Omitted Proofs of Paper 6377

A 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 $\Pr[\mathbf{E}] \geq 1 - O(1/n)$. 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 A.1 (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 A.2 (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 $X \coloneqq \sum_{i=1}^{k} X_i$. 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 A.3 (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*

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

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

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

B 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 B.1 (Lemma 3.3 and Lemma 3.4 in [Keusch, 2018], slightly reformulated)**.** *The following properties hold in $\mathcal{G}(n, d, \{w\}_1^n, \lambda) = (V, E)$.*

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

(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 4.1. *Let $\mathcal{G}(n, d, \{w\}_1^n, \lambda) = (V, E)$ be a GIRG and 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 B.1 that $\mathbb{E}[\deg(v)] = w_v$. Hence, for sufficiently large n, we get that $\log^3(n) \geq 2e\mathbb{E}[\deg(v)]$ 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 A.2 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. $\square$

C 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. 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 $V_{\leq \delta}$.

C.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 C.1. *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 C.2 (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 C.3 (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 C.4 (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 of Lemma C.1. 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 A.3 and note that the positions and weights of all vertices define a product probability space as in Theorem A.3. 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 C.3, 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 A.3 to obtain

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

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

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,

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

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 C.3, we further get that $\deg(v) = \mathcal{O}(w_v)$ for all $v \in V_{>n^{1/8}}$ and hence, by Lemma C.4 and Lemma C.2, 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. $\square$

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 C.5. *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*

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

Furthermore, let v_1, v_2, v_3 be the vertices v, s, t ordered increasingly by their weights. Then,

$$\mathbb{E}\left[\mathrm{CC}(G)\right] \leq \Pr\left[\Delta \mid v_1 \sim v_2, v_3\right].$$

Proof. We start by showing the first statement. Assume that $V = \{u_1, \dots, u_n\}$ 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 $\Gamma(v) = \{u_1, \dots, u_k\}$ 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 $k \geq 2$,

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

We proceed by showing that for any $1 \leq i < j \leq k$, we have $\mathbb{E}\left[\mathbb{1}(u_i \sim u_j) \mid \deg(v) = k\right] = \Pr\left[s \sim t \mid s, t \in \Gamma(v)\right]$. 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,

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

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

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

$$\begin{aligned}\Pr[v \sim s, t] &= \kappa_{vs}\kappa_{vt}/n^2 \\ &\geq \kappa_{v_1 v_2}\kappa_{v_1 v_3}/n^2 \\ &= \Pr[v_1 \sim v_2, v_3]\end{aligned}$$

because v_1 is the vertex of minimal weight and because the events $v_1 \sim v_2$ and $v_1 \sim v_3$ are independent. Thus,

$$\mathbb{E}[\mathrm{CC}_G(v)] \leq \Pr[\Delta \mid v \sim s, t] = \frac{\Pr[\Delta]}{\Pr[v \sim s, t]} \leq \frac{\Pr[\Delta]}{\Pr[v_1 \sim v_2, v_3]} = \Pr[\Delta \mid v_1 \sim v_2, v_3].$$

□

C.2 L_∞-norm

In this section, we analyze the clustering coefficient under L_∞-norm, which results in Theorem 3.3. To prove this theorem, we use Theorem 3.4 of [Friedrich *et al.*, 2023], which bounds the (more general) probability that the random set U_k is a clique if the ratio of the minimal and maximal weight among the vertices of U_k is at most c^d where $c > 1$ is an arbitrary constant. We show how to use this statement to bound the clustering in $G_{\leq n^{1/8}}$. For convenience, let us restate our result.

Theorem 3.3. *Asymptotically almost surely, if $d = o(\log(n))$, then the clustering coefficient of G sampled from the GIRG model with L_∞-norm fulfils*

$$\begin{aligned}\mathrm{CC}(G) &\leq 3\left(\frac{3}{4}\right)^{d\left(1-\frac{1}{\beta}\right)} + o(1) \\ &= \mathcal{O}_d\left(\left(\frac{3}{4}\right)^{d\left(1-\frac{1}{\beta}\right)}\right) + o(1).\end{aligned}$$

Proof. We use Lemma C.1 and thus only need an upper bound on $\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right]$. For this, we use Lemma C.5, and we let v, s, t be chosen u.a.r. from $G_{\leq n^{1/8}}$, and we let Δ be the event that v, s, t form a triangle. We get from Lemma C.5 that

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

Furthermore, by the second part of Lemma C.5, we may assume that v is of minimal weight among v, s, t. Accordingly, by Theorem 3.4, we may bound

$$\begin{aligned}\mathbb{E}\left[CC(G_{\leq n^{1/8}})\right] &\leq \Pr\left[\Delta \mid v \sim s, t \cap w_s, w_t \leq c^d w_v\right] + \Pr\left[\max\{w_s, w_t\} > c^d w_v\right] \\ &= \left(\frac{3c}{4}\right)^d + \Pr\left[\max\{w_s, w_t\} > c^d w_v\right].\end{aligned}$$

To bound the second term, recall that we assume $w_s, w_t \geq w_v$ and that v, s, t are vertices in $G_{\leq n^{1/8}}$. Therefore, given a value of w_v, the random variables w_s, w_t are independent and distributed as $\Pr[w_s \leq x \mid w_v] = \Pr[w_v \leq w \leq x] / \Pr\left[w_v \leq w \leq n^{1/8}\right]$ where w is a random variable following the standard Pareto distribution with parameters $w_0, 1-\beta,$. Hence, for all $w_0 \leq x \leq n^{1/8}$,

$$\begin{aligned}\Pr\left[\max\{w_s, w_t\} > c^d w_v \mid w_v = x\right] &= 1 - \left(\frac{\Pr\left[x \leq w \leq c^d x\right]}{\Pr\left[x \leq w \leq n^{1/8}\right]}\right)^2 \\ &\leq 1 - \left(\frac{\Pr\left[x \leq w \leq c^d x\right]}{\Pr[x \leq w]}\right)^2 \\ &= 1 - \left(\frac{(x/w_0)^{1-\beta} - (c^d x/w_0)^{1-\beta}}{(x/w_0)^{1-\beta}}\right)^2 \\ &= 1 - (1 - c^{d(1-\beta)})^2 \\ &= 1 - (1 - 2c^{d(1-\beta)} + c^{2d(1-\beta)}) \\ &= 2c^{d(1-\beta)} - c^{2d(1-\beta)} \leq 2c^{d(1-\beta)}.\end{aligned}$$

Therefore, by Lemma C.1, we obtain that asymptotically almost surely

$$\begin{aligned}\mathrm{CC}(G) &\leq \mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] + o(1)\\ &= \left(\frac{3c}{4}\right)^d + 2c^{d(1-\beta)} + o(1).\end{aligned}$$

Setting $c = (4/3)^{1/\beta}$, which minimizes the asymptotic behavior of the above term w.r.t. d, this yields

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

asymptotically almost surely ☐

C.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 3.1. *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 in $\mathbb{R}^d$ and afterwards use them to bound the clustering coefficient (Section C.3). We remark that, while the following statements are valid only in $\mathbb{R}^d$, we show in Section C.3 that they remain applicable for the analysis of random vectors in $\mathbb{T}^d$ that are distributed within balls under L_p-norm of sufficiently small radius. The reason for this is that random vectors within such balls behave like vectors in $\mathbb{R}^d$, so we can infer statements about their distribution by "scaling down" statements about the distibution of random vectors in unit balls in $\mathbb{R}^d$.

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 C.6 (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 C.7 (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 C.8. *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 $r \geq 0$,

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

Substituting $x_i = r \cdot y_i$ yields

$$\begin{aligned}\nu_S(r) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{1}(R \cdot (y_1, \dots, y_d) \in S(r)) r^d \mathrm{d}y_1 \dots \mathrm{d}y_d \\ &= r^d \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{1}((y_1, \dots, y_d) \in S(1)) \mathrm{d}y_1 \dots \mathrm{d}y_d \\ &= r^d \nu_S(1).\end{aligned}$$

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 C.9. *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 $S \subseteq \mathbb{R}^d_{>0}$. We note that we may express

$$\Pr[\mathbf{z} \in S] = \int_{\mathbb{R}^d_{>0}} \mathbb{1}_S(\mathbf{x}) \rho(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{2}$$

where $\mathbf{x} = (x_1, \dots, x_d)^T$. 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 $A_r = \{(r, x_2, \dots, x_d) \in \mathbb{R}^d_{>0} \mid \sum_{i=2}^d x_i^p < r^p\}$ and $A = \bigcup_{r>0} A_r$. Furthermore, we let

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

We note that this function is injective and that it has the remarkable property that for any $\mathbf{x} = (r, x_2, \dots, x_d) \in A$, $\|\varphi(\mathbf{x})\|_p = r$. 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 $A \subseteq \mathbb{R}^d_{>0}$, 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 (2) as

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

where $\mathbf{y} = (r, x_2, \dots, x_d)$ 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 $\mathbf{y} = (r, x_2, \dots, x_d) \in A$, 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

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

We define for any $r > 0$,

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

and thus obtain

$$\Pr[\mathbf{z} \in S] = \int_0^\infty \rho(r)v_S(r)\mathrm{d}r. \tag{3}$$

Now, recall that $\nu_S(R)$ is the volume of the set $S(R) = \{x \in \mathbb{R}^d \mid x/\|x\|_p \in S, \|x\|_p \leq R\}$. We show that in fact $v_S(R) = \frac{\mathrm{d}\nu_S(R)}{\mathrm{d}R}$ for all $R > 0$. This gives Equation (3) 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

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

where we used that for all $\mathbf{y} \in A$, we have $\mathbb{1}(\varphi(\mathbf{y}) \in S(R)) = \mathbb{1}(r \leq R)\mathbb{1}_S(\varphi(\mathbf{y}))$. 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 C.7.

Corollary C.10. *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 $\mathbf{z} \coloneqq \mathbf{x}/\|\mathbf{x}\|_p$. By Lemma C.9, 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 C.8 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,

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

We note that $\Pr\left[\mathbf{z}\in S_p(1)\right]=1$, and so, by Lemma C.9, we get

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

This shows

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

□

With this statement, we may now prove Lemma C.7.

Proof of Lemma C.7. 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 C.10 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 $\mathcal{X}^p$-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 $\mathcal{X}_p(d)$ and the $\mathcal{X}^p(d)$ distributions from the introduction. It is easy to see that a random vector $\mathbf{x}\sim\mathcal{X}_p(d)$ is L_p-symmetric by observing that its density function is

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

and thus only depends on the norm of $\mathbf{x}$. We further note that for the case $p=2$, $\mathcal{X}_2(d)$ is the standard d-variate normal distribution $\mathcal{N}(0,I_d)$ (where I_d is the $d\times d$ identity matrix), and that $\mathcal{X}^2(d)$ is the chi-squared distribution with d degrees of freedom. The distribution $\mathcal{X}^p(d)$ 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 $X\sim\mathcal{X}_p(1)$ 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 $\mathcal{X}^p(d)$ distribution and start with deriving its moment-generating function.

Lemma C.11. *Let $Z\sim\mathcal{X}^p(1)$. 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

$$\begin{aligned}\rho_Z(x) &= \frac{\mathrm{d}\Pr\left[Z \geq x\right]}{\mathrm{d}x} = \frac{\mathrm{d}\Pr\left[|X|^p \geq x\right]}{\mathrm{d}x} = \frac{\mathrm{d}\Pr\left[|X| \geq x^{\frac{1}{p}}\right]}{\mathrm{d}x} = \rho_{|X|}\left(x^{\frac{1}{p}}\right)\frac{\mathrm{d}x^{\frac{1}{p}}}{\mathrm{d}x} \\ &= 2\rho_X\left(x^{\frac{1}{p}}\right)\frac{\mathrm{d}x^{\frac{1}{p}}}{\mathrm{d}x} = 2\gamma e^{-\frac{1}{2}x}\frac{1}{p}x^{\frac{1}{p}-1} = \frac{x^{\frac{1}{p}-1}e^{-\frac{1}{2}x}}{2^{\frac{1}{p}}\Gamma\left(\frac{1}{p}\right)}.\end{aligned}$$

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

$$\begin{aligned}\psi_Z(\lambda) = \mathbb{E}\left[e^{\lambda Z}\right] &= \int_0^\infty \rho_Z(x)e^{\lambda x}dx \\ &= \frac{1}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\int_0^\infty x^{\frac{1}{p}-1}e^{-x(1/2-\lambda)}dx.\end{aligned}$$

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

$$\begin{aligned}\psi_Z(\lambda) &= \frac{1}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\int_0^\infty x^{\frac{1}{p}-1}e^{-x(1/2-\lambda)}dx \\ &= \frac{1}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\int_0^\infty y^{\frac{1}{p}-1}e^{-y}\frac{(1/2-\lambda)^{1-\frac{1}{p}}}{1/2-\lambda}dy \\ &= \frac{(1/2-\lambda)^{-\frac{1}{p}}}{2^{\frac{1}{p}}\Gamma\left(1/p\right)}\Gamma\left(1/p\right) \\ &= (1-2\lambda)^{-\frac{1}{p}}.\end{aligned}$$

□

Corollary C.12. *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 C.11 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 $Z \sim \chi^p(d)$ 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 C.13 (Proposition 2.9 in [Morel *et al.*, 2007])**.** *Let $X_1, \ldots, X_d$ 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 $S = \sum_{i=1}^d (X_i - \mathbb{E}\left[X_i\right])$. 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 C.14. *Let $X_1, \dots, X_d$ be i.i.d. random variables from $\mathcal{X}_p(1)$ and define the random variable $Z \coloneqq \sum_{i=1}^{d} |X_i|^p$. Note that $Z \sim \mathcal{X}^p(d)$. Then, for all $x > 0$,*

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

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

Proof. We use Theorem C.13. To show that the random variables $|X_1|^p, \dots, |X_d|^p$ fulfill the conditions of Theorem C.13, 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 C.11 to derive bounds on the moments of Y. For all integers $k \geq 0$, we note that we have $\mathbb{E}\left[Y^k\right] = \psi_Y^{(k)}(0)$, 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 $\mathbb{E}\left[Y\right] = \frac{2}{p}$ and $\mathbb{E}\left[Y^2\right] = \frac{4}{p}\left(\frac{1}{p} + 1\right)$. For $k \geq 3$ one can easily verify that

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

and hence,

$$\begin{aligned}
\mathbb{E}\left[Y^k\right] = \psi_Y^{(k)}(0) &= \mathbb{E}\left[Y^2\right]2^{k-2}\prod_{i=2}^{k-1}\left(\frac{1}{p} + i\right) \\
&= \mathbb{E}\left[Y^2\right]2^{k-2}\prod_{i=1}^{k-2}\left(\frac{1}{p} + i + 1\right) \\
&\leq \mathbb{E}\left[Y^2\right]2^{k-2}\prod_{i=1}^{k-2}(i+2) \\
&= \mathbb{E}\left[Y^2\right]2^{k-2}\frac{k!}{3!} \leq \mathbb{E}\left[Y^2\right]2^{k-1}\frac{k!}{2}. \qquad (4)
\end{aligned}$$

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 C.13 are fulfilled. Since $Z = \sum_{i=1}^{d} Y_i$ and $\mathbb{E}\left[Z\right] = \frac{2d}{p}$, we get that for all $x > 0$,

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

which shows the first statement.

For the second statement, we define $Y'_i \coloneqq -Y_i$ 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 (4) that

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

Hence, it follows from Theorem C.13 that

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

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 A.2.

Corollary 3.2. *Let $X_i, \ldots X_d$ be i.i.d. random variables from $\chi_p(1)$ and define $Z = \sum_{i=1}^{k} |X_i|^p \sim \chi^p(d)$. Then, for every $\varepsilon > 0$ and $\delta > 0$ defined by $\varepsilon = \sqrt{2\delta} + \delta$, it holds that*

$$\Pr\left[Z \geq (1+\varepsilon)\mathbb{E}[Z]\right] \leq \exp\left(-\frac{2\delta}{p} \cdot d\right)$$

and

$$\Pr\left[Z \leq (1-\varepsilon)\mathbb{E}[Z]\right] \leq \exp\left(-\frac{2\delta}{p} \cdot d\right).$$

Proof. We use Theorem C.14 and set $x = \delta\mathbb{E}[Z]$. We then obtain

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

Recalling from Theorem C.12 that $\mathbb{E}[Z] = \frac{2d}{p}$ then implies the first statement. The argumentation for the second statement is analogous. □

Bounding the clustering coefficient

Lemma C.16. *Let $G = G(n, d, \beta, w_0)$ be sampled under L_p-norm and let s, u, v be vertices chosen uniformly at random from $G_{\leq n^{1/8}}$ with $w_s \leq w_u, w_v$. There are constants $a, b > 0, c > 1$ such that for sufficiently large n and all $d \geq 1$,*

$$\Pr\left[\Delta \mid (s \sim u, v) \cap (w_u, w_v \leq c^d w_s)\right] \leq a \cdot \exp(-bd).$$

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 $r \leq 1/4$. 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 $\Pr\left[\|\mathbf{x}_u - \mathbf{x}_v\|_p \leq t_{uv}\right]$. 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 C.8), and since we choose t_{uv} such that $\nu(t_{uv}) = \lambda w_u w_v/n$, we get

$$t_{uv} = \left(\frac{\lambda w_u w_v}{\nu(1) n}\right)^{1/d}. \tag{5}$$

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

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

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 $w_v, w_s \in [w_s, c^d \cdot w_s]$, 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 C.7 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 $0 \leq R \leq 1$, we have

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

and thus,

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

Furthermore, due to the L_p-symmetry of $\mathbf{y_u}, \mathbf{y_v}$ and Lemma C.7, we assume that $\mathbf{z} = \tilde{\mathbf{z}}/\|\tilde{\mathbf{z}}\|_p$ where $\tilde{\mathbf{z}}$ is a random vector from the $\chi_p(d)$-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 $\tilde{\mathbf{z}}_u, \tilde{\mathbf{z}}_v \sim \chi_p(d)$.

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

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

where S_1 is the sum of all components in which $\mathbf{y}_{ui}$ and $\mathbf{y}_{vi}$ 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*[3] 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 $\mathbf{y}_{ui}$ and $\mathbf{y}_{vi}$ have opposite sign. Note that this implies that the term $\left|\mathbf{y}_{ui} - (w_v/w_u)^{1/d}\mathbf{y}_{vi}\right|$ is equal to $|\mathbf{y}_{ui}| + (w_v/w_u)^{1/d}\,|\mathbf{y}_{vi}|$. Furthermore, note that we may express $\mathbf{y}_{ui} = r_u \cdot \tilde{\mathbf{z}}_{ui}/\left\|\tilde{\mathbf{z}}_u\right\|_p$. 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

$$\begin{aligned}
S_1 &= \sum_{i \in I_1}\left(r_u\left|\frac{\tilde{\mathbf{z}}_{ui}}{\|\tilde{\mathbf{z}}_u\|_p}\right| + \left(\frac{w_v}{w_u}\right)^{1/d} r_v\left|\frac{\tilde{\mathbf{z}}_{vi}}{\|\tilde{\mathbf{z}}_v\|_p}\right|\right)^p \\
&\geq \sum_{i \in I_1}\left(\left(r_u\frac{|\tilde{\mathbf{z}}_{ui}|}{\|\tilde{\mathbf{z}}_u\|_p}\right)^p + \left(\frac{r_v}{c}\frac{|\tilde{\mathbf{z}}_{vi}|}{\|\tilde{\mathbf{z}}_v\|_p}\right)^p\right) \\
&= \frac{r_u^p}{\|\tilde{\mathbf{z}}_u\|_p^p}\sum_{i \in I_1}|\tilde{\mathbf{z}}_{ui}|^p + \frac{r_v^p}{c^p\|\tilde{\mathbf{z}}_v\|_p^p}\sum_{i \in I_1}|\tilde{\mathbf{z}}_{vi}|^p,
\end{aligned}$$

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

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 $\tilde{\mathbf{z}}_{ui}, \tilde{\mathbf{z}}_{vi}$ 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 A.2) 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 $\sum_{i\in I_1} |\tilde{\mathbf{z}}_{ui}|^p, \sum_{i\in I_1} |\tilde{\mathbf{z}}_{vi}|^p$ are i.i.d random variables from $\mathcal{X}^p(d)$ and $\mathcal{X}^p(|I_1|)$, respectively. Hence, Corollary 3.2 and Theorem C.12, imply that for every $\varepsilon > 0$, with overwhelming probability,

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

and

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

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

$$\begin{aligned} S_1 &\geq \frac{\delta^p}{(1+\varepsilon)2d/p}\frac{1}{2}(1-\varepsilon)^2 2d/p + \frac{\delta^p}{c^p(1+\varepsilon)2d/p}\frac{1}{2}(1-\varepsilon)^2 2d/p \\ &= \frac{\delta^p(1-\varepsilon)^2}{2(1+\varepsilon)}\left(1+\frac{1}{c^p}\right). \end{aligned}$$

We note that by choosing δ 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 $\delta < 1, c > 1$ 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 $\mathbf{y}_{ui}$ and $\mathbf{y}_{vi}$ have the same sign. This implies that $\left|\mathbf{y}_{ui} - (w_v/w_u)^{1/d}\mathbf{y}_{vi}\right| = \left||\mathbf{y}_{ui}| - (w_v/w_u)^{1/d}|\mathbf{y}_{vi}|\right|$. We can hence reformulate S_2 as

$$\begin{aligned} S_2 &= \sum_{i\in I_2}\left| r_u \left|\frac{\tilde{\mathbf{z}}_{ui}}{||\tilde{\mathbf{z}}_u||}\right| - \left(\frac{w_v}{w_u}\right)^{1/d} r_v \left|\frac{\tilde{\mathbf{z}}_{vi}}{||\tilde{\mathbf{z}}_v||}\right|\right|^p \\ &= \frac{r_u^p}{\|\tilde{\mathbf{z}}_u\|_p^p}\sum_{i\in I_2}\left||\tilde{\mathbf{z}}_{ui}| - \left(\frac{w_v}{w_u}\right)^{1/d}\frac{r_v}{r_u}\frac{\|\tilde{\mathbf{z}}_u\|_p}{\|\tilde{\mathbf{z}}_v\|_p}|\tilde{\mathbf{z}}_{vi}|\right|^p. \end{aligned}$$

We first note that, since $|I_2| = d - |I_1|$ and with overwhelming probability $|I_1| = \Theta(d)$, we have $|I_2| = \Theta(d)$ with overwhelming probability. Furthermore, we have with overwhelming probability that $r_u, r_v \geq \delta$ 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,

$$\left(\frac{w_v}{w_u}\right)^{1/d}\frac{r_v}{r_u}\frac{\|\tilde{\mathbf{z}}_u\|_p}{\|\tilde{\mathbf{z}}_v\|_p} \leq \frac{c}{\delta}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{\frac{1}{p}}. \tag{6}$$

This bound can be made smaller than 2 by choosing c, ε small enough and δ 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 $|\tilde{\mathbf{z}}_{ui}|$ is large enough and $|\tilde{\mathbf{z}}_{vi}|$ is small enough such that

$$||\tilde{\mathbf{z}}_{ui}| - 2|\tilde{\mathbf{z}}_{vi}||^p \geq \lambda$$

because $|\tilde{\mathbf{z}}_{ui}|$ and $|\tilde{\mathbf{z}}_{vi}|$ are two independent samples from $\mathcal{X}_p(1)$. Hence, the sum

$$\sum_{i\in I_2} ||\tilde{\mathbf{z}}_{ui}| - 2|\tilde{\mathbf{z}}_{vi}||^p \tag{7}$$

is with overwhelming probability lower bounded by the sum of $|I_2| = \Theta(d)$ independent Bernoulli random variables with constant success probability. Therefore, a Chernoff-Hoeffding bound (Theorem A.2) implies that with overwhelming probability,

$$\sum_{i\in I_2} ||\tilde{\mathbf{z}}_{ui}| - 2|\tilde{\mathbf{z}}_{vi}||^p = \Omega(d). \tag{8}$$

As the bound from Equation (6) is with overwhelming probability smaller than 2 for appropriate choices of c, ε, δ, we get that with overwhelming probability,

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

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

In total, we get that for every $\varepsilon' > 0$, with overwhelming probability, $S_1 + S_2 \geq 1 - \varepsilon' + \Omega(1)$ if we choose c and ε sufficiently small and δ sufficiently large. Hence, if we choose ε' small enough such that $1 - \varepsilon' + \Omega(1) > 1$, 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 first main result.

Theorem 3.1. *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 3.3, we get from Lemma C.16 that there are constants $c > 1, a, b > 0$ such that

$$\mathbb{E}\left[\mathrm{CC}(G_{\leq n^{1/8}})\right] \leq a \exp(-bd) + 2c^{d(1-\beta)}.$$

By Lemma C.1, this implies that asymptotically almost surely,

$$\mathrm{CC}(G) \leq a \exp(-bd) + 2c^{d(1-\beta)} + o(1) = \exp(-\Omega_d(d)).$$

Note that the last step holds for 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. □