and their non-geometric counterpart can be distinguished as long as $d = o(\log^{3/2}(n))$, by considering triangles among low-weight vertices. For dense SRGGs the geometry can be detected by counting so-called *signed triangles* [10]. Although for the sparse case, signed triangles have no advantage over ordinary triangles, they are much more powerful in the dense case and might hence prove useful for our model in the dense case as well.

Another crucial question is under which circumstances the underlying geometry of our model remains detectable by means of statistical testing, and when (i.e. for which values of d) our model converges in total variation distance to its non-geometric counterpart. A large body of work has already been devoted to this question for RGGs on the sphere [17, 10, 9, 33, 32] and recently also for random intersection graphs [9]. While the question when these models lose their geometry in the dense case is already largely answered, it remains open for the sparse case (where the marginal connection probability is proportional to $1/n$) and progress has only been made recently [9, 32]. It would be interesting to tightly characterize when our model loses its geometry both for the case of constant and for the case of inhomogeneous weights. Our bounds show that the number of triangles in our model for the sparse case (constant weights) is in expectation already the same as in a Erdős-Rényi graph if $d = \omega(\log^{3/2}(n))$, while on the sphere this only happens if $d = \omega(\log^3(n))$ [10]. Accordingly, we expect that our model loses its geometry earlier than the spherical model.

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 Pedro Almagro, Marián Boguñá, and M. Ángeles Serrano. Detecting the ultra low dimensionality of real networks. *Nature Communications*, 13(1):6096, 2022. doi:10.1038/s41467-022-33685-z.

3 Igor Artico, Igor E. Smolyarenko, Veronica Vinciotti, and Ernst C. Wit. How rare are power-law networks really? *Proceedings of the Royal Society A*, 476(2241):20190742, 2020. doi:10.1098/rspa.2019.0742.

4 Konstantin E. Avrachenkov and Andrei V. Bobu. Cliques in high-dimensional random geometric graphs. 5:1–24, Dec 2020. doi:10.1007/s41109-020-00335-6.

5 Thomas Bläsius and Philipp Fischbeck. On the External Validity of Average-Case Analyses of Graph Algorithms. In *30th Annual European Symposium on Algorithms (ESA 2022)*, pages 21:1–21:14, 2022. doi:10.4230/LIPIcs.ESA.2022.21.

6 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.

7 Thomas Bläsius, Tobias Friedrich, and Anton Krohmer. Cliques in hyperbolic random graphs. *Algorithmica*, 80(8):2324–2344, 2018. doi:10.1007/s00453-017-0323-3.

8 Anthony Bonato, David F. Gleich, Myunghwan Kim, Dieter Mitsche, Paweł Prałat, Yanhua Tian, and Stephen J. Young. Dimensionality of social networks using motifs and eigenvalues. *PLOS ONE*, 9(9):1–7, 2014. doi:10.1371/journal.pone.0106052.

9 Matthew Brennan, Guy Bresler, and Dheeraj Nagaraj. Phase transitions for detecting latent geometry in random graphs. (arXiv:1910.14167), Aug 2020. arXiv:1910.14167 [cs, math, stat]. URL: http://arxiv.org/abs/1910.14167.

10 Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Rácz. Testing for high-dimensional geometry in random graphs. (arXiv:1411.5713), Nov 2015. arXiv:1411.5713 [cs, math, stat]. URL: http://arxiv.org/abs/1411.5713.

11 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.

12 Fan Chung and Linyuan Lu. Connected Components in Random Graphs with Given Expected Degree Sequences. *Annals of Combinatorics*, 6(25):125–145, 2002. doi:10.1007/PL00012580.

13 Jesper Dall and Michael Christensen. Random geometric graphs. *Physical Review E*, 66:016121, 2002. doi:10.1103/PhysRevE.66.016121.

14 Fraser Daly, Alastair Haig, and Seva Shneer. Asymptotics for cliques in scale-free random graphs. (arXiv:2008.11557), Aug 2020. arXiv:2008.11557 [math]. URL: http://arxiv.org/abs/2008.11557.

15 Li Daqing, Kosmas Kosmidis, Armin Bunde, and Shlomo Havlin. Dimension of spatially embedded networks. *Nature Physics*, 7(6):481 – 484, 2011. doi:10.1038/nphys1932.

16 L. Decreusefond, E. Ferraz, H. Randriambololona, and A. Vergne. Simplicial homology of random configurations. *Advances in Applied Probability*, 46(2):325–347, Jun 2014. doi:10.1239/aap/1401369697.

17 Luc Devroye, András György, Gábor Lugosi, and Frederic Udina. High-dimensional random geometric graphs and their clique number. *Electronic Journal of Probability*, 16(none):2481–2508, Jan 2011. doi:10.1214/EJP.v16-967.

18 Vittorio Erba, Sebastiano Ariosto, Marco Gherardi, and Pietro Rotondo. Random geometric graphs in high dimension. *Physical Review E*, 102(1):012306, 2020. doi:10.1103/PhysRevE.102.012306.

19 P. Erdös and A. Rényi. On random graphs i. *Publ. Math. Debrecen*, 6, 1959.

20 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. doi:10.1214/17-AAP1314.

21 Nikolaos Fountoulakis, Pim van der Hoorn, Tobias Müller, and Markus Schepers. Clustering in a Hyperbolic Model of Complex Networks. *Electronic Journal of Probability*, 26(none):1 – 132, 2021. doi:10.1214/21-EJP583.

22 Tobias Friedrich, Andreas Göbel, Maximilian Katzmann, and Leon Schiller. Cliques in high-dimensional geometric inhomogeneous random graphs. *CoRR*, abs/2302.04113, 2023. doi:10.48550/arXiv.2302.04113.

23 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.

24 E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30:1141–1144, 1959.

25 Weiwei Gu, Aditya Tandon, Yong-Yeol Ahn, and Filippo Radicchi. Principled approach to the selection of the embedding dimension of networks. *Nature Communications*, 12(1):3772, 2021. doi:10.1038/s41467-021-23795-5.

26 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.

27 Svante Janson, Tomasz Łuczak, and Ilkka Norros. Large cliques in a power-law random graph. *Journal of Applied Probability*, 47(4):1124–1135, Dec 2010. doi:10.1239/jap/1294170524.

28 Ralph Keusch. *Geometric Inhomogeneous Random Graphs and Graph Coloring Games*. Doctoral thesis, ETH Zurich, 2018. URL: https://www.research-collection.ethz.ch/handle/20.500.11850/269658, doi:10.3929/ethz-b-000269658.

29 Christoph Koch and Johannes Lengler. Bootstrap Percolation on Geometric Inhomogeneous Random Graphs. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, pages 147:1–147:15, 2016. doi:10.4230/LIPIcs.ICALP.2016.147.

30 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.

31 Elizaveta Levina and Peter J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, page 777–784, 2004.

32 Siqi Liu, Sidhanth Mohanty, Tselil Schramm, and Elizabeth Yang. Testing thresholds for high-dimensional sparse random geometric graphs. (arXiv:2111.11316), Nov 2021. arXiv:2111.11316 [cs, math]. URL: http://arxiv.org/abs/2111.11316.

33 Suqi Liu and Miklos Z. Racz. Phase transition in noisy high-dimensional random geometric graphs. (arXiv:2103.15249), Mar 2021. arXiv:2103.15249 [cs, math, stat]. URL: http://arxiv.org/abs/2103.15249.

34 Riccardo Michielan and Clara Stegehuis. Cliques in geometric inhomogeneous random graphs. *Journal of Complex Networks*, 10(1), 2022. doi:10.1093/comnet/cnac002.

35 Mark Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64:025102, 2001. doi:10.1103/PhysRevE.64.025102.

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

37 Martin Raič. A multivariate berry–esseen theorem with explicit constants. *Bernoulli*, 25(4A), Nov 2019. arXiv:1802.06475 [math]. URL: http://arxiv.org/abs/1802.06475, doi:10.3150/18-BEJ1072.

38 Matteo Serafino, Giulio Cimini, Amos Maritan, Andrea Rinaldo, Samir Suweis, Jayanth R. Banavar, and Guido Caldarelli. True scale-free networks hidden by finite size effects. *Proceedings of the National Academy of Sciences*, 118(2), 2021. doi:10.1073/pnas.2013825118.

39 Ivan Voitalov, Pim van der Hoorn, Remco van der Hofstad, and Dmitri Krioukov. Scale-free networks well done. *Physical Review Research*, 1:033034, 2019. doi:10.1103/PhysRevResearch.1.033034.

40 Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998. doi:10.1038/30918.

41 Weihua Yang and David Rideout. High dimensional hyperbolic geometry of complex networks. *Mathematics*, 8(11), 2020. doi:10.3390/math8111861.