Real-World Networks are Low-Dimensional: Theoretical and Practical Assessment

Tobias Friedrich   TOBIAS.FRIEDRICH@HPI.DE
Andreas Göbel   ANDREAS.GOEBEL@HPI.DE
Leon Schiller   LEON.SCHILLER@STUDENT.HPI.DE
Hasso Plattner Institute, University of Potsdam

Maximilian Katzmann   MAXIMILIAN.KATZMANN@KIT.EDU
Karlsruhe Institute of Technology

arXiv:2302.06357v3 [cs.SI] 14 Aug 2024

Abstract

Detecting the dimensionality of graphs is a central topic in machine learning. While the problem has been tackled empirically as well as theoretically, existing methods have several drawbacks. On the one hand, empirical tools are computationally heavy and lack theoretical foundation. On the other hand, theoretical approaches do not apply to graphs with heterogeneous degree distributions, which is often the case for complex real-world networks.

To address these drawbacks, we consider geometric inhomogeneous random graphs (GIRGs) as a random graph model, which captures a variety of properties observed in practice. These include a heterogeneous degree distribution and non-vanishing clustering coefficient, which is the probability that two random neighbours of a vertex are adjacent. In GIRGs, n vertices are distributed on a d-dimensional torus and weights are assigned to the vertices according to a power-law distribution. Two vertices are then connected with a probability that depends on their distance and their weights.

Our first result shows that the clustering coefficient of GIRGs scales inverse exponentially with respect to the number of dimensions, when the latter is at most logarithmic in n. This gives a first theoretical explanation for the low dimensionality of real-world networks observed by Almagro et al. (2022). A key element of our proof is to show that when $d = o(\log n)$ the clustering coefficient concentrates around its expectation and that it is dominated by the clustering coefficient of the low-degree vertices.

We further use these insights to derive a linear-time algorithm for determining the dimensionality of a given GIRG. We prove that our algorithm returns the correct number of dimensions with high probability when the input is a GIRG. As a result, our algorithm bridges the gap between theory and practice, as it not only comes with a rigorous proof of correctness but also yields results comparable to that of prior empirical approaches, as indicated by our experiments on real-world instances.

Keywords: dimensionality testing, geometric inhomogeneous random graphs, clustering coefficient

1. Introduction

A key technique for understanding and analysing large complex data sets is to embed them into a low-dimensional geometric space. Hence, the search for embedding and dimensionality reduction algorithms has become an important direction in data analysis and machine learning research (Belkin and Niyogi, 2001; Sarveniazi, 2014; Camastra and Staiano, 2016; Nickel and Kiela, 2017). Embedding algorithms commonly require a metric that captures the similarities between data points, which is often abstracted using a graph whose vertices represent the data points and

two vertices are connected if they are close with respect to this metric. The algorithm then determines geometric positions for these vertices such that connected vertices are close together. Such approaches often require an a priori knowledge of the dimensionality, which is unknown in most applications. Heuristic approaches try to determine the dimensionality of a dataset by embedding it in spaces of different dimensionality and choosing the value that yields the optimal embedding (Levina and Bickel, 2004; Yin and Shen, 2018; Gu et al., 2021).

The recent work of Almagro et al. (2022) gives a new algorithm for learning the dimensionality that does not require embeddings. Instead, given a graph as input, their method counts the number of short – i.e. length 3, 4 and 5 – cycles of a graph. It then generates a search space consisting of random graphs that are generated from a geometric model of varying parameters, including the dimensionality of the space. Finally, a data-driven classifier finds the random graph of the search space that resembles the input graph the most and returns its dimensionality. A remarkable observation, that comes from using their algorithm to learn the dimensionality of real-world networks, is that the vast majority of networks has very low dimensionality, which is independent the size of the network.

A downside of the aforementioned approaches is that they rely on machine learning techniques that are computationally heavy and lack theoretical explanation. In order to argue with mathematical rigour, one requires to work with well-defined mathematical objects. A common approach to incorporate such an object is that of average-case analysis, that is, assume that the input graph comes from a well-defined random graph model. The random graph model that has been mostly considered so far in the literature is that of spherical random graphs, where vertices are generated independently and uniformly at random as points on the surface of a d-dimensional sphere and two vertices are connected if their angle is bellow a certain threshold. It can be easily shown that, as the number of dimensions increases, spherical random graphs converge to Erdős–Rényi graphs, the classical random graph model where edges are drawn independently. A series of works considers the statistical testing problem of detecting weather a given graph is a spherical random graph or an Erdős–Rényi graph and determines the parameter regime under which this can be done (Devroye et al., 2011; Bubeck et al., 2016; Brennan et al., 2020; Liu and Racz, 2021; Liu et al., 2022). Follow up works consider noisy settings (Liu and Rácz, 2021) or anisotropic geometric random graphs (Eldan and Mikulincer, 2020; Brennan et al., 2022), where each dimension has a different influence on the drawing of edges. The techniques of the aforementioned results can also be used for determining the dimensionality of the given graph (Bubeck et al., 2016, Theorem 5).

A characteristic of the random graph models considered in the aforementioned works, i.e. spherical random graphs and Erdős–Rényi graphs, is that the degree distributions of the generated graphs is concentrated around its expected value; this contrasts the power-law degree distributions observed in real-world networks (Faloutsos et al., 1999). While a latent geometric space appears to be a fundamental requirement for a random graph model that captures the high clustering coefficient (Krioukov, 2016; Boguna et al., 2021) and small diameter (Friedrich et al., 2013) observed in real-world networks, one needs to also consider the heterogeneity observed in the vertex degrees. A popular model in network theory capturing all previous properties is based on generating points on the hyperbolic plane instead of Euclidean (Boguna et al., 2010). However, it is not clear what the non-geometric counterpart to test against is in this case.

In this article we bring theory and practice closer together and provide a rigorous explanation for the very low dimensionality of real-world networks that has been observed in practice. Our proofs give new insights with which we are able to design linear-time algorithms for learning the

dimensionality of a network and show that they give the correct answer with high probability. To achieve our goal we consider the following random graph models.

Geometric inhomogeneous random graphs & Chung–Lu graphs. Geometric inhomogeneous random graphs (GIRGs), introduced by Bringmann et al. (2017) and are defined as follows.

Let $G(n, d, \beta, w_0) = (V, E)$ denote the n-vertex graph obtained in the following way. For each $v \in V$, we sample a weight w_v from the Pareto distribution $\mathcal{P}$ with parameters $w_0, 1 - \beta$ such that the CDF and density is

$$\Pr\left[w_v \leq x\right] = 1 - (x/w_0)^{1-\beta} \text{ and } \rho_{w_v}(x) = \frac{\beta - 1}{w_0^{1-\beta}} x^{-\beta},$$

respectively. We denote the sequence of the drawn weights by $\{w\}_1^n$ and assume that $\beta > 2$ such that a single weight has finite expectation (and thus the average degree in the graph is constant), but possibly infinite variance. Moreover, each vertex v is assigned a position $\mathbf{x}_v$ in the d-dimensional torus $\mathbb{T}^d$ uniformly at random according to the standard Lebesgue measure. We denote the i-th component of $\mathbf{x}_v$ by $\mathbf{x}_v(i)$. Two vertices u, v are adjacent if and only if their distance $d(\mathbf{x}_u, \mathbf{x}_v)$ is at most the *connection threshold* t_{uv}, which is defined such that the marginal connection probability of u, v is

$$\Pr\left[u \sim v\right] := \min\left\{1, \frac{\lambda w_u w_v}{n}\right\} = \frac{\kappa_{uv}}{n}, \text{ where } \kappa_{uv} := \min\left\{n, \lambda w_u w_v\right\} \tag{1}$$

and where $\lambda \in \mathbb{R}$ is a parameter that controls the average degree. We measure of the distance between two points using the L_p-norm with $1 \leq p \leq \infty$. That is, we define

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

where $|x - y|_C$ denotes the distance on the circle, i.e., $|x - y|_C = \min\{|x - y|, 1 - |x - y|\}$.

Note that L_∞ is a natural metric on the torus as $B_\infty(r)$, the ball of radius r under this norm is a (hyper-)cube and "fits" entirely into $\mathbb{T}^d$ for all $0 \leq r \leq 1$. For this reason, the connection threshold under L_∞-norm is always

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

where $\mu = 2^d/\lambda$.

The GIRG model has a natural non-geometric counterpart where the weight distribution of the vertices is the same as in GIRGs but the edges are now sampled independently, with probability

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

This inhomogeneous random graph model is known as the Chung–Lu random graph model and has been extensively studied in literature (see, e.g., Aiello et al., 2001; Chung and Lu, 2002a,b). It is important to note that – despite the fact that the connection probability of any two vertices in GIRGs

and Chung–Lu graphs is the same – these two models have important differences because edges in GIRGs do not appear independently since they further depend on the positions of the involved vertices.

Hence, for our analysis, we are now equipped with an appropriate geometric random graph model and its non-geometric counterpart. Note that, as it was shown by Friedrich et al. (2023, Theorem 1), the two models converge as the number of dimensions in the GIRG model goes to infinity, i.e. the total variation distance of the two models goes to zero. Furthermore, we observe that the GIRG model captures many quantifiable properties of real-world networks as shown by Bläsius and Fischbeck (2022). Finally, let us note that the model is very versatile as one can consider other variants with different degree distributions or metric spaces. However, our choice of the Pareto distribution for the vertex weights and of the torus for the geometric space is the one considered most frequently in literature – also in the results of Bläsius and Fischbeck (2022).

Triangles and Clustering Coefficient. The number of triangles and related properties of a graph are common statistics used in the analysis of networks (Gupta et al., 2014), especially for detecting underlying geometry. In fact a related statistic[1] was used by Bubeck et al. (2016) to efficiently test for the presence of geometry for a spherical random geometric graph.

When dealing with heterogeneous degree distributions, however, triangles that are attributed to large degree vertices potentially have a significant influence on the total number of triangles. In fact as shown by Friedrich et al. (2023), the number of triangles in GIRGs and in Chung–Lu graphs are asymptotically equivalent if $\beta \leq 7/3$, which is not an unrealistic choice for many real-world networks. Michielan et al. (2022), therefore suggest weighting each triangle by the inverse degrees of the involved vertices, a statistic they call *weighted triangles*. A normalized version of the number of triangles as well as cordless squares and pentagons was also used by Almagro et al. (2022) to determine the dimensionality of a given network.

A natural statistic, observed in many real-world networks that is however strikingly absent in non-geometric random graphs where edges are drawn independently, is the existence of a non-vanishing clustering coefficient, i.e. the probability that two randomly chosen neighbours of a vertex are adjacent. The clustering coefficient is the central focus of our analysis and we use the following common definition, also used by Keusch (2018, Definition 5.1).

Given a graph $G = (V, E)$, its *local clustering coefficient* of a vertex v is

$$\mathrm{CC}_G(v) := \begin{cases} |\{\{s,t\} \subseteq \Gamma(v) \mid s \sim t\}|/\binom{\deg(v)}{2} & \text{if } \deg(v) \geq 2 \\ 0 & \text{otherwise.} \end{cases}$$

The (global) *clustering coefficient* of G is the average of the local coefficient of each of G's vertices, that is,

$$\mathrm{CC}(G) := \frac{1}{|V|} \sum_{v \in V} \mathrm{CC}_G(v).$$

For GIRGs it was shown by Keusch (2018) that, when the vertices of the generated graph are drawn on a torus of constant dimensionality, the generated graph has a constant clustering coefficient. On the other hand, on Chung–Lu graphs it was shown that the clustering coefficient goes to 0 as n, the number of vertices of the graph, grows (van der Hofstad et al., 2017). Our first result,

---

1. The statistic used, number of signed triangles, essentially measures by how much the number of triangles are in the graph exceeds the expected value in the Erdős–Rényi graph model.

which we discuss in the next section, extends the results on the clustering coefficient of GIRGs by giving an upper bound that explicitly depends on the dimension of the underlying space. This shows that constant dimensionality is in fact needed to obtain a constant clustering coefficient. We refine our result for the case of L_∞-norm in Section 1.1 and afterwards introduce a statistical test for learning the dimension of a network based on these results (Section 1.2).

1.1. Upper bounds on the clustering coefficient of GIRGs

Our first result connects the clustering coefficient of a GIRG with the number of dimensions used to generate the positions of its vertices.

Theorem 1 *Asymptotically almost surely, if $d = o(\log(n))$, the clustering coefficient of G sampled from the GIRG model under some L_p-norm with $p \in [1, \infty]$ is*

$$\mathrm{CC}(G) = \exp(-\Omega_d(d)) + o(1).$$

We remark that, for the case of L_∞-norm, we later derive a sharper bound (see Theorem 3). Theorem 1 implies that if $d = \omega(1)$ and $d = o(\log(n))$ the clustering coefficient vanishes. As most real-world networks have a non-vanishing clustering coefficient, our theorem suggests that their dimensionality must be at most constant in the number of vertices. This can be seen as a theoretical explanation for the empirical observations of the low dimensionality of real-world networks by Almagro et al. (2022).

Besides the results of Keusch (2018) for a constant number of dimensions, the clustering coefficient of random geometric graphs (i.e., our model in the case of homogeneous weights) under the L_2-norm as a function of d was previously analysed by Dall and Christensen (2002). However, our Theorem 1 also applies to inhomogeneous degree distributions and arbitrary L_p-norms, which complicates the analysis. The main difficulty in proving Theorem 1 is that the probability that two random neighbours of a given vertex are connected is significantly influenced by their weights. To circumvent this issue we show that high-weight vertices only have a small influence on the global clustering coefficient of a power-law graph G in Section 3.1. Via an application of the method of typical bounded differences (Theorem 8 see also the article by Warnke (2016)) – a generalisation of McDiarmid's inequality McDiarmid et al. (1989) and a powerful tool to showing concentration in high dimensional spaces – we then show that the clustering coefficient of a GIRG concentrates around the expected clustering coefficient of a subgraph induced by vertices of small weight.

The bound on the clustering coefficient of the low-weight vertex subgraph follows from a bound on the probability that two random vectors $\mathbf{y}_u, \mathbf{y}_v$ uniformly distributed within the ball of radius 1 have a distance larger than a certain threshold. Intuitively, the fact that this probability decays exponentially in d is a consequence of the law of large numbers: as d grows, with large probability, about half of the components of $\mathbf{y}_u$ and $\mathbf{y}_v$ have opposite sign, which already leads to a distance between u and v that is arbitrarily close to 1 with probability converging to 1 as d grows. Taking into account that the other components of $\mathbf{y}_u$ and $\mathbf{y}_v$ also contribute at least a constant increase in distance between u and v with large probability, we get that there is an exponentially increasing probability that the distance between u and v is strictly greater than one, which suffices to show an exponential upper bound on the clustering coefficient in G. To prove this exponential decay in terms of d, we use a coupling argument based on the observation that the "direction" $\mathbf{x}/\|\mathbf{x}\|_p$ and the norm $\|\mathbf{x}\|_p$ of a random vector distributed in the unit ball under L_p-norm are independent. To analyze the normalized vector $\mathbf{x}/\|\mathbf{x}\|_p$, we define the following distribution and show that if $\mathbf{z}$ is

a vector sampled from this distribution, then $\mathbf{z}/\|\mathbf{z}\|_p$ is distributed just as $\mathbf{x}/\|\mathbf{x}\|_p$. This has the advantage that the components of $\mathbf{z}$ are now independent, allowing us to apply sharp tail bounds from which our statements follow.

The χ^p-Distribution. Let $p \in \mathbb{R}, p \geq 1$. We call a random vector $\mathbf{x} \in \mathbb{R}^d$, $\chi_p(d)$ distributed if each of its components $\mathbf{x}(i)$ is independently distributed according to the density function

$$\rho(x_i) := \gamma e^{-\frac{1}{2}|\mathbf{x}(i)|^p}$$

with the normalising constant

$$\gamma = \frac{p}{2^{1/p+1}\Gamma(1/p)},$$

where $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}\mathrm{d}x$ is the gamma function. If $\mathbf{x} \sim \chi_p(d)$, then we denote the distribution of the random variable $\left(\|\mathbf{x}\|_p\right)^p = \sum_{i=1}^d |\mathbf{x}(i)|^p$ by $\chi^p(d)$.

This distribution is a generalisation of the χ^2 distribution and a simplification of the one proposed by Livadiotis (2014). In our analysis, we determine its moment generating function. This not only gives us its expectation, which is $2d/p$, but also allows us to obtain the following concentration bound, which we use in the arguments used for the proof of Theorem 1.

Corollary 2 *Let $X_i, \ldots X_d$ be i.i.d. random variables from $\chi_p(1)$ and define $Z = \sum_{i=1}^d |X_i|^p \sim \chi^p(d)$. Then, for every $\varepsilon > 0$,*

$$\Pr\left[|Z - \mathbb{E}[Z]| \geq \varepsilon \cdot \mathbb{E}[Z]\right] \leq 2\exp\left(-\frac{2\delta}{p}d\right).$$

Where $\delta > 0$ is defined by $\varepsilon = 2(\sqrt{2\delta} + \delta)$.

We believe our analysis of the $\chi_p(d)$ and $\chi^p(d)$ distributions to be of independent interest, as many random spaces can be related to vectors drawn uniformly at random within the d-dimensional unit ball of some L_p-norm.

Improved bounds for the L_∞-norm. When using L_∞-norm as a distance measure for GIRGs we obtain more precise results and are able to further determine the base of the exponential function governing the decay of $\mathrm{CC}(G)$. Recall that the L_∞ norm is not only a natural distance measure on the torus from a mathematical point of view, but also one that yields graphs that closely resemble real-world networks (Bläsius and Fischbeck, 2022).

Theorem 3 *Assume that $d = o(\log(n))$ and $\beta \neq 3$. Then asymptotically almost surely, the clustering coefficient of G sampled from the GIRG model with L_∞-norm fulfils*

$$\mathrm{CC}(G) = \mathcal{O}_d\left(\left(\frac{3}{4}\right)^{\min\{1,\beta-2\}d}\right) + o(1)$$

and

$$\mathrm{CC}(G) = \Omega_d\left(\max\left\{\left(\frac{1}{4}\right)^{(\beta-2)d}, \left(\frac{3}{4}\right)^d\right\}\right).$$

In particular, if $\beta > 3$, we have

$$\text{CC}(G) = \Theta_d\left(\left(\frac{3}{4}\right)^d\right) + o(1).$$

This theorem shows that $\text{CC}(G)$ essentially decays as $(3/4)^d$ asymptotically in d if $\beta > 3$. Otherwise, if $\beta \in (2,3)$, we obtain slightly weaker bounds but we are in particular able to show that $\text{CC}(G)$ decays asymptotically slower than $(3/4)^d$ if β is sufficiently close to 2, which follows from the lower bound $\text{CC}(G) = \Omega_d\left((1/4)^{(\beta-2)d}\right)$. The reason for this is that the expected weight of a random neighbor of a given vertex is infinite if $\beta \in (2,3)$ which leads to an increased overall clustering coefficient.

The proof of Theorem 3 is based on an application of the following theorem by Friedrich et al. (2023, Theorem 3) that bounds the probability that a set of k vertices forms a clique conditioned on the event that said vertices form a star centered at the vertex of minimal weight under the assumption that the ratio between the minimal and maximal weight is bounded. We slightly reformulate the original statement for the sake of exposition.

Theorem 4 *Let G be a GIRG generated under L_∞-norm. Let $U = \{v, s, t\}$ be a set of 3 vertices with weights w_v, w_s, w_t such that $w_v \leq w_s \leq w_t$ and $w_v \leq cw_t$ for some constant $c > 0$. If $\left(w_t^2/(\mu n)\right)^{1/d} \leq 1/4$, we have*

$$\left(\frac{3}{4}\right)^d \leq \Pr\left[U \text{ is a triangle} \mid v \sim s, t\right] \leq c\left(\frac{3}{4}\right)^d.$$

We remark that the condition $\left(w_t^2/(\mu n)\right)^{1/d} \leq 1/4$ is needed to ensure that the connection threshold for s, t is sufficiently small such that we can ignore the topology of the underlying torus and measure distances as in $\mathbb{R}^d$.

1.2. Testing for the dimensionality

A natural further question that arises is whether one can recover the underlying dimension of a given GIRG by means of statistical testing. The previous secons suggest that the clustering coefficient is an indicator of this property, however, we have also seen that this metric is further influenced by other model parameters (i.e. w_0, λ and especially β) making it rather unsiutable for designing a rigorous test. A similar problem arises when using the total number of triangles, which is is dominated by those forming among large degree vertices independently of d if β is close to 2 as observed by Friedrich et al. (2023); Michielan et al. (2022). In Michielan et al. (2022), the authors therefore suggest to count the number of weighted triangles instead, where each triangle contributes a weight that is inversely proportional to the product of the degrees of its vertices. Weighted triangles thus counteract the effect of large degree vertices as the influence of triangles forming among such vertices is diminished. However, this approach only allows to decide whether the network has an underlying metric structure, but not its dimensionality.

We take a similar (yet more direct) approach for excluding the effect of large degree vertices and introduce a test that is further able to infer the dimension of the underlying metric space. Namely, we can show that the clustering coefficient among vertices of approximately the same weight that have at least two neighbors is highly concentrated and a direct indicator of the underlying dimension,

without being influenced by other model parameters. More precisely, using Theorem 4 together with the method of typical bounded differences (Theorem 8), we can show that the average local clustering coefficient in the induced subgraph of all vertives with weight in some interval $[w_c, cw_c]$ (where $w_c \geq w_0, 0 < c < 2/\sqrt{3}$) concentrates tightly around a value that is only dependent on d and not on β or w_0. This is formalised in the following theorem.

Theorem 5 *Let $G = G(n, d, \beta, w_0)$ be a GIRG generated under L_∞-norm. Let further $1 < c < 2/\sqrt{3} \approx 1.1547, w_c \geq w_0$ be constants, and let $\tilde{G}$ be the subgraph of G consisting of all vertices with weight in $[w_c, cw_c]$. Assume that d is an integer with $d = o(\log(n))$. Define the set S as the set of vertices in $\tilde{G}$ that have at least two neighbors in $\tilde{G}$ and the random variable $\mathrm{CC}^{(+)}(\tilde{G})$ as*

$$\mathrm{CC}^{(+)}(\tilde{G}) := \frac{1}{|S|} \sum_{v \in S} \frac{|\{\{s,t\} \in \Gamma(v) \mid s \sim t\}|}{\binom{\deg(v)}{2}} = \frac{1}{|S|} \sum_{v \in S} \mathrm{CC}_{\tilde{G}}(v).$$

Then,

$$\mathrm{CC}^{(+)}(\tilde{G}) \in \left(\frac{1}{c} \left(\frac{3}{4} \right)^d, c \left(\frac{3}{4} \right)^d \right) \pm n^{-1/5} \tag{2}$$

with probability at least $1 - 1/n$.

Theorem 5 can be viewed as a linear-time algorithm for the following statistical testing problem (assuming that w_c is constant). We are given a graph G on n vertices, its weight sequence, and an integer $d = o(\log(n))$. Under the null hypothesis, G is a GIRG generated with dimension d, whereas under the alternative hypothesis, G was generated in dimension $d_1 \neq d$ or it is a Chung–Lu graph. Here, we allow d_1 to be any integer (potentially larger than $\log(n)$). Consider the following testing procedure for this problem. Fix a constant $1 < c < 2/\sqrt{3}$ and a weight $w_c \geq w_0$. Now, consider the induced subgraph $\tilde{G}$ of G consisting of all nodes with weight in $[w_c, cw_c]$. For every node $v \in \tilde{G}$ that has at least two neighbours in $\tilde{G}$, compute its local clustering coefficient $\mathrm{CC}_{\tilde{G}}(v)$ and denote by $\mathrm{CC}^{(+)}(\tilde{G})$ the mean over all these values. We accept the null hypothesis if and only if condition (2) is met. Due to Theorem 5 the probability that this test is incorrect under both the null and alternative hypothesis goes to zero as $n \to \infty$. Furthermore, the running time of this test is linear, as we have to compute the local clustering coefficient of vertices of constant weight and as the degree of a vertex with weight cw_c is constant in expectation. Iterating this statistical test over the range of d we can recover the dimensionality of the input graph with high probability. Let us note that our result is not restricted to a constant number of dimensions but applies to the whole regime $d = o(\log(n))$, which – as Theorem 3 implies – is the only relevant one for this problem.

1.3. Application to real-world networks.

In addition to our theoretical results, we tested our algorithm in practice, both in real world networks and in GIRGs. For estimating the vertex weights, we used the maximum likelihood estimator derived in (Boguna et al., 2010, Appendix B.2). The outcome of our experiments is summarised in Figure 1. Table 1 further contains a list of the real-world networks we used for the first row of plots in Figure 1. In Figure 1, the size of the circles is proportional to the number of vertices in the induced subgraph of vertices with weight in the interval $[w_c, cw_c]$. We use $c = 1.155$ which is

roughly the maximum permitted value predicted by Theorem 5. The dashed lines represent roughly the expected value of our test statistic for a GIRG genenerate in dimension d, i.e. $(3/4)^d$. The GIRGs were generated with the sampling algorithm of Bläsius et al. (2022) using $\beta = 2.5, \alpha = 10$ and an average degree of 10. The histogram on the lower right of Figure 1 shows the frequency of each inferred dimension using the dataset of Bläsius and Fischbeck (2022) consisting of 2976 real-world networks. The dimension here is inferred by taking the weighted median of the inferred dimension from our test statistic over different values of w_c ranging from 2 to 300. The weighting is by the number of vertices in the respective subgraph induced by vertices with weight in $[w_c, cw_c]$.

The inferred dimensions we obtain are indeed similar to the results of Almagro et al. (2022, Fig. 5) with typical inferred dimensions being in the range of 1 to 10 and with social networks being generally assigned to higher dimensions than collaboration or citation networks. Some of the networks we use (all the ones we were able to find) are also contained in the dataset of Almagro et al. (2022) and here, the inferred dimensions of both approaches are very similar. For examples, consider `email-Enron`, `ca-AstroPh`, `ca-CondMat`, `ca-GrQc`, `cit-HepTh`.

We remark that the big advantage of our algorithm is that it has stronger theoretically foundations and is much more efficient. In fact, we are able to handle datasets of orders of magnitude larger than Almagro et al. (2022). In fact, our experiments conducted on the set of 2976 real-world networks by Bläsius and Fischbeck (2022) which are summarized in the histogram in Figure 1 show that a vast majority of them is assigned a dimension of at most 10. This can be seen as a further indication of the ultra-low-dimensional nature of most realistic networks, now tested on a much larger dataset as before.

Besides that, we note that the algorithm works exceptionally well for synthetic networks and this holds even if we use a soft version of GIRGs which includes an additional temperature parameter $\alpha \geq 1$ that (if close to 1) diminishes the influence of the underlying geometry (we refer to Bläsius et al. (2022) for more information). Experiments indicate that the test continues to achieve a performance similar to that predicted by theory for all $\alpha \geq 2$. However, even for smaller values of α, the inferred dimension of our test can still be seen as an upper bound on the ground truth since it is known that smaller values of α only lead to a decrease in clustering.

We further observe that some of the considered real-world networks show an overall similar behaviour as that of the GIRGs (e.g. `soc-academia`, `fb-pages-artist`, `ca-AstroPh`, `ca-CondMat`). It is however not a surprise that real-world data can be noisy and, therefore, exhibit a behaviour that differs from GIRGs. Similar difficulties were also encountered by Almagro et al. (2022) (see their supplementary material) and a similar noisy behaviour can also be observed in small generated GIRGs, where number of vertices is not hight enough for the concentration results to be strong. Nevertheless, some of the considered networks (especially the biological networks) show a rather different behavior as predicted by the GIRG model which indicates that GIRGs do not capture all properties of realistic networks and thus motivates further research.
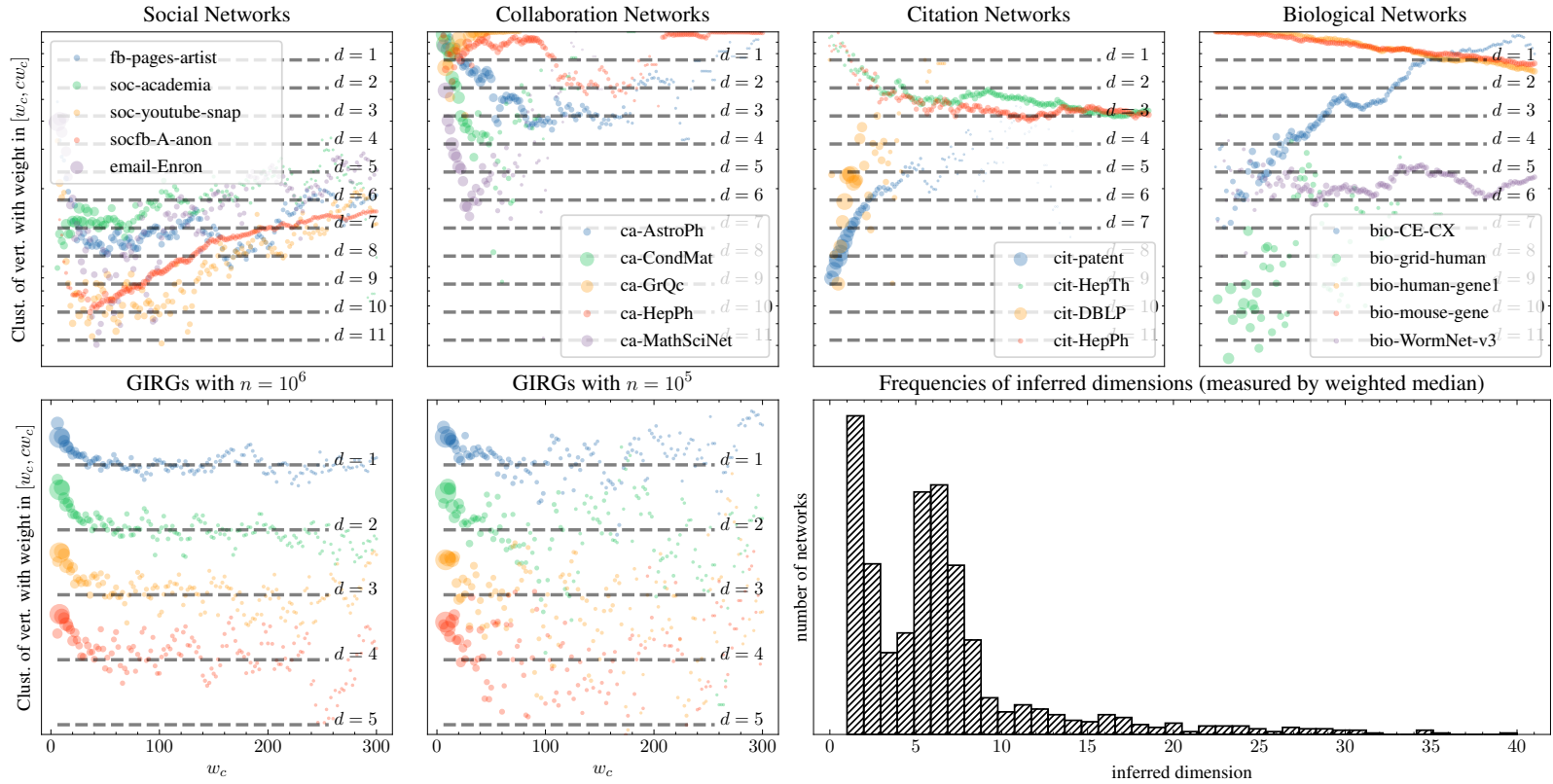
Figure 1: The clustering coefficient of the low-weight vertices for different choices of w_c in real-world and synthetic networks. The size of the circles is proportional to the number of vertices in the induced subgraph of vertices with weight in the interval $[w_c, cw_c]$ whereby we use $c = 1.155$. The dashed lines represent roughly the expected value of our test statistic for a GIRG. The histogram shows the frequency of each inferred dimension using the dataset of Bläsius and Fischbeck (2022) consisting of 2976 real-world networks.

1.4. Future work

As we previously discussed, a large body of work has been devoted to understanding in which cases (i.e. for which asymptotic behaviors of d), geometry is detectable in spherical random geometric graphs (SRGGs) for homogeneous weights. While the parameter regime where these graphs lose their geometry in the dense case, i.e. the case where the marginal connection probability of two vertices is constant and does not depend on n, is well understood (Devroye et al., 2011; Bubeck et al., 2016; Liu and Racz, 2021), it remains unclear what happens in the sparse case (where the marginal connection probability is proportional to $1/n$) and progress has been made only recently (Brennan et al., 2020; Liu et al., 2022).

On the other hand, there has not been much research devoted to studying the influence of the dimension on random geometric graphs in the case of inhomogeneous weights. We gave first results in this regard by studying how the clustering coefficient depends on d and showed that the dimension can be detected by means of statistical testing assuming that $d = o(\log(n))$. It remains to study under which conditions the geometry remains detectable if $d = \Omega(\log(n))$ and under which circumstances the model converges to its non-geometric counterpart with respect to the total variation distance of the distributions over the produced graphs as previously studied for spherical random graphs. Furthermore, it remains to study what differences arise when using the torus instead of the sphere as the underlying metric space. We expect that our model loses its geometry earlier than spherical random graphs, as the number of triangles in our model for the sparse case with constant weights is, in expectation, the same as in an Erdős-Rényi graph already if $d = \omega(\log^{3/2}(n))$ (Friedrich et al., 2023). On the sphere this only happens if $d = \omega(\log^3(n))$ (Bubeck et al., 2016).

For detecting the geometry in SRGGs Bubeck et al. (2016) have introduced the *signed triangle* statistic which proves more powerful than ordinary triangles in the dense case. It remains to study if signed triangles, or a combination of signed triangles and weighted triangles considered by Michielan et al. (2022), gives rise to a more powerful test for the case of inhomogeneous weights.

A further interesting extension is to consider an anisotropic version of our model, along the lines of the work of Eldan and Mikulincer (2020); Brennan et al. (2022). In the GIRG model, this can be naturally encoded in the distance measure used to determine the edge threshold.

Noisy settings have also been considered in the context of testing for geometry in random graphs (Liu and Rácz, 2021). Our model comes with a natural way of modelling noise in the form of and inverse temperature parameter $\alpha > 1$ (Keusch, 2018). Here, two vertices are connected with a probability that depends on both their distance and the temperature. More precisely, the connection probability of two vertices u, v fulfills

$$p_{uv} = \Theta\left(\frac{1}{\|\mathbf{x}_u - \mathbf{x}_v\|_\infty^{\alpha d}}\left(\frac{w_u w_v}{n}\right)^\alpha\right).$$

Intuitively, lower values of α diminish the influence of the underlying geometry. We expect results similar to ours to hold in the noisy setting if α is a constant. It remains to study how different (constant or superconstant) values of α influence the detectability of the underlying geometry and dimension.

2. Preliminaries

We let $G = (V, E)$ be a (random) graph on n vertices. For any value δ, let $V_{\leq\delta}$ denote the set of vertices of degree at most δ and let $G_{\leq\delta}$ denote the subgraph of G induced by $V_{\leq\delta}$. We use standard

Landau notation to describe the asymptotic behavior of functions for sufficiently large n. That is, for functions f, g, we write $f(n) = \mathcal{O}(g(n))$ if there is a constant $c > 0$ such that for all sufficiently large n, $f(n) \leq cg(n)$. Similarly, we write $f(n) = \Omega(g(n))$ if $f(n) \geq cg(n)$ for sufficiently large n. If both statements are true, we write $f(n) = \Theta(g(n))$. Regarding our study of the clustering coefficient, some results make a statement about the asymptotic behavior of a function with respect to a sufficiently large d. These are marked by $\mathcal{O}_d(\cdot), \Omega_d(\cdot), \Theta_d(\cdot)$, respectively.

2.1. Probabilistic tools

We say that an event $\mathbf{E}$ holds asymptotically almost surely if $\lim_{n \to \infty} \Pr[\mathbf{E}] = 1$ and with high probability if $\Pr[\mathbf{E}] \geq 1 - \mathcal{O}(1/n)$. The following theorem shows that the sum of independent Bernoulli random variables converges to a Poisson distributed random variable if the individual success probabilities are small.

Theorem 6 (Proposition 1 in Cam (1960)) *For $1 \leq i \leq n$, let X_i be independent Bernoulli distributed random variables such that $\Pr[X_i = 1] = p_i$. Let $\lambda_n = \sum_{i=1}^{n} p_i$, and $S = \sum_{i=1}^{n} X_i$. Then,*

$$\sum_{k=0}^{\infty} \left| \Pr[S = k] - \frac{\lambda_n^k e^{-\lambda_n}}{k!} \right| \leq 2 \sum_{i=1}^{k} p_i^2.$$

We will also use the following concentration bounds.

Theorem 7 (Theorem 2.2 in Keusch (2018), Chernoff-Hoeffding Bound) *For $1 \leq i \leq k$, let X_i be independent random variables taking values in $[0, 1]$, and let $X := \sum_{i=1}^{k} X_i$. Then, for all $0 < \varepsilon < 1$,*

(i) $\Pr[X > (1 + \varepsilon)\mathbb{E}[X]] \leq \exp\left(-\frac{\varepsilon^2}{3}\mathbb{E}[X]\right).$

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

(iii) $\Pr[X \geq t] \leq 2^{-t}$ *for all $t \geq 2e\mathbb{E}[X]$.*

While this theorem is extremely useful when dealing with sums of independent random variables, we shall further need the *method of typical bounded differences* to obtain bounds when the Chernoff-Hoeffding bound is not applicable.

Theorem 8 (Theorem 2.5 in Keusch (2018)) *Let $X_1, \ldots, X_m$ be independent random variables over $\Omega_1, \ldots, \Omega_m$. Let $X = (X_1, \ldots, X_m) \in \Omega = \prod_{i=1}^{m} \Omega_i$ and let $f : \Omega \to \mathbb{R}$ be a measurable function such that there is some $M > 0$ such that for all $\omega \in \Omega$, we have $0 \leq f(\omega) \leq M$. Let $\mathcal{B} \subseteq \Omega$ such that for some $c > 0$ and for all $\omega, \omega' \in \overline{\mathcal{B}}$ that differ in at most two components X_i, X_j, we have*

$$\left| f(\omega) - f(\omega') \right| \leq c.$$

Then, for all $t \geq 2M\Pr[\mathcal{B}]$, we have

$$\Pr[|f(X) - \mathbb{E}[f(X)]| \geq t] \leq 2 \exp\left(-\frac{t^2}{32mc^2}\right) + \left(\frac{2Mm}{c} + 1\right)\Pr[\mathcal{B}].$$

2.2. Basic properties of the GIRG model

We will need the following statements about the distribution of the degrees and weights in the GIRG model.

Lemma 9 (Lemma 3.3 and Lemma 3.4 in Keusch (2018), slightly reformulated) *The following properties hold for $G(n, d, \beta, w_0) = (V, E)$.*

(i) For all $v \in V$, we have $\mathbb{E}\left[\deg(v)\right] = \Theta(w_v)$.

(ii) With probability $1 - n^{-\omega(1)}$, we have for all $v \in V$ that $\deg(v) = \mathcal{O}(w_v + \log^2(n))$.

In fact, we need a slightly stronger version of statement (ii) above.

Lemma 10 *Let $V_{\leq \log(n)}$ be the set of all vertices with weight at most $\log(n)$. With probability at least $1 - n^{-\Omega(\log^2(n))}$, we have for all $v \in V_{\leq \log(n)}$ that $\deg(v) \leq \log^3(n)$.*

Proof We get from Lemma 9 that $\mathbb{E}\left[\deg(v)\right] = w_v$. Hence, for sufficiently large n, we get that $\log^3(n) \geq 2e\mathbb{E}\left[\deg(v)\right]$ for all $v \in V_{\leq \log(n)}$. Since the degree of a fixed vertex v conditioned on its position is a sum of independent Bernoulli distributed random variables, we may apply statement (iii) from Theorem 7 to obtain $\Pr\left[\deg(v) \geq \log^3(n)\right] \leq n^{-\Omega(\log^2(n))}$. From a union bound, we get that the probability that at least one vertex from $V_{\leq \log(n)}$ has a degree of $\log^3(n)$ or more is at most $n \cdot n^{-\Omega(\log^2(n))} = n^{-\Omega(\log^2(n))}$, which concludes the proof. ∎

3. Upper bound on the clustering coefficient of GIRGs

We proceed by pointing out our general bounding technique and then handle the case of L_∞-norm and L_p-norms with $p \in [1, \infty)$ separately.

3.1. Our bounding technique

We go on with developing a technique for upper bounding $\mathrm{CC}(G)$. The main difficulty here is that the probability that two random neighbors of a given vertex are connected grows significantly with their weight. We circumvent this issue by showing that high-weight vertices only have a small influence on the global clustering coefficient of a power-law graph G, which essentially concentrates around its expectation in an induced subgraph of small weight. We formalize this in the following lemma that is proved in a similar way as (Keusch, 2018, Theorem 4.4).

Lemma 11 *Asymptotically almost surely, we have*

$$\mathrm{CC}(G) = \mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] + o(1).$$

To prove this statement, we require the following auxiliary lemmas.

Lemma 12 (Lemma 3.5 in Keusch (2018)) *If the weight w of each vertex is sampled from the Pareto-distribution with parameters $w_0, 1 - \beta$, then for all $\eta > 0$, there is a constant $c > 0$ such that with probability $1 - n^{-\omega(\log \log(n)/\log(n))} = 1 - o(1)$, and all $w \geq w_0$, we have*

$$|V_{\geq w}| \leq cnw^{1+\eta-\beta}.$$

Lemma 13 (Lemma 3.4 in Keusch (2018)) *With probability $1 - n^{-\omega(1)}$, for all $v \in V$, we have $\deg(v) = \mathcal{O}(w_v + \log(n)^2)$.*

Lemma 14 (Lemma 3.1 in Keusch (2018)) *If for all $\eta > 0$, there is a constant $c > 0$ such that for all $w \geq w_0$, we have $|V_{\geq w}| \leq cnw^{1+\eta-\beta}$, then*

$$\sum_{v \in V_{\geq w}} w_v = \mathcal{O}(nw^{2+\eta-\beta}).$$

Proof [Proof of Lemma 11] We start by showing that

$$\mathrm{CC}(G_{\leq n^{1/8}}) = \mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] + o(1)$$

asymptotically almost surely and then how this statement transfers to the whole graph G.

To show concentration, we use Theorem 8 and note that the positions and weights of all vertices define a product probability space as in Theorem 8. We denote this space by Ω, whereby every $\omega \in \Omega$ defines a graph $G(\omega)$ on the vertex set $V_{\leq n^{1/8}}$. Note that the number of independent random variables is $m = 2n$. Thus, we may define a function $f : \Omega \to \mathbb{R}$ that maps every $\omega \in \Omega$ to $\mathrm{CC}(G(\omega))$. We consider the "bad" event

$$\mathcal{B} = \{\omega \in \Omega \mid \text{the maximum degree in } G(\omega) \text{ is at least } n^{1/4}\}.$$

By Lemma 13, we get that $\Pr[\mathcal{B}] = n^{-\omega(1)}$. Now, let $\omega, \omega' \in \overline{\mathcal{B}}$ such that they differ in at most two coordinates. We observe that changing the weight or coordinates of one vertex v only influences the clustering coefficient of v itself or vertices that are neighbors of v before or after the change. Since v has at most $n^{1/4}$ neighbors in both $G(\omega)$ and $G(\omega')$, the change affects at most $2n^{1/4}$ vertices. Two such changes can hence only increase or decrease the clustering coefficient of $G(\omega)$ by at most $4n^{1/4}/n$, and so we have $|f(\omega) - f(\omega')| \leq 4n^{-3/4}$. We note that the choice $t = n^{-1/8}$ fulfills the condition $t \geq 2M\Pr[\mathcal{B}]$ since $M = 1$ and $\Pr[\mathcal{B}] = n^{-\omega(1)}$. Thus, we may apply Theorem 8 to obtain

$$\Pr\left[\left|\mathrm{CC}(G_{\leq n^{1/8}}) - \mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right]\right| \geq n^{-1/8}\right]$$
$$\leq 2\exp\left(-\frac{n^{-1/4}}{32 \cdot 2n \cdot 16n^{-3/2}}\right) + \left(\frac{4n}{n^{-3/4}} + 1\right)n^{-\omega(1)} = n^{-\omega(1)}.$$

This shows that with high probability, $|\mathrm{CC}(G_{\leq n^{1/8}}) - \mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right]| = o(1)$.

In order to transfer this finding to the entire graph G, we note that each additional vertex we add to $G_{\leq n^{1/8}}$ has (local) clustering of at most one and each edge, we add to a vertex $v \in V_{\leq n^{1/8}}$ can only increase its clustering by at most one as well. Hence,

$$\mathrm{CC}(G) \leq \frac{1}{n}\left(|V_{\leq n^{1/8}}|\mathrm{CC}(G_{\leq n^{1/8}}) + |V_{>n^{1/8}}| + \sum_{v \in V_{>n^{1/8}}} \deg(v)\right)$$
$$\leq \mathrm{CC}(G_{\leq n^{1/8}}) + \frac{|V_{>n^{1/8}}|}{n} + \frac{1}{n}\sum_{v \in V_{>n^{1/8}}} \deg(v).$$

To bound this term, we note that the probability that a random vertex v has weight greater than $n^{1/8}$ is proportional to $n^{(1-\beta)/8} = o(n^{-1/8})$. Hence, the expected size of $V_{>n^{1/8}}$ is $o(n^{7/8})$ and by a Chernoff bound, we get that $|V_{>n^{1/8}}| \leq 2\mathbb{E}\left[V_{>n^{1/8}}\right]$ with high probability, implying $|V_{>n^{1/8}}|/n = o(1)$ with high probability By Lemma 13, we further get that $\deg(v) = \mathcal{O}(w_v)$ for all $v \in V_{>n^{1/8}}$ and hence, by Lemma 14 and Lemma 12, we get

$$\sum_{v \in V_{>n^{1/8}}} \deg(v) = \mathcal{O}\left(\sum_{v \in V_{>n^{1/8}}} w_v\right) = \mathcal{O}(n^{1+(2+\eta-\beta)/8}) = o(n)$$

asymptotically almost surely and for some sufficiently small $\eta > 0$ from which our statement follows. ∎

We further require the following lemma, which formalizes that the clustering coefficient of a vertex v can equally be seen as the probability that two randomly chosen neighbors of v are adjacent.

Lemma 15 *Let v, s, t be three vertices from G, chosen uniformly at random. Denote by Δ the event that v, s, t form a triangle. We have*

$$\mathbb{E}\left[\mathrm{CC}(G)\right] = \Pr\left[\Delta \mid v \sim s, t\right] \Pr\left[\deg(v) \geq 2\right].$$

Furthermore, let $\hat{v}, \hat{s}, \hat{t}$ be the vertices v, s, t ordered increasingly by their weights. Then,

$$\mathbb{E}\left[\mathrm{CC}(G)\right] \leq \Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right].$$

Proof We start by showing the first statement. Assume that $V = \{u_1, \ldots, u_n\}$ and observe that, by linearity of expectation,

$$\mathbb{E}\left[\mathrm{CC}(G)\right] = \frac{1}{n} \sum_{u \in V} \mathbb{E}\left[\mathrm{CC}_G(u)\right] = \mathbb{E}\left[\mathrm{CC}_G(u_1)\right]$$

as every vertex has the same expected local clustering assuming that its weight is an independent sample from the Pareto distribution. It thus suffices to show that $\mathbb{E}\left[\mathrm{CC}_G(v)\right] \leq \Pr\left[\Delta \mid v \sim s, t\right]$. For this, recall that $\Gamma(v) = \{u_1, \ldots, u_k\}$ is the (random) set of neighbors of v numbered from 1 to k in some random order. Observe that $\deg(v) = |\Gamma(v)|$ and recall that the random variable $\mathrm{CC}_G(v)$ is defined as

$$\mathrm{CC}_G(v) = \frac{1}{\binom{|\Gamma(v)|}{2}} \sum_{i<j} \mathbb{1}(u_i \sim u_j),$$

where $\mathbb{1}(s \sim t)$ is an indicator random variable that is 1 if and only if s and t are connected. By linearity of expectation, we get that, for any $k \geq 2$,

$$\mathbb{E}\left[\mathrm{CC}_G(v) \mid |\Gamma(v)| = k\right] = \frac{1}{\binom{k}{2}} \sum_{i<j} \Pr\left[u_i \sim u_j \mid \deg(v) = k\right].$$

15

We proceed by showing that for any $1 \leq i < j \leq k$, we have $\mathbb{E}\left[\mathbb{1}\left(u_i \sim u_j\right) \mid \deg(v) = k\right] = \Pr\left[s \sim t \mid s, t \in \Gamma(v)\right]$. To this end, let Ω be the global sample space consisting of all possible n-vertex graphs and two of its vertices s, t chosen u.a.r. Let further $\mathcal{B} \subset \Omega$ be the set of all outcomes where $\deg(v) = k$ and where $s = u_i$ and $t = u_j$. We have,

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{1}\left(u_i \sim u_j\right) \mid \deg(v) = k\right] &= \Pr\left[u_i \sim u_j \mid \deg(v) = k\right] \\
&= \Pr_{\mathcal{B}}\left[s \sim t\right] \\
&= \Pr_{\Omega}\left[s \sim t \mid \mathcal{B}\right] \\
&= \Pr\left[s \sim t \mid (s = u_i) \cap (t = u_j) \cap (\deg(v) = k)\right] \\
&= \frac{\Pr\left[(s \sim t) \cap (s = u_i) \cap (t = u_j) \cap (\deg(v) = k) \mid s, t \in \Gamma(v)\right]}{\Pr\left[(s = u_i) \cap (t = u_j) \cap (\deg(v) = k) \mid s, t \in \Gamma(v)\right]} \\
&= \Pr\left[s \sim t \mid s, t \in \Gamma(v)\right],
\end{aligned}
$$

where the second to last equality holds because the events $s \sim t$ and $s = u_i \cap t = u_j \cap \deg(v) = k$ are independent if we condition on $s, t \in \Gamma(v)$. This implies

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{CC}_G(v) \mid \deg(v) = k\right] &= \frac{1}{\binom{k}{2}} \sum_{i<j} \Pr\left[u_i \sim u_j \mid |\Gamma(v)| = k\right]. \\
&= \Pr\left[s \sim t \mid s, t \in \Gamma(v)\right] \\
&= \Pr\left[\Delta \mid v \sim s, t\right].
\end{aligned}
$$

If $k = |\Gamma(v)| < 2$, we have that $\mathrm{CC}_G(v) = 0$, implying that in total,

$$
\mathbb{E}\left[\mathrm{CC}(G)\right] = \Pr\left[\Delta \mid v \sim s, t\right] \Pr\left[\deg(v) \geq 2\right]
$$

as desired.

For the second part, recall that we defined for all $i, j \in V$ the quantity $\kappa_{ij} = \min\{\lambda w_i w_j, n\}$ and note that

$$
\begin{aligned}
\Pr\left[v \sim s, t\right] &= \frac{\min\{\lambda w_v w_s, n\}}{n} \frac{\min\{\lambda w_v w_t, n\}}{n} \\
&\geq \frac{\min\{\lambda w_{\hat{v}} w_{\hat{s}}, n\}}{n} \frac{\min\{\lambda w_{\hat{v}} w_{\hat{t}}, n\}}{n} = \Pr\left[\hat{v} \sim \hat{s}, \hat{t}\right]
\end{aligned}
$$

because $\hat{v}$ is the vertex of minimal weight and because the events $\hat{v} \sim \hat{s}$ and $\hat{v} \sim \hat{t}$ are independent. Thus,

$$
\mathbb{E}\left[\mathrm{CC}_G(v)\right] \leq \Pr\left[\Delta \mid v \sim s, t\right] = \frac{\Pr\left[\Delta\right]}{\Pr\left[v \sim s, t\right]} \leq \frac{\Pr\left[\Delta\right]}{\Pr\left[\hat{v} \sim \hat{s}, \hat{t}\right]} = \Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right].
$$

$\blacksquare$

3.2. L_∞-norm

In this section, we analyse the clustering coefficient under L_∞-norm, which results in Theorem 3. To prove this theorem, we use Theorem 4.

Theorem 3 *Assume that $d = o(\log(n))$ and $\beta \neq 3$. Then asymptotically almost surely, the clustering coefficient of G sampled from the GIRG model with L_∞-norm fulfils*

$$\mathrm{CC}(G) = \mathcal{O}_d\left(\left(\frac{3}{4}\right)^{\min\{1,\beta-2\}d}\right) + o(1)$$

and

$$\mathrm{CC}(G) = \Omega_d\left(\max\left\{\left(\frac{1}{4}\right)^{(\beta-2)d}, \left(\frac{3}{4}\right)^d\right\}\right).$$

In particular, if $\beta > 3$, we have

$$\mathrm{CC}(G) = \Theta_d\left(\left(\frac{3}{4}\right)^d\right) + o(1).$$

Proof (1) Upper Bounds. We use Lemma 11 and thus only need an upper bound on $\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right]$. For this, we use Lemma 15, and we let v, s, t be three random vertices in $G_{\leq n^{1/8}}$ conditioned on the event that v is of minimal weight among v, s, t. If we denote by Δ be the event that v, s, t form a triangle, we get from Lemma 15 that $\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] \leq \Pr\left[\Delta \mid v \sim s, t\right]$. Accordingly, by Theorem 4, we may bound

$$\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] \leq \mathbb{E}\left[\min\left\{1, \frac{\max\{w_s, w_t\}}{w_v}\left(\frac{3}{4}\right)^d\right\} \,\middle|\, v \sim s, t\right].$$

To bound the expectation above, we analyze the distribution of $\max\{w_s, w_t\}$. Taking into account, that we consider $v, s, t \in G_{\leq n^{1/8}}$ conditioned on the event that $w_v \leq w_s, w_t$, a standard calculation shows that

$$\Pr\left[\max\{w_s, w_t\} \geq x \mid v \sim s, t\right]$$
$$= (1 + o(1))\left(1 - \left(1 - \left(\frac{x}{w_v}\right)^{2-\beta}\right)^2\right) \leq (2 + o(1))\left(\frac{x}{w_v}\right)^{2-\beta}$$

where the factor of $1 + o(1)$ comes from conditioning on $v, s, t \in G_{\leq n^{1/8}}$ and the exponent of $2 - \beta$ is due to the weight bias arising from conditioning on $v \sim s, t$. Therefore, if $\beta > 3$, the above random variable has finite expectation and we can bound

$$\mathbb{E}\left[\min\left\{1, \frac{\max\{w_s, w_t\}}{w_v}\left(\frac{3}{4}\right)^d\right\} \,\middle|\, v \sim s, t\right] \leq \left(\frac{3}{4}\right)^d \mathbb{E}\left[\frac{\max\{w_s, w_t\}}{w_v} \,\middle|\, v \sim s, t\right]$$
$$\leq (2 + o(1))\left(\frac{3}{4}\right)^d \int_1^\infty (\beta - 2)x^{\beta-2}\mathrm{d}x$$
$$\leq (2 + o(1))\frac{\beta - 2}{\beta - 3}\left(\frac{3}{4}\right)^d = \mathcal{O}_d\left(\left(\frac{3}{4}\right)^d\right),$$

which finishes the case $\beta > 3$. For the case $\beta < 3$, we instead bound

$$
\mathbb{E}\left[\min\left\{1, \frac{\max\{w_s, w_t\}}{w_v}\left(\frac{3}{4}\right)^d\right\}\,\middle|\, v \sim s, t\right]
$$

$$
\leq (2 + o(1))\left(\frac{3}{4}\right)^d \int_1^{\left(\frac{4}{3}\right)^d} (\beta - 2)x^{2-\beta}\mathrm{d}x + \Pr\left[\frac{\max\{w_s, w_t\}}{w_v} \geq \left(\frac{4}{3}\right)^d \,\middle|\, v \sim s, t\right]
$$

$$
\leq (2 + o(1))\frac{\beta - 2}{3 - \beta}\left(\frac{3}{4}\right)^d\left(\frac{4}{3}\right)^{d(3-\beta)} + (2 + o(1))\left(\frac{3}{4}\right)^{d(\beta-2)} = \mathcal{O}_d\left(\left(\frac{3}{4}\right)^{(\beta-2)d}\right)
$$

as desired.

(2) Lower Bounds. First of all, we show that $\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] = \Omega_d((3/4)^d)$. To this end, we let v, s, t be three random vertices in $G_{\leq n^{1/8}}$ and we let $\hat{v}, \hat{s}, \hat{t}$ be the vertices v, s, t reordered by their weights such that the weight of $\hat{v}$ is minimal. Then,

$$
\Pr[\Delta \mid v \sim s, t] = \frac{\Pr[\Delta]}{\Pr[v \sim s, t]} = \frac{\Pr[\hat{v} \sim \hat{s}, \hat{t}]}{\Pr[v \sim s, t]}\Pr[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}] = \frac{w_{\hat{v}}}{w_v}\Pr[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}]. \quad (3)
$$

Furthermore, by Lemma 15, we have

$$
\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] = \Pr[\Delta \mid v \sim s, t]\Pr[\deg(v) \geq 2]
$$

$$
\geq \Pr[\Delta \cap (w_{\hat{s}}, w_{\hat{t}} \leq cw_{\hat{v}}) \mid v \sim s, t]\Pr[\deg(v) \geq 2]
$$

$$
= \Pr[\Delta \mid (v \sim s, t) \cap (w_{\hat{s}}, w_{\hat{t}} \leq cw_{\hat{v}})]
$$

$$
\cdot \Pr[w_{\hat{s}}, w_{\hat{t}} \leq cw_{\hat{v}} \mid v \sim s, t]\Pr[\deg(v) \geq 2]
$$

where c is an arbitry constant greater than 1 (that does not depend on d). Now it is easy to see that $\Pr[w_{\hat{s}}, w_{\hat{t}} \leq cw_{\hat{v}} \mid v \sim s, t]$ and $\Pr[\deg(v) \geq 2]$ are both constant. To bound the remaining factor, we note that $w_{\hat{v}}/w_v \geq 1/c$ conditioned on $w_{\hat{s}}, w_{\hat{t}} \leq cw_{\hat{v}}$, so we can use (3) to obtain

$$
\Pr[\Delta \mid (v \sim s, t) \cap (w_{\hat{s}}, w_{\hat{t}} \leq cw_{\hat{v}})] \geq \frac{1}{c}\Pr[\Delta \mid (\hat{v} \sim \hat{s}, \hat{t}) \cap (w_{\hat{s}}, w_{\hat{t}} \leq cw_{\hat{v}})] \geq \frac{1}{c}\left(\frac{3}{4}\right)^d
$$

where in the last step, we used Theorem 4 which is now applicable because $\hat{v}$ is of minimal weight among v, s, t. Together, this shows that $\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] = \Omega_d((3/4)^d)$ as desired.

However, if β is very close to 2, we derive a better lower bound. To this end, we note that $\Pr[\Delta \mid v \sim s, t] = 1$ if $w_s, w_t \geq 2^d w_v$, which is easy to show using the respective connection thresholds. Hence, $\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] \geq \Pr[w_s, w_t \geq 2^d w_v \mid v \sim s, t]\Pr[\deg(v) \geq 2]$. Now, using that w_v is at most a constant with constant probability and then applying similar calculations regarding the weight distribution of s, t as in part (1) of this proof yields that

$$
\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] \geq \Pr[w_s, w_t \geq 2^d w_v \mid v \sim s, t]\Pr[\deg(v) \geq 2]
$$

$$
\geq C(2^{d(2-\beta)})^2
$$

$$
= \Omega_d((1/4)^{(\beta-2)d})
$$

for some constant $C > 0$, as desired. ∎

3.3. General L_p-norms

In this section, we generalize the previous result to other L_p-norms for $1 \leq p < \infty$. We show that, in the threshold model, one also obtains an upper bound on the clustering coefficient that decreases exponentially with d and holds with high probability. Although we do not have an explicit bound for the base of this exponential function, this result illustrates that using a different norm does not drastically change the behavior of the clustering coefficient.

Theorem 1 *Asymptotically almost surely, if $d = o(\log(n))$, the clustering coefficient of G sampled from the GIRG model under some L_p-norm with $p \in [1, \infty]$ is*

$$\mathrm{CC}(G) = \exp(-\Omega_d(d)) + o(1).$$

We start with deriving probability theoretic methods for analyzing random vectors uniformly distributed in the unit ball under L_p-norm and afterwards use them to bound the clustering coefficient (Section 3.3.2).

3.3.1. PROBABILITY-THEORETIC METHODS

We start by introducing the following useful property of the distribution of a random vector $\mathbf{x} \in \mathbb{R}^d$, which will afterwards allow us view $\mathbf{x} = \|\mathbf{x}\|_p \frac{\mathbf{x}}{\|\mathbf{x}\|_p}$ where $\|\mathbf{x}\|_p$ and $\frac{\mathbf{x}}{\|\mathbf{x}\|_p}$ are independent. In the following we show this formally and analyze the distribution of these random variables. We start with $\frac{\mathbf{x}}{\|\mathbf{x}\|_p}$ and define the following useful property of a random vector.

Definition 16 (L_p-Symmetry) *Let $\mathbf{x} \in \mathbb{R}^d$ be a random vector with density function $\rho : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$. We refer to ρ and $\mathbf{x}$ as L_p-symmetric if for all $\mathbf{y}, \mathbf{z} \in \mathbb{R}^d$ with $\|\mathbf{y}\|_p = \|\mathbf{z}\|_p$, we have $\rho(\mathbf{y}) = \rho(\mathbf{z})$. As this implies that ρ only depends on the norm $r \in \mathbb{R}$ of its argument, we also denote with $\rho(r)$ the value of ρ for any $\mathbf{z} \in \mathbb{R}^d$ with $\|\mathbf{z}\|_p = r$.*

It is easy to see that $\mathbf{x} \sim B_p(1)$ has the above property. We shall further see that any two L_p-symmetric random vectors $\mathbf{y}, \mathbf{y}'$ are equivalent in the sense that their "directions" $\mathbf{y}/\|\mathbf{y}\|_p$ and $\mathbf{y}'/\|\mathbf{y}'\|_p$ are identically distributed. This allows us to sample the random vector $\mathbf{x}/\|\mathbf{x}\|_p$ from an arbitrary L_p-symmetric distribution.

Lemma 17 (Equivalence of L_p-Symmetric Density Functions) *Let $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{R}^d$ be two random vectors with density functions ρ and $\tilde{\rho}$ respectively, both L_p-symmetric. Then, the random vectors $\mathbf{z} := \frac{\mathbf{x}}{\|\mathbf{x}\|_p}$ and $\tilde{\mathbf{z}} := \frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|_p}$ are identically distributed.*

Before we prove this lemma, we introduce some further notation and some auxiliary statements. Let $S \subseteq S_p(1)$ be some subset of the (surface of the) unit sphere under L_p-norm. We define the set $S(r) = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{x}/\|\mathbf{x}\|_p \in S, \|\mathbf{x}\|_p \leq r\}$, which contains all vectors from $\mathbb{R}^d$ with norm at most r that are in S when projected to $S_p(1)$. We further denote by $\nu(r)$ the volume of the unit ball of radius r and by $\nu_S(r)$ the volume of the set $S(r)$. We start by showing the following useful property.

Lemma 18 *Let $S \subseteq S_p(1)$, and let $S(r)$, $\nu_S(r)$, and $\nu(r)$ be defined as above. We have*

$$\nu_S(r) = r^d \cdot \nu_S(1) = \nu(r) \frac{\nu_S(1)}{\nu(1)}.$$

Proof We note that for any $r \geq 0$,

$$\nu_S(r) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \mathbb{1}((x_1, \ldots, x_d) \in S(r)) \mathrm{d}x_1 \ldots \mathrm{d}x_d.$$

Substituting $x_i = r \cdot y_i$ yields

$$\nu_S(r) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \mathbb{1}(r \cdot (y_1, \ldots, y_d) \in S(r)) r^d \mathrm{d}y_1 \ldots \mathrm{d}y_d$$
$$= r^d \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \mathbb{1}((y_1, \ldots, y_d) \in S(1)) \mathrm{d}y_1 \ldots \mathrm{d}y_d$$
$$= r^d \nu_S(1).$$

which shows the first part of our statement. For the second part, we observe that $\nu(r) = \nu_{S_p(1)}(r)$, and thus immediately obtain $\nu(r) = r^d \nu(1)$. Hence, $r^d = \nu(r)/\nu(1)$, implying that $\nu_S(r) = \nu(r) \frac{\nu_S(1)}{\nu(1)}$. ∎

We continue by showing that we can express the probability of the event $\mathbf{x}/\|\mathbf{x}\|_p \in S$ for any L_p-symmetric random vector $\mathbf{x}$ in the following way.

Lemma 19 *Let $\mathbf{x} \in \mathbb{R}^d$ be a random vector with L_p-symmetric density function ρ and let $S \subseteq S_p(1)$. We have*

$$\Pr\left[\frac{\mathbf{x}}{\|\mathbf{x}\|_p} \in S \right] = \int_0^{\infty} \rho(r) \frac{\mathrm{d}\nu_S(r)}{\mathrm{d}r} \mathrm{d}r.$$

Proof We define for any $\mathbf{x} \in \mathbb{R}^d$ the indicator function

$$\mathbb{1}_S(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x}/\|\mathbf{x}\|_p \in S \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, we define $\mathbf{z} := \mathbf{x}/\|\mathbf{x}\|_p$. For simplicity, we assume that S is located in only one of the 2^d orthants of the standard d-dimensional cartesian coordinate system, the argumentation for the case where S spans multiple orthants are analogously obtained by splitting S into parts that each span one orthant, and afterwards summing over them. Therefore, in the following, we assume that $S \subseteq \mathbb{R}_{>0}^d$. We note that we may express

$$\Pr[\mathbf{z} \in S] = \int_{\mathbb{R}_{>0}^d} \mathbb{1}_S(\mathbf{x}) \rho(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{4}$$

where $\mathbf{x} = (x_1, \ldots, x_d)^T$. We get from (Spivak, 1998, Theorem 3-13, page 67) that if $A \subset \mathbb{R}^d$ is an open set and if $\varphi : A \to \mathbb{R}^d$ is an injective, continuously differentiable function such that $\det(J\varphi(\mathbf{x})) \neq 0$ for all $\mathbf{x} \in A$, then if $f : \varphi(A) \to \mathbb{R}^d$ is integrable,

$$\int_{\varphi(A)} f(\mathbf{x}) \mathrm{d}\mathbf{x} = \int_A f(\varphi(\mathbf{y})) |\det(J\varphi(\mathbf{y}))| \mathrm{d}\mathbf{y},$$

where $J\varphi(\mathbf{x})$ denotes the Jacobian matrix of φ at the point $\mathbf{x}$. We define A_r as the open set $A_r = \{(r, x_2, \ldots, x_d) \in \mathbb{R}_{>0}^d \mid \sum_{i=2}^d x_i^p < r^p\}$ and $A = \bigcup_{r>0} A_r$. Furthermore, we let

$$\varphi : A \to \mathbb{R}^d, (r, x_2, \ldots, x_d) \mapsto \left(\left(r^p - \sum_{i=2}^d x_i^p \right)^{1/p}, x_2, \ldots, x_d \right).$$

We note that this function is injective and that it has the remarkable property that for any $\mathbf{x} = (r, x_2, \ldots, x_d) \in A$, $\|\varphi(\mathbf{x})\|_p = r$. Furthermore, we have $J\varphi_{ij} = 0$ for $i, j \geq 2, i \neq j$, $J\varphi_{ij} = 1$ for $i = j \geq 2$ and

$$J\varphi_{11} = \frac{\partial}{\partial r} \left(r^p - \sum_{i=2}^d x_i^p \right)^{1/p} = r^{p-1} \left(r^p - \sum_{i=2}^d x_i^p \right)^{1/p-1}.$$

Furthermore, for all $i \geq 2$, we have

$$J\varphi_{1i} = \frac{\partial}{\partial x_i} \left(r^p - \sum_{i=2}^d x_i^p \right)^{1/p} = -x_i^{p-1} \left(r^p - \sum_{i=2}^d x_i^p \right)^{1/p-1}.$$

Hence, φ is continuously differentiable. Moreover, since $A \subseteq \mathbb{R}_{>0}^d$, we get that for all $1 \leq i \leq d$ and $\mathbf{x} \in A$, we have $J\varphi_{1i} \neq 0$ and $J\varphi_{ii} \neq 0$, but for all $i, j \geq 2, i \neq j$, we have $J\varphi_{ij} = 0$. For this reason the columns of $J\varphi(\mathbf{x})$ are not linearly dependent and so $\det(d\varphi(\mathbf{x})) \neq 0$. In the following, we denote $|\det(J\varphi(\mathbf{x}))|$ with $g(\mathbf{x})$. We can hence transform Equation (4) as

$$\begin{aligned}
\Pr[\mathbf{z} \in S] &= \int_{\mathbb{R}_{>0}^d} \mathbb{1}_S(\mathbf{x})\rho(\mathbf{x})\mathrm{d}\mathbf{x} \\
&= \int_A \mathbb{1}_S(\varphi(\mathbf{y}))\rho(\varphi(\mathbf{y}))g(\mathbf{y})\mathrm{d}\mathbf{y} \\
&= \int_0^\infty \ldots \int_0^\infty \mathbb{1}_S(\varphi(\mathbf{y}))\mathbb{1}(\mathbf{y} \in A)\rho(\varphi(\mathbf{y}))g(\mathbf{y})\mathrm{d}x_d \ldots \mathrm{d}x_2\mathrm{d}r,
\end{aligned}$$

where $\mathbf{y} = (r, x_2, \ldots, x_d)$ and $\mathbb{1}(\mathbf{y} \in A)$ is an indicator function, which is equal to 1 if $\mathbf{y} \in A$ and 0 otherwise. We note that for any $\mathbf{y} = (r, x_2, \ldots, x_d) \in A$, we have $\|\varphi(\mathbf{y})\|_p = r$. Since $\rho(\mathbf{x})$ is L_p-symmetric it only depends on the norm of $\mathbf{x}$, hence $\rho(\varphi(\mathbf{y}))$ only depends on the first component r of $\mathbf{y}$. We may therefore rewrite $\rho(\varphi(\mathbf{y})) = \rho(r)$ and rearrange

$$\Pr[\mathbf{z} \in S] = \int_0^\infty \rho(r) \int_0^\infty \ldots \int_0^\infty \mathbb{1}_S(\varphi(\mathbf{y}))\mathbb{1}(\mathbf{y} \in A)g(\mathbf{y})\mathrm{d}x_d \ldots \mathrm{d}x_2\mathrm{d}r.$$

We define for any $r > 0$,

$$v_S(r) := \int_0^\infty \ldots \int_0^\infty \mathbb{1}_S(\varphi(\mathbf{y}))\mathbb{1}(\mathbf{y} \in A)g(\mathbf{y})\mathrm{d}x_d \ldots \mathrm{d}x_2$$

and thus obtain

$$\Pr[\mathbf{z} \in S] = \int_0^\infty \rho(r)v_S(r)\mathrm{d}r. \tag{5}$$

Now, recall that $\nu_S(R)$ is the volume of the set $S(R) = \{x \in \mathbb{R}^d \mid x/\|x\|_p \in S, \|x\|_p \le R\}$. We show that in fact $v_S(R) = \frac{\mathrm{d}\nu_S(R)}{\mathrm{d}R}$ for all $R > 0$. This gives Equation (5) an intuitive interpretation as integrating ρ over r along the sphere radius r under L_p-norm. Note that

$$\nu_S(R) = \int_{\mathbb{R}^d} \mathbb{1}(\mathbf{x} \in S(R))\mathrm{d}\mathbf{x}.$$

Now, with the same argumentation as above (and by omitting ρ), we obtain

$$\begin{aligned}
\nu_S(R) &= \int_0^\infty \dots \int_0^\infty \mathbb{1}(\varphi(\mathbf{y}) \in S(R))\mathbb{1}(\mathbf{y} \in A)g(\mathbf{y})\mathrm{d}x_d \dots \mathrm{d}x_2 \mathrm{d}r \\
&= \int_0^\infty \mathbb{1}(r \le R) \int_0^\infty \dots \int_0^\infty \mathbb{1}_S(\varphi(\mathbf{y}))\mathbb{1}(\mathbf{y} \in A)g(\mathbf{y})\mathrm{d}x_d \dots \mathrm{d}x_2 \mathrm{d}r \\
&= \int_0^\infty \mathbb{1}(r \le R)v_S(r)\mathrm{d}r = \int_0^R v_S(r)\mathrm{d}r
\end{aligned}$$

where we used that for all $\mathbf{y} \in A$, we have $\mathbb{1}(\varphi(\mathbf{y}) \in S(R)) = \mathbb{1}(r \le R)\mathbb{1}_S(\varphi(\mathbf{y}))$. Applying the Leibnitz integral rule, we get $\frac{\mathrm{d}\nu_S(R)}{\mathrm{d}R} = v_S(R)$, which finishes the proof. ∎

The above two statements imply the following corollary, which in turn implies Lemma 17.

Corollary 20 *Let $\mathbf{x}$ be an L_p-symmetric random vector and let $S \subseteq S_p(1)$. We have*

$$\Pr\left[\frac{\mathbf{x}}{\|\mathbf{x}\|_p} \in S\right] = \frac{\nu_S(1)}{\nu(1)}.$$

Proof define $\mathbf{z} := \mathbf{x}/\|\mathbf{x}\|_p$. By Lemma 19, we may express

$$\Pr[\mathbf{z} \in S] = \int_0^\infty \rho(r)\frac{\mathrm{d}\nu_S(r)}{\mathrm{d}r}\mathrm{d}r.$$

Furthermore, we have by Lemma 18 that $\nu_S(R) = \nu(R)\frac{\nu_S(1)}{\nu(1)}$ and hence,

$$\frac{\mathrm{d}\nu_S(R)}{\mathrm{d}R} = \frac{\nu_S(1)}{\nu(1)}\frac{\mathrm{d}\nu(R)}{\mathrm{d}R}.$$

Accordingly,

$$\begin{aligned}
\Pr[\mathbf{z} \in S] &= \int_0^\infty \rho(r) \cdot \frac{\mathrm{d}\nu_S(R)}{\mathrm{d}r}\mathrm{d}r \\
&= \frac{\nu_S(1)}{\nu(1)}\int_0^\infty \rho(r) \cdot \frac{\mathrm{d}\nu(R)}{\mathrm{d}r}\mathrm{d}r.
\end{aligned}$$

We note that $\Pr[\mathbf{z} \in S_p(1)] = 1$, and so, by Lemma 19, we get

$$\int_0^\infty \rho(r) \cdot \frac{\mathrm{d}\nu(R)}{\mathrm{d}r}\mathrm{d}r = \Pr[\mathbf{z} \in S_p(1)] = 1.$$

This shows

$$\Pr\left[\mathbf{z} \in S\right] = \frac{\nu_S(1)}{\nu(1)}.$$

∎

With this statement, we may now prove Lemma 17.

Proof [Proof of Lemma 17.] We show that for any $S \subseteq S_p(1)$, we have that $\Pr\left[\mathbf{z} \in S\right] = \Pr\left[\tilde{\mathbf{z}} \in S\right]$. Because $\mathbf{x}$ and $\tilde{\mathbf{x}}$ are both L_p-symmetric, we get by Theorem 20 that both $\Pr\left[\mathbf{z} \in S\right]$ and $\Pr\left[\tilde{\mathbf{z}} \in S\right]$ are equal to $\frac{\nu_S(1)}{\nu(1)}$, which directly implies the desired statement. ∎

The χ^p-Distribution In addition to the distribution of $\mathbf{x} \sim B_p(1)$, we need another L_p-symmetric distribution. For this purpose recall the definitions of the $\chi_p(d)$ and the $\chi^p(d)$ distributions from the introduction. It is easy to see that a random vector $\mathbf{x} \sim \chi_p(d)$ is L_p-symmetric by observing that its density function is

$$\rho_{\mathbf{x}}(\mathbf{x}) = \prod_{i=1}^{d} \gamma e^{-\frac{1}{2}|\mathbf{x}(i)|^p} = \gamma^d e^{-\frac{1}{2}\sum_{i=1}^{d}|\mathbf{x}(i)|^p} = \gamma^d e^{-\frac{1}{2}(\|\mathbf{x}\|_p)^p}$$

and thus only depends on the norm of $\mathbf{x}$. We further note that for the case $p = 2$, $\chi_2(d)$ is the standard d-variate normal distribution $\mathcal{N}(0, I_d)$ (where I_d is the $d \times d$ identity matrix), and that $\chi^2(d)$ is the chi-squared distribution with d degrees of freedom. The distribution $\chi^p(d)$ can hence be seen as a generalization of the chi-squared distribution to other L_p-norms.

We further verify that γ is indeed the correct normalization constant. For this, let $X \sim \chi_p(1)$ and observe that

$$1 = \int_{-\infty}^{\infty} \rho_X(x)\mathrm{d}x = \gamma \cdot 2 \int_{0}^{\infty} e^{-\frac{1}{2}x^p}\mathrm{d}x.$$

With the substitution $x = (2y)^{\frac{1}{p}}$, we obtain

$$\gamma \cdot 2 \int_{0}^{\infty} e^{-\frac{1}{2}x^p}\mathrm{d}x = \gamma \cdot 2 \int_{0}^{\infty} \frac{2^{1/p}}{p} y^{1/p-1} e^{-y}\mathrm{d}y = \gamma \frac{2^{1/p+1}\Gamma\left(\frac{1}{p}\right)}{p}.$$

We hence get

$$\gamma = \frac{p}{2^{\frac{1}{p}+1}\Gamma\left(\frac{1}{p}\right)},$$

Note that for $p = 2$, one does indeed obtain the correct normalization constant of the standard normal distribution $\mathcal{N}(0, 1)$, which is equal to $1/\sqrt{2\pi}$.

We continue with deriving a tail bound on the $\chi^p(d)$ distribution and start with deriving its moment-generating function.

Lemma 21 *Let $Z \sim \chi^p(1)$. Let ψ_Z be the moment generating function of Z, defined as*

$$\psi_Z : \mathbb{R}_0^+ \to \mathbb{R}, \psi_Z(\lambda) = \mathbb{E}\left[e^{\lambda Z}\right].$$

Then, for every $\lambda < \frac{1}{2}$, we have

$$\psi_Z(\lambda) = (1 - 2\lambda)^{-\frac{1}{p}}.$$

Proof Let $X \sim \chi_p(1)$ and note that we may write $Z = |X|^p$. Recall that the probability density of X is $\rho_X(x) = \gamma e^{-\frac{1}{2}|x|^p}$. Denote by ρ_Z the density function of Z and observe that

$$\rho_Z(x) = \frac{\mathrm{d}\mathrm{Pr}\left[Z \geq x\right]}{\mathrm{d}x} = \frac{\mathrm{d}\mathrm{Pr}\left[|X|^p \geq x\right]}{\mathrm{d}x} = \frac{\mathrm{d}\mathrm{Pr}\left[|X| \geq x^{\frac{1}{p}}\right]}{\mathrm{d}x} = \rho_{|X|}\left(x^{\frac{1}{p}}\right)\frac{\mathrm{d}x^{\frac{1}{p}}}{\mathrm{d}x}$$

$$= 2\rho_X\left(x^{\frac{1}{p}}\right)\frac{\mathrm{d}x^{\frac{1}{p}}}{\mathrm{d}x} = 2\gamma e^{-\frac{1}{2}x}\frac{1}{p}x^{\frac{1}{p}-1} = \frac{x^{\frac{1}{p}-1}e^{-\frac{1}{2}x}}{2^{\frac{1}{p}}\Gamma\left(\frac{1}{p}\right)}.$$

Note that, in the fifth equality, we used that $\rho_{|X|}(x) = 2\rho_X(x)$. We continue by deriving the moment-generating function of the random variable Z. We obtain

$$\psi_Z(\lambda) = \mathbb{E}\left[e^{\lambda Z}\right] = \int_0^\infty \rho_Z(x)e^{\lambda x}dx$$

$$= \frac{1}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\int_0^\infty x^{\frac{1}{p}-1}e^{-x(1/2-\lambda)}dx.$$

We note that this integral exists for $\lambda < \frac{1}{2}$. With the substitution $x = y(1/2 - \lambda)^{-1}$, it transforms to

$$\psi_Z(\lambda) = \frac{1}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\int_0^\infty x^{\frac{1}{p}-1}e^{-x(1/2-\lambda)}dx$$

$$= \frac{1}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\int_0^\infty y^{\frac{1}{p}-1}e^{-y}\frac{(1/2-\lambda)^{1-\frac{1}{p}}}{1/2-\lambda}dy$$

$$= \frac{(1/2-\lambda)^{-\frac{1}{p}}}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\Gamma\left(1/p\right)$$

$$= (1-2\lambda)^{-\frac{1}{p}}.$$

∎

Corollary 22 *Let $Z \sim \chi^p(d)$. Then,*

$$\mathbb{E}\left[Z\right] = \frac{2d}{p}.$$

Proof Let $X \sim \chi^p(1)$. We get $\mathbb{E}\left[Z\right] = d \cdot \mathbb{E}\left[X\right]$ as Z is the sum of d independent random variables distributed identically as X. We further note that the expectation of X is equal to the derivative of its moment-generating function at $\lambda = 0$. We get from Lemma 21 that

$$\frac{\mathrm{d}\psi_X(\lambda)}{\mathrm{d}\lambda} = \frac{2}{p}(1-2\lambda)^{-\frac{1}{p}-1}$$

and hence, $\mathbb{E}\left[X\right] = \frac{2}{p}$.

∎

We continue by showing that a random variable $Z \sim \mathcal{X}^p(d)$ is concentrated around its expected value. Under the hood, our bounds are obtained in the same way as the Chernoff-Hoeffding bounds, namely by applying Markov's inequality to the moment generating function of Z. However, instead of doing this directly, we take a shortcut by applying the following variant of Bernstein's inequality that is proven by Massart in Morel et al. (2007).

Theorem 23 (Proposition 2.9 in Morel et al. (2007)) *Let $X_1, \dots, X_d$ be independent, real-valued random variables. Assume that there exist constants $v, c > 0$ such that*

$$\sum_{i=1}^{d} \mathbb{E}\left[X_i^2\right] \leq v$$

and that for all integers $k \geq 3$,

$$\sum_{i=1}^{d} \mathbb{E}\left[|X_i|^k\right] \leq \frac{k!}{2}vc^{k-2}.$$

Let $S = \sum_{i=1}^{d}(X_i - \mathbb{E}[X_i])$. Then, for every $x > 0$,

$$\Pr\left[S \geq \sqrt{2vx} + cx\right] \leq \exp(-x).$$

With this, we are able to show the following.

Theorem 24 *Let $X_1, \dots, X_d$ be i.i.d. random variables from $\mathcal{X}_p(1)$ and define the random variable $Z := \sum_{i=1}^{d} |X_i|^p$. Note that $Z \sim \mathcal{X}^p(d)$. Then, for all $x > 0$,*

(i) $\Pr\left[Z \geq \mathbb{E}[Z] + 2\sqrt{2\mathbb{E}[Z]x} + 2x\right] \leq \exp(-x)$

(ii) $\Pr\left[Z \leq \mathbb{E}[Z] - 2\sqrt{2\mathbb{E}[Z]x} - 2x\right] \leq \exp(-x)$.

Proof We use Theorem 23. To show that the random variables $|X_1|^p, \dots, |X_d|^p$ fulfill the conditions of Theorem 23, we derive bounds on its moments. For any $X \sim \mathcal{X}_p(1)$, define $Y = |X|^p$. We use the moment generating function from Lemma 21 to derive bounds on the moments of Y. For all integers $k \geq 0$, we note that we have $\mathbb{E}[Y^k] = \psi_Y^{(k)}(0)$, where $\psi_Y^{(k)}$ denotes the k-th derivative of ψ_Y. We note that

$$\psi_Y'(\lambda) = \frac{2}{p}\left(1 - 2\lambda\right)^{-\frac{1}{p}-1}$$

and

$$\psi_Y''(\lambda) = \frac{4}{p}\left(\frac{1}{p} + 1\right)\left(1 - 2\lambda\right)^{-\frac{1}{p}-2},$$

from which we derive $\mathbb{E}[Y] = \frac{2}{p}$ and $\mathbb{E}[Y^2] = \frac{4}{p}\left(\frac{1}{p} + 1\right)$. For $k \geq 3$ one can easily verify that

$$\psi_Y^{(k)}(\lambda) = (1 - 2\lambda)^{-\frac{1}{p}-k} \mathbb{E}[Y^2] 2^{k-2} \prod_{i=2}^{k-1}\left(\frac{1}{p} + i\right)$$

25

and hence,

$$\mathbb{E}\left[Y^k\right] = \psi_Y^{(k)}(0) = \mathbb{E}\left[Y^2\right] 2^{k-2} \prod_{i=2}^{k-1}\left(\frac{1}{p}+i\right)$$

$$= \mathbb{E}\left[Y^2\right] 2^{k-2} \prod_{i=1}^{k-2}\left(\frac{1}{p}+i+1\right)$$

$$\leq \mathbb{E}\left[Y^2\right] 2^{k-2} \prod_{i=1}^{k-2}(i+2) = \mathbb{E}\left[Y^2\right] 2^{k-2}\frac{k!}{3!} \leq \mathbb{E}\left[Y^2\right] 2^{k-1}\frac{k!}{2}. \quad (6)$$

Recall that we have $\mathbb{E}\left[Y^2\right] = \frac{4}{p}\left(\frac{1}{p}+1\right)$ and hence, $\mathbb{E}\left[Y^2\right] \leq \frac{8}{p}$ due to $p \geq 1$. If we define $Y_i = |X_i|^p$ and set $v = 8d/p, c = 2$, we have that

$$\sum_{i=1}^{d}\mathbb{E}\left[Y_i^2\right] \leq \frac{8d}{p} = v$$

and thus, for all $k \geq 3$,

$$\sum_{i=1}^{d}\mathbb{E}\left[Y_i^k\right] \leq d\mathbb{E}\left[Y^2\right] 2^{k-1}\frac{k!}{2} \leq \frac{k!}{2}vc^{k-2},$$

which shows that the conditions of Theorem 23 are fulfilled. Since $Z = \sum_{i=1}^{d}Y_i$ and $\mathbb{E}[Z] = \frac{2d}{p}$, we get that for all $x > 0$,

$$\Pr\left[Z - \mathbb{E}[Z] \geq \sqrt{16d/p \cdot x} + 2x\right] = \Pr\left[Z \geq \mathbb{E}[Z] + 2\sqrt{2\mathbb{E}[Z] \cdot x} + 2x\right]$$
$$\leq \exp(-x),$$

which shows the first statement.

For the second statement, we define $Y_i' := -Y_i$ and note that $-Z = \sum_{i=1}^{d}Y_i'$. Furthermore, we have that $\mathbb{E}\left[Y_i'^2\right] = \mathbb{E}\left[Y_i^2\right]$ and $\mathbb{E}\left[|Y_i'|^k\right] = \mathbb{E}\left[Y_i^k\right]$ for all integers $k \geq 0$. We have that

$$\sum_{i=1}^{d}\mathbb{E}\left[Y_i'^2\right] = \sum_{i=1}^{d}\mathbb{E}\left[Y_i^2\right] \leq \frac{8d}{p} = v$$

and for all $k \geq 3$, we get from Equation (6) that

$$\sum_{i=1}^{d}\mathbb{E}\left[|Y_i'|^k\right] = \sum_{i=1}^{d}\mathbb{E}\left[Y_i^k\right] \leq \frac{k!}{2}vc^{k-2}.$$

Hence, it follows from Theorem 23 that

$$\Pr\left[-Z + \mathbb{E}[Z] \geq \sqrt{16d/p \cdot x} + 2x\right] = \Pr\left[Z \leq \mathbb{E}[Z] - 2\sqrt{2\mathbb{E}[Z]x} - 2x\right]$$
$$\leq \exp(-x),$$

which implies the second statement. ■

We can slightly reformulate this bound such that it is more convenient to work with them. Observe the similarity of the following bounds with the Chernoff-Hoeffding bound from Theorem 7.

Corollary 2 *Let $X_i, \ldots X_d$ be i.i.d. random variables from $\mathcal{X}_p(1)$ and define $Z = \sum_{i=1}^d |X_i|^p \sim \mathcal{X}^p(d)$. Then, for every $\varepsilon > 0$,*

$$\Pr\left[|Z - \mathbb{E}\left[Z\right]| \geq \varepsilon \cdot \mathbb{E}\left[Z\right]\right] \leq 2\exp\left(-\frac{2\delta}{p}d\right).$$

Where $\delta > 0$ is defined by $\varepsilon = 2(\sqrt{2\delta} + \delta)$.

Proof We use Theorem 24 and set $x = \delta\mathbb{E}\left[Z\right]$. We then obtain

$$\Pr\left[Z \geq \mathbb{E}\left[Z\right] + \mathbb{E}\left[Z\right] \cdot 2\sqrt{2\delta} + \mathbb{E}\left[Z\right] \cdot 2\delta\right] = \Pr\left[Z \geq \mathbb{E}\left[Z\right](1 + 2\sqrt{2\delta} + 2\delta)\right]$$
$$\leq \exp(-\delta\mathbb{E}\left[Z\right]).$$

Recalling from Theorem 22 that $\mathbb{E}\left[Z\right] = \frac{2d}{p}$ then implies that $\Pr\left[Z \geq (1 + \varepsilon)\mathbb{E}\left[Z\right]\right] \leq \exp(-2d\delta/p)$ for $\varepsilon = 2(\sqrt{2\delta} + \delta)$. The argumentation for the second statement is analogous. ■

3.3.2. BOUNDING THE CLUSTERING COEFFICIENT

We use the insights gained so far to prove a bound on the probability that two random neighbors of a vertex s that have bounded weight are adjacent.

Lemma 25 *Let $G = G(n, d, \beta, w_0)$ be a GIRG sampled under L_p-norm. After sampling the weights, let s, u, v be three vertices in $G_{\leq n^{1/8}}$ with $w_s \leq w_u, w_v$ and $w_u, w_v \leq c^d w_s$ for some $c > 1$. Let Δ be the event that $\{s, u, v\}$ form a triangle. Then, there exists a choice for c such that there are constants $a, b > 0, c > 1$ such that for sufficiently large n and all $d \geq 1, d = o(\log(n))$,*

$$\Pr\left[\Delta \mid s \sim u, v\right] \leq a \cdot \exp(-bd).$$

Here, the randomness originates from the assignment of coordiates in $\mathbb{T}^d$ to s, u, v.

Proof Recall that $B_p(r)$ is the ball of radius r under L_p norm. We assume that n is large enough such that the ball of volume $\lambda w_s^2 c^{2d}/n$ has a radius of $r \leq 1/4$. Note that this is possible since $d = o(\log(n))$. With this we may simply measure the distance of two points $\mathbf{x}, \mathbf{y} \in B_p(r)$ as $\|\mathbf{x} - \mathbf{y}\|_p$ and assume that t_{uv} is precisely the radius of the ball of volume $\lambda w_u w_v/n$.

Now, assuming $s \sim u, v$ and $w_u, w_v \leq c^d w_s$, the vertices u, v are uniformly distributed within the balls $B_p(t_{sv})$ and $B_p(t_{su})$ (centered at the position of s), respectively. Assuming the position of s is the origin of our coordinate system, we denote by $\mathbf{x}_u, \mathbf{x}_v$ the (random) positions of u, v. Hence, the probability that u and v are connected is simply $\Pr[\|\mathbf{x}_u - \mathbf{x}_v\|_p \leq t_{uv}]$. If we denote by $\nu(r)$ the volume of the ball $B_p(r)$, we further note that $\nu(r) = r^d\nu(1)$ (cf. Lemma 18), and since we choose t_{uv} such that $\nu(t_{uv}) = \lambda w_u w_v/n$, we get

$$t_{uv} = \left(\frac{\lambda w_u w_v}{\nu(1)n}\right)^{1/d}. \tag{7}$$

27

In the following, we derive an upper bound for $\Pr\left[\|\mathbf{x}_u - \mathbf{x}_v\|_p \leq t_{uv}\right]$. We note that we can equivalently describe the random variables $\mathbf{x}_u, \mathbf{x}_v$ as $\mathbf{x}_u = t_{us}\mathbf{y}_u$ and $\mathbf{x}_v = t_{vs}\mathbf{y}_v$, where $\mathbf{y}_u$ and $\mathbf{y}_v$ are i.i.d. random vectors uniformly distributed according to the standard Lebesgue measure in $B_p(1)$. With this, we reformulate the probability $\Pr\left[\|\mathbf{x}_u - \mathbf{x}_v\|_p \leq t_{uv}\right]$ as

$$\Pr\left[\|\mathbf{x}_u - \mathbf{x}_v\|_p \leq t_{uv}\right] = \Pr\left[\|t_{us}\mathbf{y}_u - t_{vs}\mathbf{y}_v\|_p \leq t_{uv}\right]$$
$$= \Pr\left[\|\mathbf{y}_u - (t_{vs}/t_{us})\mathbf{y}_v\|_p \leq t_{uv}/t_{us}\right]$$
$$= \Pr\left[\|\mathbf{y}_u - (w_v/w_u)^{1/d}\mathbf{y}_v\|_p \leq (w_v/w_s)^{1/d}\right].$$

To find an upper bound for this probability, we instead lower bound the probability of the event that

$$\left\|\mathbf{y}_u - (w_v/w_u)^{1/d}\mathbf{y}_v\right\|_p > (w_v/w_s)^{1/d}.$$

Since $w_v, w_s \in [w_s, c^d w_s]$, we have $(w_v/w_s)^{1/d} \leq c$ and hence, it suffices to lower bound

$$\Pr\left[\left\|\mathbf{y}_u - (w_v/w_u)^{1/d}\mathbf{y}_v\right\|_p > c\right]$$

or equivalently

$$\Pr\left[\left(\left\|\mathbf{y}_u - (w_v/w_u)^{1/d}\mathbf{y}_v\right\|_p\right)^p > c^p\right].$$

For this, we start by investigating the properties of the random vectors $\mathbf{y}_u, \mathbf{y}_v \sim B_p(1)$. Recall from Lemma 17 that we may equivalently express the random vector $\mathbf{y} \sim B_p(1)$ as $\mathbf{y} = \|\mathbf{y}\|_p \cdot \mathbf{y}/\|\mathbf{y}\|_p$ where $\|\mathbf{y}\|_p$ and $\mathbf{y}/\|\mathbf{y}\|_p$ are independent. Accordingly, $\mathbf{y}$ is identically distributed as the product of a random variable r identically distributed as $\|\mathbf{y}\|_p$, and a random vector $\mathbf{z}$ identically distributed as $\mathbf{y}/\|\mathbf{y}\|_p$.

We note that r and $\|\mathbf{y}\|_p$ are distributed such that for any $0 \leq \zeta \leq 1$, we have

$$\Pr\left[\|\mathbf{y}\|_p \leq \zeta\right] = \frac{\nu_p(\zeta)}{\nu_p(1)} = \zeta^d$$

and thus,

$$\Pr\left[\|\mathbf{y}\|_p \geq \zeta\right] = 1 - \zeta^d.$$

Furthermore, due to the L_p-symmetry of $\mathbf{y_u}, \mathbf{y_v}$ and Lemma 17, we assume that $\mathbf{z} = \tilde{\mathbf{z}}/\|\tilde{\mathbf{z}}\|_p$ where $\tilde{\mathbf{z}}$ is a random vector from the $\mathcal{X}_p(d)$-distibution.

In the following, we hence assume that $\mathbf{y}_u = r_u \cdot \tilde{\mathbf{z}}_u/\|\tilde{\mathbf{z}}_u\|_p$, and $\mathbf{y}_v = r_v \cdot \tilde{\mathbf{z}}_v/\|\tilde{\mathbf{z}}_v\|_p$, for suitable, independent random variables r_u, r_v and independent random vectors $\tilde{\mathbf{z}}_u, \tilde{\mathbf{z}}_v \sim \mathcal{X}_p(d)$.

With this observation, we find a lower bound for

$$\Pr\left[\left(\left\|\mathbf{y}_u - (w_v/w_u)^{1/d}\mathbf{y}_v\right\|_p\right)^p > c^p\right].$$

We first rewrite the term $\left(\left\| \mathbf{y}_u - (w_v/w_u)^{1/d} \mathbf{y}_v \right\|_p \right)^p$ as

$$\left(\left\| \mathbf{y}_u - \left(\frac{w_v}{w_u} \right)^{1/d} \mathbf{y}_v \right\|_p \right)^p = \sum_{i=1}^{d} \left| \mathbf{y}_u(i) - \left(\frac{w_v}{w_u} \right)^{1/d} \mathbf{y}_v(i) \right|^p = S_1 + S_2,$$

where S_1 is the sum of all components in which y_{ui} and y_{vi} have opposite sign, and S_2 is the sum of all remaining components. We show that there are constants $a, b > 0, c > 1$ such that $S_1 + S_2$ is greater than c^p with probability at least $1 - a \cdot \exp(-bd)$. In this section, we refer to an event as happening *with overwhelming probability*[2] if there are constants $a, b > 0$ such that the event happens with probability at least $1 - a \cdot \exp(-bd)$. Note that, if two events $\mathbf{E}_1$ and $\mathbf{E}_2$ happen with overwhelming probability, then also $\mathbf{E}_1 \cap \mathbf{E}_2$ happens with overwhelming probability as, by a union bound, we have $\Pr\left[\overline{\mathbf{E}_1 \cap \mathbf{E}_2}\right] \leq a \cdot \exp(-bd) + a' \cdot \exp(-b'd)$ for some $a, a', b, b' > 0$ and thus $\Pr\left[\mathbf{E}_1 \cap \mathbf{E}_2\right] \geq 1 - 2\max\{a, a'\} \exp(-\max\{b, b'\}d)$.

We start with giving a lower bound for S_1. Let I_1 be the set of all component indices i in which $\mathbf{y}_u(i)$ and $\mathbf{y}_v(i)$ have opposite sign. Note that this implies that

$$\left| \mathbf{y}_u(i) - (w_v/w_u)^{1/d}\mathbf{y}_v(i) \right| = |\mathbf{y}_u(i)| + (w_v/w_u)^{1/d} |\mathbf{y}_v(i)| .$$

Furthermore, note that we may express $\mathbf{y}_u(i) = r_u \cdot \tilde{\mathbf{z}}_u(i)/ \|\tilde{\mathbf{z}}_u\|_p$. Since $w_u \leq w_s \cdot c^d$ and $w_v \geq w_s$, we further have $(w_v/w_u)^{1/d} \geq 1/c$ and can thus rewrite S_1 as

$$S_1 = \sum_{i \in I_1} \left(r_u \left| \frac{\tilde{\mathbf{z}}_u(i)}{\|\tilde{\mathbf{z}}_u\|_p} \right| + \left(\frac{w_v}{w_u} \right)^{1/d} r_v \left| \frac{\tilde{\mathbf{z}}_v(i)}{\|\tilde{\mathbf{z}}_v\|_p} \right| \right)^p$$

$$\geq \sum_{i \in I_1} \left(\left(r_u \frac{|\tilde{\mathbf{z}}_u(i)|}{\|\tilde{\mathbf{z}}_u\|_p} \right)^p + \left(\frac{r_v}{c} \frac{|\tilde{\mathbf{z}}_v(i)|}{\|\tilde{\mathbf{z}}_v\|_p} \right)^p \right)$$

$$= \frac{r_u^p}{\|\tilde{\mathbf{z}}_u\|_p^p} \sum_{i \in I_1} |\tilde{\mathbf{z}}_u(i)|^p + \frac{r_v^p}{c^p \|\tilde{\mathbf{z}}_v\|_p^p} \sum_{i \in I_1} |\tilde{\mathbf{z}}_v(i)|^p,$$

where, in the second step, we used the inequality $(a+b)^p \geq a^p + b^p$ for all $a, b > 0$ and $p \geq 1$. Now, we can apply tail bounds on the random variables in the above expression. We start with observing that the probability that $\tilde{z}_{ui}, \tilde{z}_{vi}$ have a opposite sign is exactly $1/2$. Hence, the set I_1 is a subset of component indices where each component is independently chosen with probability $1/2$. A Chernoff-Hoeffding bound (Theorem 7) therefore implies that for every $\varepsilon > 0$, with overwhelming probability,

$$\frac{1}{2}d(1 - \varepsilon) \leq |I_1| \leq \frac{1}{2}d(1 + \varepsilon).$$

We further note that the random variables $\|\tilde{\mathbf{z}}_u\|_p^p, \|\tilde{\mathbf{z}}_v\|_p^p$, and $\sum_{i \in I_1} |\tilde{\mathbf{z}}_u(i)|^p, \sum_{i \in I_1} |\tilde{\mathbf{z}}_v(i)|^p$ are i.i.d random variables from $\mathcal{X}^p(d)$ and $\mathcal{X}^p(|I_1|)$, respectively. Hence, Corollary 2 and Theorem 22, imply that for every $\varepsilon > 0$, with overwhelming probability,

$$(1 - \varepsilon)\frac{2d}{p} \leq \|\tilde{\mathbf{z}}_u\|_p^p, \|\tilde{\mathbf{z}}_v\|_p^p \leq (1 + \varepsilon)\frac{2d}{p}$$

---

2. Note that this is a stricter notion of what is commonly referred to as "with overwhelming probability" in literature.

and

$$(1 - \varepsilon)\frac{2|I_1|}{p} \leq \sum_{i \in I_1} |\tilde{\mathbf{z}}_u(i)|^p, \sum_{i \in I_1} |\tilde{\mathbf{z}}_v(i)|^p \leq (1 + \varepsilon)\frac{2|I_1|}{p}.$$

Moreover, we note that the probability $\Pr[r_u \geq \zeta] = 1 - \zeta^d$ for every $0 < \zeta < 1$, so we have $r_u, r_v \geq \zeta$ with overwhelming probability. In total, this implies that with overwhelming probability,

$$S_1 \geq \frac{\zeta^p}{(1 + \varepsilon)2d/p}\frac{1}{2}(1 - \varepsilon)\frac{2d}{p}(1 - \varepsilon) + \frac{\zeta^p}{c^p(1 + \varepsilon)2d/p}\frac{1}{2}(1 - \varepsilon)\frac{2d}{p}(1 - \varepsilon)$$

$$= \frac{\zeta^p(1 - \varepsilon)^2}{2(1 + \varepsilon)}\left(1 + \frac{1}{c^p}\right).$$

We note that by choosing ζ sufficiently large, and c and ε sufficiently small, we can push this lower bound to every number smaller than 1. That is, we have shown that that for every $\varepsilon' > 0$, there are constants $\zeta < 1, c > 1$ such that with overwhelming probability, $S_1 \geq 1 - \varepsilon'$.

We go on with lower bounding S_2. Analogously to I_1, let I_2 be the set of all component indices i in which y_{ui} and y_{vi} have the same sign. This implies that

$$\left|\mathbf{y}_u(u) - (w_v/w_u)^{1/d}\mathbf{y}_v(i)\right| = \left|\left|\mathbf{y}_u(i)\right| - (w_v/w_u)^{1/d}\left|\mathbf{y}_v(i)\right|\right|.$$

We can hence reformulate S_2 as

$$S_2 = \sum_{i \in I_2}\left|r_u\left|\frac{\tilde{\mathbf{z}}_u(i)}{\|\tilde{\mathbf{z}}_u\|}\right| - \left(\frac{w_v}{w_u}\right)^{1/d}r_v\left|\frac{\tilde{\mathbf{z}}_v(i)}{\|\tilde{\mathbf{z}}_v\|}\right|\right|^p$$

$$= \frac{r_u^p}{\|\tilde{\mathbf{z}}_u\|_p^p}\sum_{i \in I_2}\left|\left|\tilde{\mathbf{z}}_u(i)\right| - \left(\frac{w_v}{w_u}\right)^{1/d}\frac{r_v}{r_u}\frac{\|\tilde{\mathbf{z}}_u\|_p}{\|\tilde{\mathbf{z}}_v\|_p}\left|\tilde{\mathbf{z}}_v(i)\right|\right|^p.$$

We first note that, since $|I_2| = d - |I_1|$ and with overwhelming probability $|I_1| = \Theta_d(d)$, we have $|I_2| = \Theta_d(d)$ with overwhelming probability. Furthermore, we have with overwhelming probability that $r_u, r_v \geq \zeta$ and that both $\|\tilde{\mathbf{z}}_u\|_p^p$ and $\|\tilde{\mathbf{z}}_v\|_p^p$ are between $(1 - \varepsilon)2d/p$ and $(1 + \varepsilon)2d/p$ just like in the above paragraph. Together with $(w_v/w_u)^{1/d} \leq c$, this implies that with overwhelming probability,

$$\left(\frac{w_v}{w_u}\right)^{1/d}\frac{r_v}{r_u}\frac{\|\tilde{\mathbf{z}}_u\|_p}{\|\tilde{\mathbf{z}}_v\|_p} \leq \frac{c}{\zeta}\left(\frac{1 + \varepsilon}{1 - \varepsilon}\right)^{\frac{1}{p}}. \tag{8}$$

This bound can be made smaller than 2 by choosing c, ε small enough and ζ large enough. Furthermore, we get that for every $1 \leq i \leq d$ and any constant $\lambda > 0$, there is a constant probability of the event $\mathbf{E}_\lambda$ that $|\tilde{\mathbf{z}}_u(i)|$ is large enough and $|\tilde{\mathbf{z}}_v(i)|$ is small enough such that

$$\left|\left|\tilde{\mathbf{z}}_u(i)\right| - 2\left|\tilde{\mathbf{z}}_v(i)\right|\right|^p \geq \lambda$$

because $|\tilde{\mathbf{z}}_u(i)|$ and $|\tilde{\mathbf{z}}_v(i)|$ are two independent samples from $\chi_p(1)$[3]. Hence, the sum

$$\sum_{i \in I_2}\left|\left|\tilde{\mathbf{z}}_u(i)\right| - 2\left|\tilde{\mathbf{z}}_v(i)\right|\right|^p \tag{9}$$

---

3. This is the crucial step of this proof in which our coupling between $\mathbf{y}$ and $\tilde{\mathbf{z}}$ turns out to be useful. This is because the components of $\tilde{\mathbf{z}}$ are indeed independent whereas the components of $\mathbf{y}$ are not.

is with overwhelming probability lower bounded by the sum of $|I_2| = \Theta_d(d)$ independent Bernoulli random variables with constant success probability. Therefore, a Chernoff-Hoeffding bound (Theorem 7) implies that with overwhelming probability,

$$\sum_{i \in I_2} ||\tilde{\mathbf{z}}_{\mathbf{u}}(i)| - 2\,|\tilde{\mathbf{z}}_{\mathbf{v}}(i)||^p = \Omega_d(d). \tag{10}$$

As the bound from Equation (8) is with overwhelming probability smaller than 2 for appropriate choices of c, ε, ζ, we get that with overwhelming probability that

$$\sum_{i \in I_2} \left| |\tilde{\mathbf{z}}_u(i)| - \left(\frac{w_v}{w_u}\right)^{1/d} \frac{r_v}{r_u} \frac{\|\tilde{\mathbf{z}}_u\|_p}{\|\tilde{\mathbf{z}}_v\|_p} |\tilde{\mathbf{z}}_v(i)| \right|^p = \Omega(d).$$

As we further get that $r_u^p / \|\tilde{\mathbf{z}}_u\|_p^p = \mathcal{O}_d(1/d)$, with overwhelming probability, we have in total that $S_2 = \Omega_d(1)$ with overwhelming probability where the leading constant does not depend on c, ζ, ε.

In total, we get that for every $\varepsilon' > 0$, with overwhelming probability, $S_1 + S_2 \geq 1 - \varepsilon' + \Omega_d(1)$ if we choose c and ε sufficiently small and ζ sufficiently large (i.e. sufficiently close to 1). Hence, if we choose ε' small enough, there is a $c > 1$ such that with overwhelming probability, $S_1 + S_2 \geq c^p$. This implies our statement. ∎

This lemma directly implies our main result.

Theorem 1 *Asymptotically almost surely, if $d = o(\log(n))$, the clustering coefficient of G sampled from the GIRG model under some L_p-norm with $p \in [1, \infty]$ is*

$$\mathrm{CC}(G) = \exp(-\Omega_d(d)) + o(1).$$

Proof Similarly as in the proof of Theorem 3, we use Lemma 11 and Lemma 15 to conclude that a.a.s.,

$$\mathrm{CC}(G) = \mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] + o(1) \leq \Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right] + o(1)$$

where $\hat{v}, \hat{s}, \hat{t}$ are three random vertices from $G_{\leq n^{1/8}}$ conditioned on $w_{\hat{v}} \leq w_{\hat{s}}, w_{\hat{t}}$. To bound, the above probability, fix any constant $c > 1$ and note that

$$\Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right] \leq \Pr\left[\Delta \mid (\hat{v} \sim \hat{s}, \hat{t}) \cap (\max\{w_{\hat{s}}, w_{\hat{t}}\} \leq c^d w_{\hat{v}})\right]$$
$$+ \Pr\left[\max\{w_{\hat{s}}, w_{\hat{t}}\} \geq c^d w_{\hat{v}} \mid \hat{v} \sim \hat{s}, \hat{t}\right].$$

Now we let c be as in Lemma 25 and replace the first term by the bound derived in said lemma (Lemma 25). The second term can be bounded by $(2 + o(1))c^{d(2-\beta)}$ which follows from the distribution of $\max\{w_{\hat{s}}, w_{\hat{t}}\}$ as derived in the proof of Theorem 3. In total, this means that there are constants $a, b > 0, c > 1$ (from Lemma 25) such that

$$\Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right] \leq a \exp(-bd) + (2 + o(1))c^{d(2-\beta)} = \exp(-\Omega_d(d))$$

as desired. Note that the last step above holds since for sufficiently large d there is a constant δ such that the above term is upper bounded by $\exp(-\delta d)$, which concludes the proof. ∎

4. Testing for the dimensionality

We turn to the question of how one can recover the dimension d of $G(n, d, \beta, w_0)$ generated under L_∞-norm. We show that this is accomplished by a simple test statistic that computes a modified version of the clustering coefficient in a subgraph of G consisting of all vertices with weight in $[w_c, cw_c]$ for some constants $w_c \geq w_0, 1 < c < 2/\sqrt{3}$. We show that the value of this test statistic is well concentrated around its expectation such that it allows us to test whether G came from dimension d as long as $d = o(\log(n))$. As a side result, this shows that all graphs in the low-dimensional regime form a clear dichotomy.

Formally, we consider the following statistical testing problem. We are given a graph G on n vertices, its weight sequence, and an integer $d = o(\log(n))$. Under the null hypothesis, G is a GIRG generated in the weight sampling model with dimension d, whereas under the alternative hypothesis, G was generated in dimension $d_1 \neq d$ or it is a Chung–Lu graph. Here, we allow d_1 to be any integer (potentially larger than $\log(n)$). As a proof of concept, we propose the following testing procedure for this problem. Fix a constant $1 < c < 2/\sqrt{3}$ and a weight $w_c \geq w_0$. Now, consider the induced subgraph $\tilde{G}$ of G consisting of all vertices with weight in $[w_c, cw_c]$. For every vertex $v \in \tilde{G}$ that has at least two neighbors in $\tilde{G}$, we compute its local clustering coefficient $CC_{\tilde{G}}(v)$ and denote by $\mathrm{CC}^{(+)}(\tilde{G})$ the mean over all these values. We accept the null hypothesis if and only if condition (2) is met. We show that the probability that this test makes a mistake under both the null and alternative hypothesis goes to zero as $n \to \infty$ and we capture this in the following statment.

Theorem 5 *Let $G = G(n, d, \beta, w_0)$ be a GIRG generated under L_∞-norm. Let further $1 < c < 2/\sqrt{3} \approx 1.1547, w_c \geq w_0$ be constants, and let $\tilde{G}$ be the subgraph of G consisting of all vertices with weight in $[w_c, cw_c]$. Assume that d is an integer with $d = o(\log(n))$. Define the set S as the set of vertices in $\tilde{G}$ that have at least two neighbors in $\tilde{G}$ and the random variable $\mathrm{CC}^{(+)}(\tilde{G})$ as*

$$\mathrm{CC}^{(+)}(\tilde{G}) := \frac{1}{|S|} \sum_{v \in S} \frac{|\{\{s, t\} \in \Gamma(v) \mid s \sim t\}|}{\binom{\deg(v)}{2}} = \frac{1}{|S|} \sum_{v \in S} \mathrm{CC}_{\tilde{G}}(v).$$

Then,

$$\mathrm{CC}^{(+)}(\tilde{G}) \in \left(\frac{1}{c} \left(\frac{3}{4} \right)^d, c \left(\frac{3}{4} \right)^d \right) \pm n^{-1/5} \tag{2}$$

with probability at least $1 - 1/n$.

Proof We start by estimating the expectation of $\mathrm{CC}^{(+)}(\tilde{G})$. It is not hard to see that by linearity of expectation

$$\mathbb{E}\left[|S| \, \mathrm{CC}^{(+)}(\tilde{G}) \right] = \sum_{v \in G} \mathbb{E}\left[\mathbb{1}(v \in S) \, \mathrm{CC}_{\tilde{G}}(v) \right]$$

$$= \sum_{v \in G} \Pr\left[v \in S \right] \mathbb{E}\left[\mathrm{CC}_{\tilde{G}}(v) \mid v \in S \right]$$

$$= \mathbb{E}\left[|S| \right] \Pr\left[\Delta \mid v \sim s, t \right]$$

where v, s, t are three random vertices in $\tilde{G}$, and Δ is the event that v, s, t are a triangle. Notice that we used that $\mathbb{E}\left[\mathrm{CC}_{\tilde{G}}(v) \mid v \in S\right] = \Pr[\Delta \mid v \sim s, t]$ here as established (in the proof of) Lemma 15; equality holds here because conditioning on $v \in S$ is the same as conditioning on $|\Gamma(v)| \geq 2$.

Our proof now proceeds in two steps: (1) we show that $\Pr[\Delta \mid v \sim s, t]$ is in the interval $(c^{-1}(3/4)^d, c(3/4)^d)$, and (2) we show that $\mathrm{CC}^{(+)}(\tilde{G})$ concentrates around its expectation using the method of typical bounded differences.

For part (1), we apply Theorem 4 and note that – since v, s, t are in $\tilde{G}$ – the weights of v, s, t differ by at most a factor of c. However, Theorem 4 only yields a bound on the probability of Δ if v is the vertex of minimal weight among v, s, t. However, if we define $\hat{v}, \hat{s}, \hat{t}$ to be the vertices v, s, t reordered such that $\hat{v}$ is of minimal weight, we can express $\Pr[\Delta \mid v \sim s, t]$ as

$$\frac{\Pr[\Delta]}{\Pr[v \sim s, t]} = \frac{\Pr\left[\hat{v} \sim \hat{s}, \hat{t}\right]}{\Pr[v \sim s, t]} \Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right] = \frac{w_{\hat{v}}}{w_v} \cdot \Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right]$$

where the last equality holds because $\Pr[v \sim s, t] = \lambda^2 w_v^2 w_s w_t / n^2$ (we can ignore the minimum in eq. (1) here because the weights are constant). Since the fraction in the above equation is at least $1/c$ and at most 1, and since we can bound $\Pr\left[\Delta \mid \hat{v} \sim \hat{s}, \hat{t}\right]$ using Theorem 4, we conclude that

$$\Pr[\Delta \mid v \sim s, t] \in \left(\frac{1}{c}\left(\frac{3}{4}\right)^d, c\left(\frac{3}{4}\right)^d\right)$$

For the second part of the proof, we first show that $|S|$ is linear in n with high probability. Afterwards, we apply the same procedure to $|S|\,\mathrm{CC}^{(+)}(\tilde{G})$. We start by showing that there is a constant $\alpha > 0$ such that $|S|$ is at least αn with probability $1 - n^{-\omega(1)}$. Consider a fixed vertex v from G and denote the number of its neighbors in $\tilde{G}$ by X_v. We note that every vertex in G has a constant probability of being in $\tilde{G}$ and a probability of at least $\lambda w_0^2/n$ to connect to v. X_v is therefore lower bounded by the sum of n independent Bernoulli random variables with success probability in $\Theta(1/n)$. Denote this sum by $\tilde{X}_v$ and note that $\mathbb{E}[\tilde{X}_v] = \Theta(1)$. By (Cam, 1960, Proposition 1), the $\tilde{X}_v$ thus converges to a Poisson distributed random variable with constant expectation. Accordingly, $\Pr[\tilde{X}_v \geq 2]$ is constant as well. This shows that every vertex in G has at least a constant probability of having two neighbors in $\tilde{G}$. As the probability that v is in $\tilde{G}$ is constant as well, this implies that $\mathbb{E}[|S|] = \Omega(n)$. We continue with showing concentration of this random variable using Theorem 8. We note that the random variables $x_1, x_2, \ldots, x_n$ (the positions of all vertices), and $w_1, w_2, \ldots, w_n$ (the weights of all vertices) are independent and define a product probability space Ω such that each $\omega \in \Omega$ defines a graph $G(\omega)$ and a corresponding test graph $\tilde{G}(\omega)$. We further define $f(\omega)$ as the value of $|S|$ in $G(\omega)$ and consider the "bad" event

$$\mathcal{B} = \{\omega \in \Omega \mid \text{the max degree in } \tilde{G}(\omega) \text{ is greater than } \log^3(n)\}.$$

By Lemma 10, $\mathcal{B}$ happens with probability $n^{-\omega(1)}$ since all vertices in $\tilde{G}$ have at most constant weight. Now, let $\omega, \omega' \in \overline{\mathcal{B}}$ be such that they differ in at most two coordinates. Changing the weight or coordinate of one vertex can only decrease the number of vertices in $\tilde{G}$ with at least two neighbors by at most $2\log^3(n)$ as the weight or coordinate change only influences vertices that are neighbours of the changed vertex before or after the change. Accordingly, two coordinate or weight changes

33

can only change $|S|$ by at most $c' := 4\log^3(n)$. Using $t = n^{3/4}$ further satisfies the condition $t \geq 2M\Pr[\mathcal{B}]$ as $M \leq n$ and $\Pr[\mathcal{B}] = n^{-\omega(1)}$. As $m = 2n$, we get,

$$\Pr\left[|S| - \mathbb{E}[|S|]| \geq n^{3/4}\right] \leq 2\exp\left(-\frac{n^{1/2}}{32 \cdot 2 \cdot 16\log^6(n)}\right) + \left(\frac{n^2}{\log^3(n)} + 1\right)n^{-\omega(1)}$$
$$= n^{-\omega(1)}.$$

Similarly, we can show concentration of $f(S) = |S| \cdot \mathrm{CC}^{(+)}(\tilde{G}) = \sum_{v \in S} \mathrm{CC}_{\tilde{G}}(v)$. Again, changing the coordinate or weight of any two vertices can only increase or decrease the local clustering coefficient of at most $4\log^3(n)$ vertices by a value of at most one. Hence, we can again choose $c' := 4\log^3(n)$ and $t = n^{3/4}$ to obtain that

$$\Pr\left[|f(S) - \mathbb{E}[f(S)]| \geq n^{3/4}\right] \leq n^{-\omega(1)}.$$

Combining these two concentration results, we get that

$$f(S) = \mathbb{E}[|S|]\Pr[\Delta \mid v \sim s, t] \pm n^{3/4} \text{ and}$$
$$|S| = \mathbb{E}[|S|] \pm n^{3/4}$$

both hold with probability $1 - o(1/n)$. Dividing by $|S|$ and using $\mathbb{E}[|S|] = \Theta(n)$ then yield that

$$\frac{f(S)}{|S|} = \mathrm{CC}^{(+)}(\tilde{G}) = \frac{\mathbb{E}[|S|]}{\mathbb{E}[|S|] \mp n^{3/4}}\Pr[\Delta \mid v \sim s, t] \pm \frac{n^{3/4}}{\mathbb{E}[|S|] \mp n^{3/4}}$$
$$= \Pr[\Delta \mid v \sim s, t] \pm n^{-1/5}.$$

Using our estimate for $\Pr[\Delta \mid v \sim s, t]$ from part (1) of this proof concludes the argument. ∎

Using this, we immediately get that the probability that our test makes a mistake assuming that the null hypothesis is true is only $n^{-\omega(1)}$. Under the alternative hypothesis, assume that d_1 is the ground truth dimension G came from, and assume further without loss of generality that $d_1 \geq d+1$. We have to show that asymptotically,

$$\frac{1}{c}\left(\frac{3}{4}\right)^d - n^{-1/5} > c\left(\frac{3}{4}\right)^{d+1} + n^{-1/5}$$
$$\Leftrightarrow 1 > \frac{3}{4}c^2 + 2\left(\frac{4}{3}\right)^d n^{-1/5}.$$

As $c < 2/\sqrt{3}$ and $d = o(\log(n))$, this inequality is true for sufficiently large n. To see this, observe that

$$\left(\frac{4}{3}\right)^d n^{-1/5} = \exp\left(\ln\left(\frac{4}{3}\right)d - \frac{1}{5}\ln(n)\right) = o(1).$$

Acknowledgments

References

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.

Pedro Almagro, Marián Boguñá, and M Ángeles Serrano. Detecting the ultra low dimensionality of real networks. *Nature communications*, 13(1):6096, 2022.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in neural information processing systems*, 14, 2001.

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)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

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.

Marian Boguna, Fragkiskos Papadopoulos, and Dmitri Krioukov. Sustaining the internet with hyperbolic mapping. *Nature Communications*, 1(1):62, Dec 2010. ISSN 2041-1723. doi: 10.1038/ncomms1063. arXiv: 1009.0267.

Marian Boguna, Ivan Bonamassa, Manlio De Domenico, Shlomo Havlin, Dmitri Krioukov, and M Ángeles Serrano. Network geometry. *Nature Reviews Physics*, 3(2):114–135, 2021.

Matthew Brennan, Guy Bresler, and Dheeraj Nagaraj. Phase transitions for detecting latent geometry in random graphs. *Probability Theory and Related Fields*, 178(3-4):1215–1289, 2020.

Matthew Brennan, Guy Bresler, and Brice Huang. Threshold for detecting high dimensional geometry in anisotropic random geometric graphs. *arXiv preprint arXiv:2206.14896*, 2022.

Karl Bringmann, Ralph Keusch, and Johannes Lengler. Sampling Geometric Inhomogeneous Random Graphs in Linear Time. pages 20:1–20:15, 2017. ISSN 1868-8969. doi: 10.4230/LIPIcs.ESA.2017.20.

Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures and Algorithms*, 49(3):503–532, 2016.

Lucien Le Cam. An approximation theorem for the poisson binomial distribution. *Pacific Journal of Mathematics*, 10(4):1181–1197, Jan 1960. ISSN 0030-8730.

Francesco Camastra and Antonino Staiano. Intrinsic dimension estimation: Advances and open problems. *Information Sciences*, 328:26–41, 2016.

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, 2002a. doi: 10.1073/pnas.252631999.

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

Jesper Dall and Michael Christensen. Random geometric graphs. *Physical Review E*, 66(1):016121, Jul 2002. ISSN 1063-651X, 1095-3787. doi: 10.1103/PhysRevE.66.016121. arXiv:cond-mat/0203026.

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. ISSN 1083-6489, 1083-6489. doi: 10.1214/EJP.v16-967.

Ronen Eldan and Dan Mikulincer. Information and dimensionality of anisotropic random geometric graphs. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2017-2019 Volume I*, pages 273–324. Springer, 2020.

Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM computer communication review*, 29(4):251–262, 1999.

Tobias Friedrich, Thomas Sauerwald, and Alexandre Stauffer. Diameter and broadcast time of random geometric graphs in arbitrary dimensions. *Algorithmica*, 67:65–88, 2013.

Tobias Friedrich, Andreas Göbel, Maximilian Katzmann, and Leon Schiller. Cliques in high-dimensional geometric inhomogeneous random graphs. 2023. doi: 10.48550/ARXIV.2302.04113. URL https://arxiv.org/abs/2302.04113.

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.

Rishi Gupta, Tim Roughgarden, and Comandur Seshadhri. Decompositions of triangle-dense graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 471–482, 2014.

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.

Dmitri Krioukov. Clustering implies geometry in networks. *Physical review letters*, 116(20): 208302, 2016.

Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

Elizaveta Levina and Peter Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17, 2004.

Siqi Liu, Sidhanth Mohanty, Tselil Schramm, and Elizabeth Yang. Testing thresholds for high-dimensional sparse random geometric graphs. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 672–677, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392648. doi: 10.1145/3519935.3519989.

Suqi Liu and Miklós Z Rácz. Phase transition in noisy high-dimensional random geometric graphs. *arXiv preprint arXiv:2103.15249*, 2021.

Suqi Liu and Miklos Z Racz. A probabilistic view of latent space graphs and phase transitions. *arXiv preprint arXiv:2110.15886*, 2021.

George Livadiotis. Chi-p distribution: characterization of the goodness of the fitting using lpnorms. *Journal of Statistical Distributions and Applications*, 1(1):4, Jun 2014. ISSN 2195-5832. doi: 10.1186/2195-5832-1-4.

Colin McDiarmid et al. On the method of bounded differences. *Surveys in combinatorics*, 141(1): 148–188, 1989.

Riccardo Michielan, Nelly Litvak, and Clara Stegehuis. Detecting hyperbolic geometry in networks: Why triangles are not enough. *Physical Review E*, 106(5):054303, 2022.

J.M. Morel, P. Massart, J. Picard, F. Takens, and B. Teissier. *Concentration Inequalities and Model Selection:*. Concentration Inequalities and Model Selection: Ecole D'Eté de Probabilités de Saint-Flour XXXIII - 2003. Springer, 2007. ISBN 978-3-540-48497-4.

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30, 2017.

Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015. URL https://networkrepository.com.

Alireza Sarveniazi. An actual survey of dimensionality reduction. *American Journal of Computational Mathematics*, 2014, 2014.

Michael Spivak. *Calculus on manifolds: a modern approach to classical theorems of advanced calculus*. Mathematics monograph series. Perseus Books, Cambridge, Mass, 27. printing edition, 1998. ISBN 978-0-8053-9021-6.

Remco van der Hofstad, A. J. E. M. Janssen, Johan S. H. van Leeuwaarden, and Clara Stegehuis. Local clustering in scale-free networks with hidden variables. *Physical Review E*, 95(2):022307, Feb 2017. doi: 10.1103/PhysRevE.95.022307.

Lutz Warnke. On the method of typical bounded differences. *Combinatorics, Probability and Computing*, 25:269–299, 2016.

Zi Yin and Yuanyuan Shen. On the dimensionality of word embedding. *Advances in neural information processing systems*, 31, 2018.

Appendix A. Experimental Data

Table 1: The networks used in the plots in the first row of Figure 1 and their basic attributes. All found in the network repository Rossi and Ahmed (2015) and the SNAP dataset Leskovec and Krevl (2014).

Name	$|V|$	$|E|$	Category
ca-AstroPh	18.7 k	198.1 k	collaboration
ca-CondMat	23.1 k	93.4 k	collaboration
ca-GrQc	5.2 k	14.5 k	collaboration
ca-HepPh	12 k	118.5 k	collaboration
ca-MathSciNet	332.7 k	820.6 k	collaboration
cit-patent	3.7 M	16.5 M	citation
cit-HepTh	27.7 k	352.8 k	citation
cit-DBLP	12.6 k	49.7 k	citation
cit-HepPh	34.5 k	421.6 k	citation
fb-pages-artists	50.5 k	819.1 k	social
soc-academia	200.2 k	1.4 M	social
soc-youtube-snap	1.1 M	3 M	social
socfb-A-anon	3.1 M	23.7 M	social
email-Enron	36.7 k	183.8 k	social
bio-CE-CX	15.2 k	246 k	biological
bio-human-gene1	21.9 k	12.3 M	biological
bio-mouse-gene	43.1 k	14.5 M	biological
bio-WormNet-v3	16.3 k	762.8 k	biological
bio-grid-human	9.4 k	62.4 k	biological