

SUB-TREE COUNTS ON HYPERBOLIC RANDOM GEOMETRIC GRAPHS

TAKASHI OWADA,^{*} *Purdue University*
D. YOGESHWARAN ,^{**} *Indian Statistical Institute*

Abstract

The hyperbolic random geometric graph was introduced by Krioukov *et al.* (*Phys. Rev. E* **82**, 2010). Among many equivalent models for the hyperbolic space, we study the d -dimensional Poincaré ball ($d \geq 2$), with a general connectivity radius. While many phase transitions are known for the expectation asymptotics of certain subgraph counts, very little is known about the second-order results. Two of the distinguishing characteristics of geometric graphs on the hyperbolic space are the presence of tree-like hierarchical structures and the power-law behaviour of the degree distribution. We aim to reveal such characteristics in detail by investigating the behaviour of sub-tree counts. We show multiple phase transitions for expectation and variance in the resulting hyperbolic geometric graph. In particular, the expectation and variance of the sub-tree counts exhibit an intricate dependence on the degree sequence of the tree under consideration. Additionally, unlike the thermodynamic regime of the Euclidean random geometric graph, the expectation and variance may exhibit different growth rates, which is indicative of power-law behaviour. Finally, we also prove a normal approximation for sub-tree counts using the Malliavin–Stein method of Last *et al.* (*Prob. Theory Relat. Fields* **165**, 2016), along with the Palm calculus for Poisson point processes.

Keywords: Hyperbolic spaces; random geometric graphs; Poisson point process; sub-tree counts; central limit theorem; Malliavin–Stein method

2020 Mathematics Subject Classification: Primary 60F05; 60D05
Secondary 05C80; 51M10

1. Introduction

In this article, we study random geometric graphs on the d -dimensional Poincaré ball, a canonical model for negatively curved spaces in geometry [15, 41]. Random geometric graphs were first studied in the 1960s, beginning with [24], as a model of radio communications. Since then, they have been a thriving topic of research in probability, statistical physics, and wireless networks (see [35, 38, 42, 3, 4, 26]). In its simplest form, the random geometric graph can be constructed by taking a set of independent and identically distributed (i.i.d.) points $\mathcal{X}_n = \{X_1, \dots, X_n\}$ on a metric space as its vertex set, and placing an edge between any two distinct points within a distance r_n . Most of the studies on such graphs assume that the underlying metric space is Euclidean or a compact and convex Euclidean subset. However,

Received 25 December 2020; revision received 3 January 2022.

^{*} Postal address: Department of Statistics, Purdue University, West Lafayette, IN 47907, USA. Email address: owada@purdue.edu

^{**} Postal address: Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 8th Mile, Mysore Road, Bangalore, 560059, INDIA. Email address: d.yogesh@isibang.ac.in

© The Author(s), 2022. Published by Cambridge University Press on behalf of Applied Probability Trust.

various applications, especially those in topological data analysis, necessitate more general metric spaces. For example, the extension to a compact manifold without boundary has recently been investigated in [8, 9, 18]. A crude one-line summary of these studies is that the behaviour of the graph on a ‘nice’ d -dimensional manifold is similar to that on a d -dimensional Euclidean space, though the proofs are technically more challenging and boundary effects need to be accounted for carefully.

Going beyond the class of nice d -dimensional manifolds with positive curvature and the Euclidean space with zero curvature, we focus on ‘negatively’ curved spaces in this paper. Such an investigation on hyperbolic spaces was initiated in [30]. Among many equivalent models for the hyperbolic space, we focus our attention on the widely used d -dimensional Poincaré ball. Apart from mathematical curiosity, a good reason for investigating hyperbolic geometric graphs lies in their properties such as sparsity, power-law degree distribution, small-world phenomena, and clustering. Such properties are often observed in complex networks. For more details, see the introductions in [30, 25, 20], where the graph is often referred to as the disk model or the KPKVB model after the authors of [30]. However, we always call it the hyperbolic random geometric graph. In a broader sense, our work is an addition to the developing literature on random structures of hyperbolic spaces (see [12, 6, 5, 34, 31, 40, 17, 21]).

The rest of the article is organized as follows. Sections 1.1 and 1.2 provide simulations on hyperbolic geometric graphs, and a preview of our results and their implications. In Section 2, we introduce our setup in detail and state all our results. We compare our results with those in the existing literature in Section 2.4, and Section 2.5 sketches results for analogous models in the Euclidean case. This is followed by the proofs in Section 3, where we introduce basic lemmas on the hyperbolic metric and measure and also the abstract normal approximation bound or second-order Poincaré inequality from [32] needed for our central limit theorem (CLT).

1.1. Simulations on hyperbolic geometric graphs

We begin by introducing the Poincaré ball and the hyperbolic random geometric graph. Though there are other models of hyperbolic spaces, they are all isometric to the Poincaré ball (see [15, Section 7]); therefore, our results are essentially independent of the model selection. The d -dimensional Poincaré ball, denoted by

$$B_d^{(\zeta)} := \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_1^2 + \dots + x_d^2 < 1\} \quad (1)$$

with negative curvature $-\zeta^2$, is a d -dimensional open unit ball equipped with the Riemannian metric

$$ds^2 := \frac{4}{\zeta^2} \frac{|dx|^2}{(1 - |x|^2)^2} = \frac{4}{\zeta^2} \frac{dx_1^2 + \dots + dx_d^2}{(1 - x_1^2 - \dots - x_d^2)^2}, \quad (2)$$

where $|\cdot|$ denotes the Euclidean norm. Then, the hyperbolic distance between the origin (the same as the origin in the Euclidean plane) and $x \in B_d^{(\zeta)}$ is given by $d_\zeta(0, x) = \zeta^{-1} \log((1 + |x|)/(1 - |x|))$, where d_ζ is the hyperbolic distance induced by (2). Though $B_d^{(\zeta)}$ is topologically the same as any open Euclidean ball, what matters to us is the metric itself. In the notation of (2), the Euclidean metric can be specified as $ds^2 = dx_1^2 + \dots + dx_d^2$, and so, on compact sets of the unit ball, the two metric elements differ only by a constant. This ensures that various geometric ingredients in the Euclidean case can be carried over to the Poincaré disk when restricted to a compact set of the unit ball. However, as $|x| \uparrow 1$ in (2), one would expect that the geometry of the Poincaré disk significantly differs from that of the Euclidean disk; see Figure 1

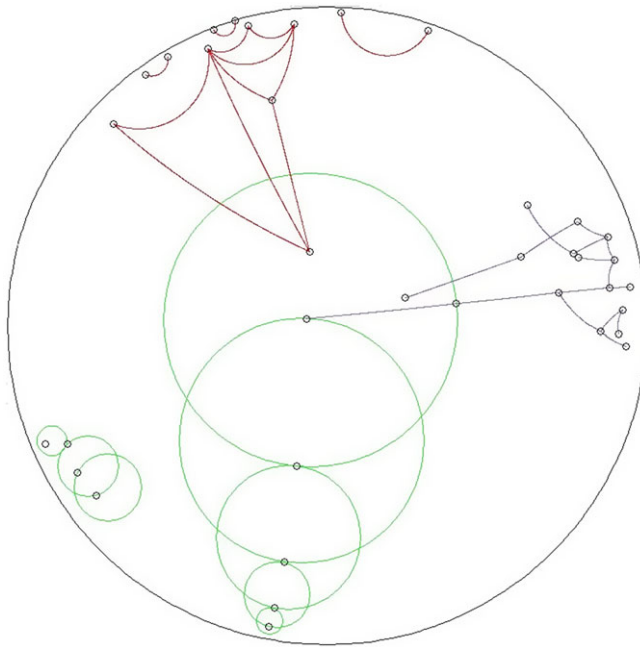


FIGURE 1. Geodesic line segments and triangles (red lines), geodesic line segments of unit length (violet lines), and unit circles with centres on the Poincaré disk (green circles) with $\zeta = 1$. These figures were drawn using the applet NonEuclid [16].

for hyperbolic lines and circles on the Poincaré disk with $d = 2$. The reader may refer to [41] for a rigorous discussion of basic features of hyperbolic spaces and geometry.

Let $B(0, R)$ denote the hyperbolic ball of radius R centred at the origin. In addition to the curvature parameter $-\zeta^2$ in (2), our model involves a second curvature parameter $-\alpha^2$ of another d -dimensional Poincaré ball $B_d^{(\alpha)}$. We first choose a sequence of radii $R_n \rightarrow \infty$ as $n \rightarrow \infty$ and select $N_n \stackrel{d}{=} \text{Poisson}(n)$ i.i.d. ‘uniform’ points X_1, \dots, X_{N_n} in $B(0, R_n) \subset B_d^{(\alpha)}$. Here, we mean ‘uniform’ with respect to the hyperbolic volume measure with curvature $-\alpha^2$. We then represent X_i in the hyperbolic polar coordinates as (r_i, Θ_i) , where r_i is the hyperbolic distance with respect to the curvature $-\alpha^2$, and $\Theta_i \in [0, \pi]^{d-2} \times [0, 2\pi)$ is a uniform random vector specifying angles. Next, we consider $X_i = (r_i, \Theta_i)$ as the points of $B_d^{(\zeta)}$ by keeping their polar coordinates fixed, and connect any two points X_i, X_j if $0 < d_\zeta(X_i, X_j) < R_n$. In other words, we sample points uniformly in a growing ball in $B_d^{(\alpha)}$ with respect to the hyperbolic volume measure with curvature $-\alpha^2$, and form a geometric graph on $B_d^{(\zeta)}$ with respect to the hyperbolic metric with curvature $-\zeta^2$. We denote the resulting graph by $HG_n(R_n; \alpha, \zeta)$, for which four parameters are involved: the dimension d , two curvature parameters α and ζ , and the radii regime R_n . As will be clarified in our later analyses, by choosing two distinct curvature parameters, one can construct a larger parametric family of random graphs with a richer set of features. We refer the reader to Remark 1 for a discussion about the dependence of our results on the individual parameters, not just their ratio, unlike many earlier results. After Definition 1, we briefly remark on the relationship of this model to the random geometric graph on a stationary Poisson process.

Figure 2 presents five simulation results with $d = 2$, $n = 1000$, $\zeta = 1$, $R_n = 2 \log 1000 = 13.82$, and different choices of α . It is useful to keep in mind that for a small α , the space is closer to the Euclidean space, in the sense that more points are scattered near the centre. For a large α , the points near the boundary mainly dominate the asymptotics. Furthermore, by the geometry of a hyperbolic space, the points near the centre easily connect to many other points, and so the presence of such points will change the connectivity structure. Additionally, we may notice a phase transition in the connectivity at $\alpha = 1$, due to the appearance of points closer to the centre. Some of the above observations will be crystallized later as rigorous theorems, while many more still await exploration in future works. We now preview some of our results and place them in the context of the growing literature on hyperbolic random geometric graphs.

1.2. Tree-like hierarchical structures: a preview of our results and some implications

One of the main characteristics of hyperbolic random geometric graphs is the presence of *tree-like hierarchical structures*, implying that the vertices on $B_d^{(\zeta)}$ are classified into large groups of smaller subgroups, which themselves consist of further smaller subgroups (see [30]). This is reflected in our simulations, illustrating that, as compared to the Euclidean spaces, the non-amenability of a negatively curved space enables one to embed many more trees in the space.

The study of the hyperbolic random geometric graph started with [30]. Since then, various features of the resulting graph have been investigated intensively by many authors. Indeed, there are a number of publications on the subject, in areas including the structure of large components [10, 22, 29], the probability of connectivity [11], clique-counts [7], the degree sequence and clustering [14, 25, 21], the bootstrap percolation process [13], the typical distances and diameter of components [1, 27, 36], and the spectral properties of the graph [28]. More recently, [23] has proved expectation and variance asymptotics as well as CLTs for the number of isolated and extreme points for the hyperbolic random geometric graph.

The earlier studies cited above, however, treat only a special case, namely,

$$d = 2 \text{ and } R_n = 2\zeta^{-1} \log(n/\nu) \quad (3)$$

for some $\nu \in (0, \infty)$, or similar variants. In contrast, the current study assumes a set of conditions more general than (3). For example, we allow for a general dimension d , and we do not put any restrictive conditions on R_n as in (3). Under this general setup, the primary objective of this paper is to uncover the spatial distribution of the tree-like structure via sub-tree counts, i.e., the number of copies of a given tree in $HG_n(R_n; \alpha, \zeta)$. In graph-theoretic language, we count the number of injective graph homomorphisms from a fixed sub-tree to $HG_n(R_n; \alpha, \zeta)$. Our main results in Theorems 1, 2, and 3 establish expectation asymptotics, variance asymptotics, and the normal approximation bounds for the sub-tree counts. All of our asymptotic results crucially depend on a degree sequence of the tree under consideration as well as the two curvature parameters $-\zeta^2$ and $-\alpha^2$.

More specifically, the present paper studies a more general sub-tree count, denoted by $S_n^{(\gamma)}$, $\gamma \in (0, 1]$, which counts sub-trees only on the space $B(0, R_n) \setminus \overset{\circ}{B}(0, (1 - \gamma)R_n)$. In other words, $S_n^{(\gamma)}$ counts sub-trees in an annulus with inner radius $(1 - \gamma)R_n$ and outer radius R_n . In particular, $S_n := S_n^{(1)}$ denotes the sub-tree counts for the ‘entire’ space $B(0, R_n)$. Our results will reveal that if α is sufficiently large, i.e., the space $B_d^{(\alpha)}$ is sufficiently hyperbolic (for example, $\alpha = 1.2$ in Figure 2), then the asymptotics of $S_n^{(\gamma)}$ does not depend on γ . In other words, the

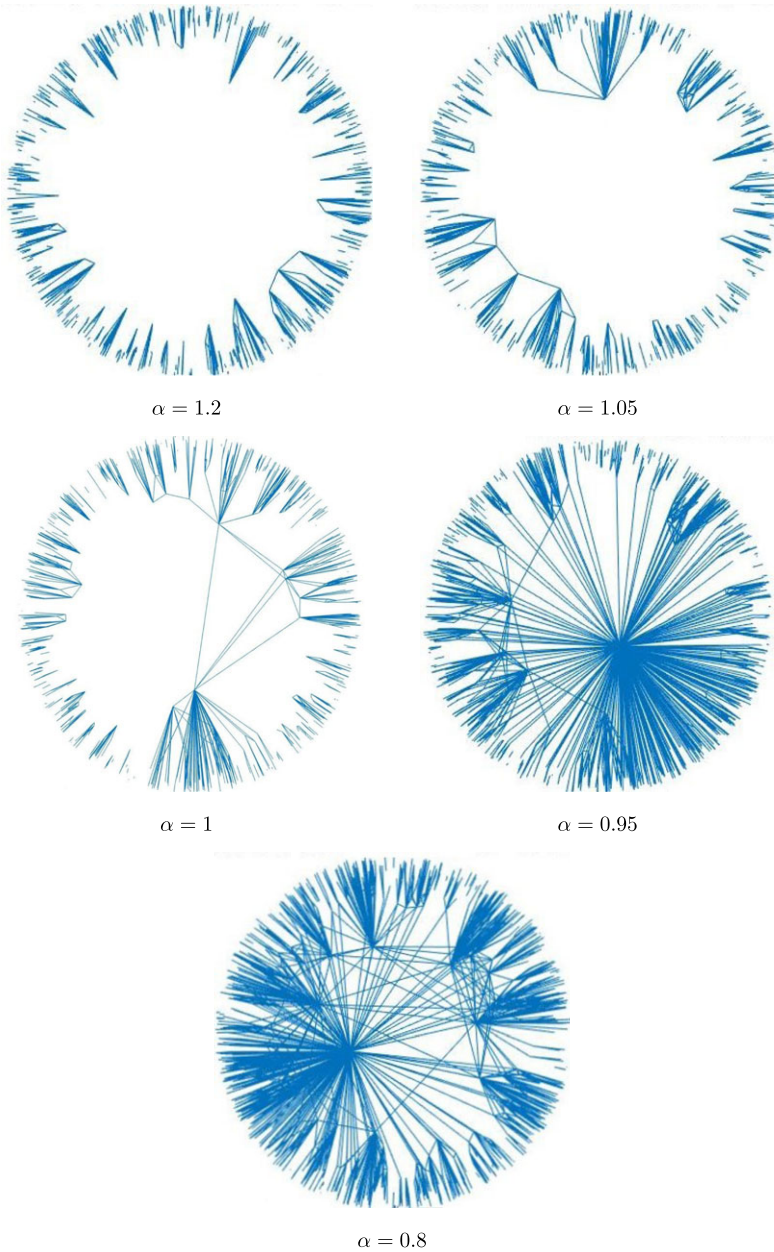


FIGURE 2. Simulations of $HG_{1000}(R_{1000}; \alpha, 1)$ for $d = 2$ with $R_{1000} = 2 \log 1000 = 13.82$ and different values of α . Isolated vertices have been omitted.

spatial distribution of sub-trees is mostly determined by those near the boundary of $B(0, R_n)$. In this case, S_n is well approximated by $S_n^{(\gamma)}$ for any $\gamma \in (0, 1)$. On the other hand, if α is small, as in Figure 2 with $\alpha = 0.8$, the distribution of sub-trees will be affected by those near the centre of the space, and $S_n^{(\gamma)}$ no longer approximates S_n . To relate the above phenomena

TABLE 1. Summary of related results for $d = 2$, $\zeta = 1$, $R_n = 2 \log(n/\nu)$, $\nu \in (0, \infty)$. Here K_k denotes the number of k -cliques in $HG_n(R_n; \alpha, 1)$, $\mathbb{P}_{conn} = \mathbb{P}(HG_n(R_n; \alpha, 1) \text{ is connected})$, and $\mathbb{P}_{perc} = \mathbb{P}(HG_n(R_n; \alpha, 1) \text{ percolates})$, where, by percolation, we mean the existence of a giant component, i.e., a component of size $\Theta(n)$.

Regime/Properties	\mathbb{P}_{conn}	\mathbb{P}_{perc}	$\mathbb{E}(K_k), k \geq 2$	$\mathbb{E}(S_n^{(\gamma)}), \gamma \in (0, 1) \setminus \{1/2\}$
Results from	[11]	[10]	[7], [14]	Corollary 2
$\alpha < 1/2$	1	1	Not Known	$\sim n^{1+2\gamma(k-1-k\alpha)}$
$1/2 < \alpha \leq 1 - 1/k$	0	1	$\Theta(n^{(1-\alpha)k})$	$\sim n^{1+\gamma \sum_{i=1}^k (d_i - 2\alpha)_+}$
$1 - 1/k < \alpha < 1$	0	1	$\Theta(n)$	$\sim n^{1+\gamma \sum_{i=1}^k (d_i - 2\alpha)_+}$
$1 < \alpha < d_{(k)}/2$	0	0	$\Theta(n)$	$\sim n^{1+\gamma \sum_{i=1}^k (d_i - 2\alpha)_+}$
$d_{(k)}/2 < \alpha$	0	0	$\Theta(n)$	$\sim n$

to the context of previous studies, Corollary 2 restates our results in the special case when $R_n = 2(\zeta(d - 1))^{-1} \log(n/\nu)$. If $d = 2$, this clearly reduces to (3). Interested readers wishing to understand this special case can proceed directly to Section 2.4. In particular, Table 1 compares a part of our results with those of earlier works on connectivity, percolation, and clique counts. If we take the tree to be a single edge, then S_n denotes the number of edges. Our results restricted to this case are consistent with the power-law behaviour for degree distributions. Our results also indicate a similar power-law behaviour for more general sub-tree counts, revealing the dependence of the exponents on the degree sequence of the tree under consideration and the parameters α, ζ . This is also explained in detail in Section 2.4.

Let us now add a few words on our proofs. The expectation and variance asymptotics for sub-tree counts involve the Mecke formula for Poisson point processes and various estimates for the measure and metric on the Poincaré ball. By virtue of the tree assumption, the relative angles between points will exhibit sufficient independence (see Lemma 2). For the CLT, we use the abstract normal approximation result (see [32]) based on the Malliavin–Stein method. To use this result, we derive detailed bounds on the first- (add-one cost) and second-order difference operators of the sub-tree counts. In many applications to Euclidean stochastic geometric functionals, the first- and second-order difference operators are bounded by a constant, but in our setup, the bounds for these operators are unbounded in n . Nevertheless, we manage to apply [32] by carefully bounding the difference operators, so that their growth is dominated by our variance lower bounds (see Remark 3(iii)).

2. Our setup and results

2.1. The Poincaré ball

Our underlying metric space is the d -dimensional Poincaré ball $B_d^{(\zeta)}$ in (1), which is equipped with the Riemannian metric (2), where $-\zeta^2$ represents the negative (Gaussian) curvature with $\zeta > 0$. This is a standard model in the literature, but earlier studies (e.g. [10, 22, 29, 11, 7, 14, 25, 13, 27, 36, 28, 23, 1, 19, 20]) treated only the special case $d = 2$, whereas we allow for higher dimensions as well. We mention some of the basic properties of this metric space, and some more will be stated in Section 3.1. For more details on the Poincaré ball, we refer the reader to [41, 2], while [15] is a good resource for a quick reading. In what follows,

we often represent the point $x \in B_d^{(\zeta)}$ in terms of ‘hyperbolic’ polar coordinates: for $x \in B_d^{(\zeta)}$, we write $x = (r, \theta_1, \dots, \theta_{d-1})$, where $r \geq 0$ is the radial part of x , defined by

$$r = \frac{1}{\zeta} \log \frac{1 + |x|}{1 - |x|},$$

and $(\theta_1, \dots, \theta_{d-1}) \in C_d := [0, \pi]^{d-2} \times [0, 2\pi)$ is the angular part of x . Let d_ζ denote the hyperbolic distance induced by (2); then it satisfies $d_\zeta(0, x) = r$.

Using the hyperbolic polar coordinates $(r, \theta_1, \dots, \theta_{d-1})$, the metric (2) can be rewritten as

$$ds^2 = dr^2 + \left(\frac{\sinh(\zeta r)}{\zeta}\right)^2 \left(d\theta_1^2 + \sum_{k=2}^{d-1} \prod_{i=1}^{k-1} \sin^2 \theta_i d\theta_k^2\right), \tag{4}$$

from which we obtain the volume element,

$$dV = \left(\frac{\sinh(\zeta r)}{\zeta}\right)^{d-1} \prod_{i=1}^{d-2} \sin^{d-i-1} \theta_i dr d\theta_1 \dots d\theta_{d-1}.$$

We now generate ‘quasi-uniform’ random points on a sequence of growing compact subsets of the Poincaré ball. First, we choose a deterministic sequence $R_n, n \geq 1$, which grows to infinity as $n \rightarrow \infty$. We assume that the angular part of the random points is uniformly chosen from the density,

$$\pi(\theta_1, \dots, \theta_{d-1}) = \frac{\prod_{i=1}^{d-2} \sin^{d-i-1} \theta_i}{2 \prod_{i=1}^{d-1} \kappa_{d-i-1}}, \quad (\theta_1, \dots, \theta_{d-1}) \in C_d, \tag{5}$$

where we have set $\kappa_m = \int_0^\pi \sin^m \theta d\theta$, and trivially, $\kappa_0 = \pi$. We use the symbol π to denote the angular density as well as the famed constant, but the context makes it clear which of them we refer to. Given another curvature parameter $\alpha > 0$, we assume that the density of the radial part is

$$\rho_{n,\alpha}(r) = \frac{\sinh^{d-1}(\alpha r)}{\int_0^{R_n} \sinh^{d-1}(\alpha s) ds}, \quad 0 \leq r \leq R_n. \tag{6}$$

This density is described as the ratio of the surface area of $B(0, r)$ to the volume of $B(0, R_n)$, where $B(0, r)$ and $B(0, R_n)$ are both defined on $B_d^{(\alpha)}$. So the density (6) can be regarded as a uniform density (for the radial part) on $B_d^{(\alpha)}$. Combining (5) and (6), we can generate uniform random points on $B_d^{(\alpha)}$. Since their radial coordinates are bounded by R_n , we can also consider them as points in $B(0, R_n) \subset B_d^{(\alpha)}$. By using the same hyperbolic polar coordinates, i.e., the same radial and angular coordinates, we now represent these points in $B_d^{(\zeta)}$ as well, and construct the hyperbolic geometric graph in $B_d^{(\zeta)}$. Specifically, we connect any two distinct points X_i, X_j if $0 < d_\zeta(X_i, X_j) \leq R_n$. Obviously, (6) is no longer a uniform density (for the radial part) on $B_d^{(\zeta)}$, unless $\zeta = \alpha$; it is called ‘quasi-uniform’.

Though the probability density in (6) looks a little complicated, we shall mostly resort to the following useful approximation. Set $T := R_n - d_\zeta(0, X)$, where X is a random point with density $\rho_{n,\alpha} \otimes \pi$. Denote by $\bar{\rho}_{n,\alpha}(t)$ the density of T ; i.e.,

$$\bar{\rho}_{n,\alpha}(t) = \frac{\sinh^{d-1}(\alpha(R_n - t))}{\int_0^{R_n} \sinh^{d-1}(\alpha s) ds}, \quad 0 \leq t \leq R_n. \tag{7}$$

In the sequel, we often denote a random point X with density $\rho_{n,\alpha} \otimes \pi$ by its hyperbolic polar coordinates $X = (T, \Theta)$, where Θ is the angular part, and the radial part is described by T rather than $d_\zeta(0, X)$. The approximation result below was established for $d = 2$ in [14]. The proof is in Section 3.

Lemma 1.

(i) As $n \rightarrow \infty$, we have

$$\bar{\rho}_{n,\alpha}(t) \leq (1 + o(1))\alpha(d - 1)e^{-\alpha(d-1)t}$$

uniformly for $0 \leq t < R_n$.

(ii) For every $0 < \lambda < 1$, we have, as $n \rightarrow \infty$,

$$\bar{\rho}_{n,\alpha}(t) = (1 + o(1))\alpha(d - 1)e^{-\alpha(d-1)t}$$

uniformly for $0 \leq t \leq \lambda R_n$.

More concretely, this lemma claims that if t is restricted to $[0, \lambda R_n]$ for some fixed $\lambda \in (0, 1)$, then $\bar{\rho}_{n,\alpha}(t)$ is asymptotically and uniformly equal to $\alpha(d - 1)e^{-\alpha(d-1)t}$. If there is no such restriction on the range of t , we can only bound $\bar{\rho}_{n,\alpha}(t)$, up to constant factors, by the same quantity. For the exact asymptotics of $\bar{\rho}_{n,\alpha}(t)$, we apply the result in Case (ii), while Case (i) is used for bounding $\bar{\rho}_{n,\alpha}(t)$.

Now we define our main object of interest, the hyperbolic random geometric graph. The first ingredient is the Poisson point process on $B_d^{(\zeta)}$. For every $n \geq 1$, let $(X_i, i \geq 1)$ be a sequence of i.i.d. random points on $B_d^{(\zeta)}$ with common density $\rho_{n,\alpha} \otimes \pi$. Letting N_n be a Poisson random variable with mean n , independent of (X_i) , one can construct the Poisson point process $\mathcal{P}_n = \{X_1, X_2, \dots, X_{N_n}\}$ whose intensity measure is $n(\rho_{n,\alpha} \otimes \pi)$.

Definition 1. (Hyperbolic random geometric graph) *Let $R_n \rightarrow \infty$ as $n \rightarrow \infty$. The hyperbolic random geometric graph $HG_n(R_n; \alpha, \zeta)$ is a simple, undirected graph whose vertex set is the Poisson point process \mathcal{P}_n with intensity measure $n(\rho_{n,\alpha} \otimes \pi)$, and whose edge set is $\{(X_i, X_j): X_i, X_j \in \mathcal{P}_n, 0 < d_\zeta(X_i, X_j) \leq R_n\}$.*

Finally, we remark that even if $\alpha = \zeta$, the process \mathcal{P}_n is not the restriction of a stationary Poisson process to $B_d^{(\alpha)}$ unless $|B(0, R_n)| = n$. A random geometric graph on a stationary Poisson process is a different model, which is actually beyond the scope of this article.

2.2. Expectation and variance asymptotics for sub-tree counts

We introduce the notion of a graph homomorphism to define our sub-tree counts. Suppose H is a simple graph on $[k] := \{1, \dots, k\}$ with edge set $E(H)$. Given another simple graph $G = (V(G), E(G))$, a graph homomorphism from H to G refers to a function $f: [k] \rightarrow V(G)$ such that if $(i, j) \in E(H)$, then $(f(i), f(j)) \in E(G)$, i.e., the adjacency relation is preserved. We denote by $\mathcal{C}(H, G)$ the number of injective graph homomorphisms from H to G , i.e., of functions f as above that are injective homomorphisms. Informally, $\mathcal{C}(H, G)$ counts the number of copies of H in G . This can be represented easily as follows:

$$\mathcal{C}(H, G) = \sum_{(v_1, \dots, v_k) \in V(G)}^{\neq} \prod_{(i,j) \in E(H)} \mathbf{1}\{ (v_i, v_j) \in E(G) \}, \tag{8}$$

where \sum^{\neq} denotes the sum over distinct k -tuples v_1, \dots, v_k and $\mathbf{1}\{\cdot\}$ is an indicator function. Notice that the subgraphs counted by $\mathcal{C}(H, G)$ are not necessarily induced subgraphs in G isomorphic to H . By (8), it is easy to derive a monotonicity property: if H_1, H_2 are simple graphs on $[k]$ such that $E(H_1) \subset E(H_2)$, then $\mathcal{C}(H_2, G) \leq \mathcal{C}(H_1, G)$.

Let us return to our setup in Definition 1. If $N_n \geq k$, we denote a collection of k -tuples of distinct elements in \mathcal{P}_n by

$$\mathcal{P}_{n,\neq}^k := \{(X_{i_1}, \dots, X_{i_k}) \in \mathcal{P}_n^k : i_j \in \{1, \dots, N_n\}, i_j \neq i_\ell \text{ for } j \neq \ell\}. \tag{9}$$

Set $\mathcal{P}_{n,\neq}^k = \emptyset$ if $N_n < k$. Define the annulus $D_\gamma(R_n) := B(0, R_n) \setminus \mathring{B}(0, (1 - \gamma)R_n)$ for $0 < \gamma \leq 1$, where \mathring{A} denotes the interior of a set A . Constructing the hyperbolic random geometric graph on $\mathcal{P}_n \cap D_\gamma(R_n)$ as in Definition 1, we denote it by $HG_n^{(\gamma)}(R_n; \alpha, \zeta)$. For $k \geq 2$, let Γ_k be a tree with nodes $[k]$ and edge set E . Our interest lies in the sub-tree counts $S_n^{(\gamma)} := \mathcal{C}(\Gamma_k, HG_n^{(\gamma)}(R_n; \alpha, \zeta))$ for $\gamma \in (0, 1]$. Recalling that $T_i = R_n - d_\zeta(0, X_i)$ and using (8), one can represent $S_n^{(\gamma)}$ as

$$S_n^{(\gamma)} = \sum_{(X_1, \dots, X_k) \in \mathcal{P}_{n,\neq}^k} \prod_{(i,j) \in E} \mathbf{1}\{0 < d_\zeta(X_i, X_j) \leq R_n, T_i, T_j \leq \gamma R_n\}. \tag{10}$$

In particular, we write $S_n := S_n^{(1)}$. Then $S_n = \mathcal{C}(\Gamma_k, HG_n(R_n; \alpha, \zeta))$. Our first result gives the asymptotic growth rate of $\mathbb{E}(S_n^{(\gamma)})$ for $\gamma \in (0, 1]$.

Theorem 1. *Let Γ_k be a tree on k vertices ($k \geq 2$) with degree sequence d_1, \dots, d_k , and let $S_n^{(\gamma)}$ be the sub-tree counts in (10). For $\gamma \in (0, 1) \setminus \{\frac{1}{2}\}$, we have*

$$\mathbb{E}(S_n^{(\gamma)}) \sim \left(\frac{2^{d-1}}{\kappa_{d-2}}\right)^{k-1} \alpha^k (d-1) n^k e^{-\zeta(d-1)(k-1)R_n/2} \prod_{i=1}^k a_n^{(\gamma)}(d_i), \quad n \rightarrow \infty, \tag{11}$$

where

$$a_n^{(\gamma)}(p) := \int_0^{\gamma R_n} e^{\zeta(d-1)(p-2\alpha/\zeta)t/2} dt, \quad p \in \mathbb{N}_+.$$

For $\gamma = 1/2$, we have

$$\mathbb{E}(S_n^{(\gamma)}) = \Theta(n^k e^{-\zeta(d-1)(k-1)R_n/2} \prod_{i=1}^k a_n^{(1/2)}(d_i)), \quad n \rightarrow \infty. \tag{12}$$

Furthermore, let $d_{(1)} \leq d_{(2)} \leq \dots \leq d_{(k)}$ be the degree sequence of Γ_k arranged in ascending order. If $2\alpha/\zeta > d_{(k)}$, then, for all $\gamma \in (0, 1]$, we have

$$\mathbb{E}(S_n) \sim \mathbb{E}(S_n^{(\gamma)}) \sim \left(\frac{2^{d-1}}{(d-1)\kappa_{d-2}}\right)^{k-1} \alpha^k \prod_{i=1}^k \left(\alpha - \frac{\zeta d_i}{2}\right)^{-1} n^k e^{-\zeta(d-1)(k-1)R_n/2}, \quad n \rightarrow \infty. \tag{13}$$

Remark 1. Many of the previous studies cited in Section 1.2 set either α or ζ to be 1, because in their setup, the asymptotics of the hyperbolic random geometric graph is completely determined by the ratio α/ζ . The crucial assumption in these studies is

$$R_n = 2[\zeta(d-1)]^{-1} \log(n/\nu) \text{ with } d = 2, \tag{14}$$

for some positive constant ν . In the current study, however, we do not put any stringent assumptions on R_n as in (14). Consequently, as seen in (11)–(13), ζ contributes, as an individual parameter, to the asymptotics of sub-tree counts via the term $e^{-\zeta(d-1)(k-1)R_n/2}$.

Theorem 1 indicates that the asymptotics of $\mathbb{E}(S_n^{(\gamma)})$ are crucially determined by the underlying curvatures $-\zeta^2$ and $-\alpha^2$. To make the implication of the theorem more transparent, we fix ζ and think of α as a parameter. If $\alpha > 0$ is sufficiently large, i.e., the space $B_d^{(\alpha)}$ is sufficiently hyperbolic, sub-trees are asymptotically dominated by the contributions near the boundary of $B(0, R_n)$. More specifically, if $2\alpha/\zeta > d_{(k)}$, then $a_n^{(\gamma)}(d_i), i = 1, \dots, k$, all converge to a positive constant, and thus the growth rate of $\mathbb{E}(S_n^{(\gamma)})$ does not depend on γ , implying that the spatial distribution of sub-trees is completely determined by those near the boundary of $B(0, R_n)$. In fact, for each $\gamma \in (0, 1)$, the growth rate of $\mathbb{E}(S_n^{(\gamma)})$ coincides with that of $\mathbb{E}(S_n)$.

On the other hand, if α becomes smaller, i.e., the space $B_d^{(\alpha)}$ becomes flatter, then the spatial distribution of sub-trees begins to be affected by those scattered away from the boundary of $B(0, R_n)$. For example, if $d_{(k-1)} < 2\alpha/\zeta < d_{(k)}$, we see that as $n \rightarrow \infty$,

$$a_n^{(\gamma)}(d_{(k)}) \sim \frac{1}{d-1} \left(\frac{\zeta d_{(k)}}{2} - \alpha \right)^{-1} e^{\zeta(d-1)(d_{(k)}-2\alpha/\zeta)\gamma R_n/2} \rightarrow \infty,$$

while $a_n^{(\gamma)}(d_i), i = 1, \dots, k-1$, all tend to a positive constant. In this case, the growth rate of $\mathbb{E}(S_n^{(\gamma)})$ is no longer independent of γ , and the growth rate becomes faster as $\gamma \nearrow 1$, i.e., as the inner radius of the corresponding annulus shrinks. Moreover, if α becomes even smaller, so that $d_{(k-2)} < 2\alpha/\zeta < d_{(k-1)}$, then $a_n^{(\gamma)}(d_{(k-1)})$ also asymptotically contributes to the growth rate of $\mathbb{E}(S_n^{(\gamma)})$. Finally, if $0 < 2\alpha/\zeta < d_{(1)}$, then all of the terms $a_n^{(\gamma)}(d_i)$ contribute to the growth rate of $\mathbb{E}(S_n^{(\gamma)})$.

The corollary below claims that if an underlying tree Γ_k is a k -star such that $d_{(k)} = k-1, d_{(k-1)} = \dots = d_{(1)} = 1$, then even more can be said about the asymptotics of $\log \mathbb{E}(S_n)$.

Corollary 1. *Let Γ_k be a k -star with $k \geq 2$. Moreover, assume that R_n satisfies*

$$\frac{R_n}{\log n} \rightarrow c, \quad n \rightarrow \infty, \quad \text{for some } c \in [0, \infty]$$

(note that $c = 0$ or $c = \infty$ is possible). Let $a \vee b = \max\{a, b\}$ for $a, b \in \mathbb{R}$.

(i) *If $2\alpha/\zeta > k-1$,*

$$\begin{aligned} \mathbb{E}(S_n) \sim & \left(\frac{2^{d-1}}{(d-1)\kappa_{d-2}} \right)^{k-1} \alpha^k (\alpha - \zeta(k-1)/2)^{-1} \\ & \times (\alpha - \zeta/2)^{-(k-1)} n^k e^{-\zeta(d-1)(k-1)R_n/2}, \quad n \rightarrow \infty. \end{aligned} \quad (15)$$

(ii) *If $1 < 2\alpha/\zeta \leq k-1$,*

$$\frac{\log \mathbb{E}(S_n)}{R_n \vee \log n} \rightarrow k(c \vee 1)^{-1} - \alpha(d-1)(1 \vee c^{-1})^{-1}, \quad n \rightarrow \infty.$$

(iii) *If $0 < 2\alpha/\zeta \leq 1$,*

$$\frac{\log \mathbb{E}(S_n)}{R_n \vee \log n} \rightarrow k(c \vee 1)^{-1} - (d-1)(\alpha k - \zeta(k-1)/2)(1 \vee c^{-1})^{-1}, \quad n \rightarrow \infty.$$

Our ultimate goal is to establish the CLT for the sub-tree counts $S_n^{(\gamma)}$ for $0 < \gamma \leq 1$. Before describing it, however, it is important to investigate the variance asymptotics. The theorem below provides an asymptotic lower bound for $\mathbb{V}ar(S_n^{(\gamma)})$ up to a constant factor. Similarly to Theorem 1, if $\alpha/\zeta > d_{(k)}$, the lower bound of $\mathbb{V}ar(S_n^{(\gamma)})$ is independent of γ , whereas it depends on γ when $\alpha/\zeta \leq d_{(k)}$. Furthermore, if $\alpha/\zeta > d_{(k)}$, we can obtain the exact growth rate of $\mathbb{V}ar(S_n)$. For two sequences (a_n) and (b_n) , we write $a_n = \Omega(b_n)$ if there exists a positive constant C such that $a_n/b_n \geq C$ for all $n \geq 1$. Furthermore, C^* denotes a generic positive constant, which may vary between lines and does not depend on n .

Theorem 2. *Let Γ_k be a tree on k vertices ($k \geq 2$) with degree sequence d_1, \dots, d_k , and let $S_n^{(\gamma)}$ be the sub-tree counts in (10). For $0 < \gamma < 1$,*

$$\mathbb{V}ar(S_n^{(\gamma)}) = \Omega\left(n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} a_n^{(\gamma)} (2d_{(k)}) \prod_{i=1}^{k-1} a_n^{(\gamma)}(d_{(i)})^2 \right. \\ \left. \vee n^k e^{-\zeta(d-1)(k-1)R_n/2} \prod_{i=1}^k a_n^{(\gamma)}(d_i) \right). \tag{16}$$

Suppose further that $\alpha/\zeta > d_{(k)}$; then, for all $\gamma \in (0, 1]$, we have

$$\mathbb{V}ar(S_n) \sim \mathbb{V}ar(S_n^{(\gamma)}) \sim C^* \left[n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} \vee n^k e^{-\zeta(d-1)(k-1)R_n/2} \right], \quad n \rightarrow \infty. \tag{17}$$

The main point of (17) is that the growth rate of $\mathbb{V}ar(S_n)$ is determined by how rapidly R_n grows to infinity. To see this in more detail, we consider a special case for which $R_n = c \log n$ for some $c \in (0, \infty)$. Then we can observe the following phase transitions:

$$\mathbb{V}ar(S_n) \sim \begin{cases} C^* n^{2k-1-c\zeta(d-1)(k-1)} & \text{if } 0 < c < 2\zeta^{-1}(d-1)^{-1}, \\ C^* n & \text{if } c = 2\zeta^{-1}(d-1)^{-1}, \\ C^* n^{k-c\zeta(d-1)(k-1)/2} & \text{if } c > 2\zeta^{-1}(d-1)^{-1}. \end{cases}$$

Moreover, Theorems 1 and 2 imply that, in the case of $\alpha/\zeta < d_{(k)} < 2\alpha/\zeta$, $\mathbb{V}ar(S_n^{(\gamma)})$ grows super-linearly, while the expectation exhibits only linear growth. A similar phenomenon has been observed for isolated points in [23, Theorem 1.1]. This is indicative of a power-law behaviour in the sub-tree counts.

2.3. Central limit theorem for sub-tree counts

Having derived variance bounds, we proceed to prove the CLT for $S_n^{(\gamma)}$. Before stating the normal approximation result, we need to define the two metrics to be used: the Wasserstein distance d_W and the Kolmogorov distance d_K . Let Y_1, Y_2 be two random variables, and let $Lip(1)$ be the set of Lipschitz functions $h : \mathbb{R} \rightarrow \mathbb{R}$ with Lipschitz constant of at most 1. Then we define the two metrics as follows:

$$d_W(Y_1, Y_2) = \sup_{h \in Lip(1)} | \mathbb{E}(h(Y_1)) - \mathbb{E}(h(Y_2)) |,$$

$$d_K(Y_1, Y_2) = \sup_{x \in \mathbb{R}} | \mathbb{P}(Y_1 \leq x) - \mathbb{P}(Y_2 \leq x) |.$$

Although we have defined d_W, d_K as distances between two random variables, they are actually distances between two probability distributions. Let N denote the standard normal random variable, and let \Rightarrow denote weak convergence in \mathbb{R} . In our proof, we derive more detailed, albeit complicated, bounds that indicate the scope for improvement.

Theorem 3. *Let Γ_k be a tree on k vertices ($k \geq 2$) with degree sequence d_1, \dots, d_k , and let $S_n^{(\gamma)}$ be the sub-tree counts in (10). Assume further that R_n satisfies*

$$ne^{-\zeta(d-1)R_n/2} \rightarrow c \in (0, \infty]. \tag{18}$$

For every $0 < a < 1/2$, there exists $0 < \gamma_0 < 1/2$ such that for all $0 < \gamma < \gamma_0$, we have

$$d_W \left(\frac{S_n^{(\gamma)} - \mathbb{E}(S_n^{(\gamma)})}{\sqrt{\text{Var}(S_n^{(\gamma)})}}, N \right) = O(n^{-a}), \quad d_K \left(\frac{S_n^{(\gamma)} - \mathbb{E}(S_n^{(\gamma)})}{\sqrt{\text{Var}(S_n^{(\gamma)})}}, N \right) = O(n^{-a}). \tag{19}$$

Furthermore, if $\alpha/\zeta > d_{(k)}$, then

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \Rightarrow N, \quad n \rightarrow \infty. \tag{20}$$

Here, we claim that for a small enough γ , the CLT holds for $S_n^{(\gamma)}$, and if $\alpha/\zeta > d_{(k)}$, the CLT holds for S_n as well. Heuristically, a larger α/ζ means that the space $B_d^{(\alpha)}$ is more hyperbolic relative to $B_d^{(\zeta)}$ and hence contains more points in the boundary. So in that case, S_n is well approximated by $S_n^{(\gamma)}$. Note that the CLT for $S_n^{(\gamma)}$ possibly holds even when the variance grows super-linearly in n . It is still unclear whether or not the CLT holds for S_n under super-linear growth of the variance. In [23, Theorem 1.3], it is shown that the CLT for ‘isolated points’ cannot hold under super-linear growth of the variance. Finally, Remark 4 at the end of Section 3.3.3 gives a more precise quantification of the bounds in (19).

Remark 2. If the constant c in (18) is finite and $2\alpha/\zeta > d_{(k)}$, we see from (13) that $\mathbb{E}(S_n)$ is asymptotically equal to n , up to constant factors. This implies that S_n behaves similarly to the subgraph counts in the thermodynamic regime of the Euclidean setting (more detailed arguments on this point are given in Section 2.5). If $c = \infty$ in (18), R_n grows more slowly, so that the spatial distribution of Γ_k becomes denser. For the sparse phase asymptotics, the first author’s ongoing work studies the case in which R_n diverges faster than (18), so that $\mathbb{E}(S_n)$ is asymptotically equal to a positive constant. In this case, the tree-like structure rarely occurs, and S_n will be approximated by a Poisson random variable.

2.4. Special case and comparison with previous studies

The objective of this section is to restate our results in the special case

$$R_n = 2[\zeta(d-1)]^{-1} \log(n/\nu) \tag{21}$$

for some positive constant $\nu \in (0, \infty)$. By doing so, we can make an easy comparison with previous studies. As mentioned before, most earlier works (e.g. [10, 22, 29, 11, 7, 14, 25, 13, 27, 36, 28, 23, 1, 19, 20]) are concerned only with the case

$$d = 2 \quad \text{and} \quad R_n = 2\zeta^{-1} \log(n/\nu) \quad \text{for some } \nu \in (0, \infty), \tag{22}$$

or its variants; hence, (21) serves as a higher-dimensional version of (22). Assuming (21) and $2\alpha/\zeta \notin \mathbb{N}$, it is easy to see that as $n \rightarrow \infty$,

$$a_n^{(\gamma)}(p) \sim |(d-1)(\alpha - \zeta p/2)|^{-1} \left(\frac{n}{v}\right)^{\gamma(p-2\alpha/\zeta)_+}, \quad p \in \mathbb{N}_+, \tag{23}$$

where $(a)_+ = a$ if $a > 0$ and $(a)_+ = 0$ otherwise.

Corollary 2. *Let $R_n = 2[\zeta(d-1)]^{-1} \log(n/v)$, let Γ_k be a tree on k vertices ($k \geq 2$) with degree sequence d_1, \dots, d_k , and let $S_n^{(\gamma)}$ be a sub-tree count as in (10). Suppose that $2\alpha/\zeta$ is not an integer. For $\gamma \in (0, 1) \setminus \{\frac{1}{2}\}$, we have, as $n \rightarrow \infty$,*

$$\mathbb{E}(S_n^{(\gamma)}) \sim \left(\frac{2^{d-1}}{(d-1)\kappa_{d-2}}\right)^{k-1} \alpha^k \prod_{j=1}^k \left|\alpha - \frac{\zeta}{2}d_j\right|^{-1} v^{k-1-\gamma \sum_{i=1}^k (d_i-2\alpha/\zeta)_+} n^{1+\gamma \sum_{i=1}^k (d_i-2\alpha/\zeta)_+},$$

and for $\gamma = 1/2$, we have

$$\mathbb{E}\left(S_n^{(\gamma)}\right) = \Theta\left(n^{1+2^{-1} \sum_{i=1}^k (d_i-2\alpha/\zeta)_+}\right). \tag{24}$$

Furthermore, as for the variance asymptotics, we have for $\gamma \in (0, 1)$

$$\text{Var}(S_n^{(\gamma)}) = \Omega\left(n^{1+2\gamma(d_{(k)}-\alpha/\zeta)_++2\gamma \sum_{i=1}^{k-1} (d_{(i)}-2\alpha/\zeta)_+} \sqrt{n^{1+\gamma \sum_{i=1}^k (d_i-2\alpha/\zeta)_+}}\right).$$

If $2\alpha/\zeta > d_{(k)}$, we have for $\gamma \in (0, 1]$

$$\mathbb{E}(S_n) \sim \mathbb{E}(S_n^{(\gamma)}) \sim \left(\frac{2^{d-1}}{(d-1)\kappa_{d-2}}\right)^{k-1} \alpha^k \prod_{j=1}^k \left(\alpha - \frac{\zeta}{2}d_j\right)^{-1} v^{k-1}n, \quad n \rightarrow \infty. \tag{25}$$

If $\alpha/\zeta > d_{(k)}$, we have for $\gamma \in (0, 1]$

$$\text{Var}(S_n) \sim \text{Var}(S_n^{(\gamma)}) \sim C^*n, \quad n \rightarrow \infty.$$

To avoid repetition, we do not state the corresponding CLT. However, it holds under the assumptions of Corollary 2. Note that (24) is a partial generalization of [14, Claim 5.2] to higher dimensions.

Now we compare the results in Corollary 2 with those in the existing literature. If we choose $k = 2$ in (25), then S_n represents the number of edges with $d_{(2)} = 1$. Moreover, $S_n/2n$ denotes the empirical degree distribution. Then (25) asserts that $\mathbb{E}(S_n/2n) \rightarrow C^*$ if $\alpha/\zeta > 1/2$. The convergence of the expected typical degree to a constant indicates that this regime is, in nature, the thermodynamic regime. Furthermore, when $1/2 < \alpha/\zeta < 1$, the expected degree distribution converges to a constant, but the variance diverges. This is consistent with the power-law behaviour for degree distributions with exponent $2\alpha/\zeta + 1$, which itself was predicted in [30]. Such a power-law behaviour for degree distributions was proven in [25, 19] for $d = 2$. For $2\alpha/\zeta \leq 1$, the expected average degree grows to infinity, which is again consistent with the conjecture that the degree distribution has a power-law behaviour with exponent 2 (see [30]).

In Table 1, we summarize some of the results in the existing literature and those in this paper in the case $d = 2$, $\zeta = 1$, $R_n = 2 \log(n/v)$, $v \in (0, \infty)$. As seen in Table 1, our results demonstrate various phase transitions. For example, for $1 < \alpha < d_{(k)}/2$, the expected sub-tree counts grow super-linearly in n , whereas the expected number of k -cliques is linear, and there is no ‘giant component’. This justifies our observation in Section 1.2 that hyperbolic random geometric graphs contain many tree-like hierarchical structures, as compared to their Euclidean counterparts.

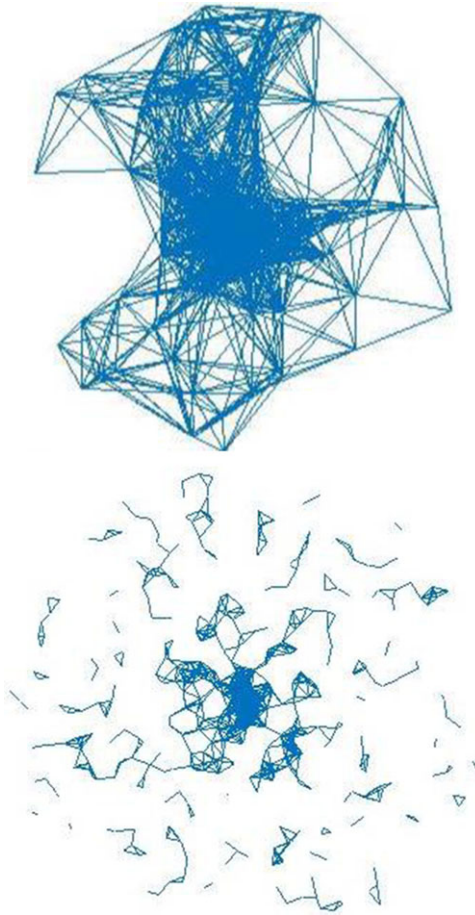


FIGURE 3. Simulations of $EG_{1,100}$ with $\pi r_{100}^2 = 100$ and $EG_{2,500}$ with $\pi r_{500}^2 = 500$, $s_{500} = 1$ for $d = 2$.

2.5. Comparison to subgraph counts of Euclidean random geometric graphs

In this section, we sketch asymptotic results for subgraph counts of analogous random geometric graph models in the Euclidean space. We claim that the behaviour of the subgraph counts of the Euclidean analogues is significantly different from that of the counts in the previous sections. For example, in contrast to Theorem 1, the growth rates for the subgraph counts in the Euclidean case do not depend on the degree sequence of the subgraph under consideration.

For simplicity, we restrict ourselves to the case $\alpha = \zeta = 1$ and $R_n = 2(d-1)^{-1} \log(n/\nu)$ for some $\nu > 0$ as in Section 2.4. Then, by Corollary 2, we have that $\mathbb{E}(E_n) = \Theta(n)$, where E_n denotes the number of edges in $HG_n(R_n; 1, 1)$. We now consider two possible Euclidean analogues in the setting of Corollary 2. One of them is defined by following the construction of $HG_n(R_n; 1, 1)$, and the other is defined to ensure that the expected degree is of the same order as $HG_n(R_n; 1, 1)$. It is a moot point as to which is the right analogue, and hence we make comparisons to both. See Figure 3 for the simulations of these Euclidean graphs.

The two analogues are constructed as follows:

1. Consider Poisson(n) points distributed uniformly in a sequence of growing Euclidean balls of radius $r_n \rightarrow \infty$, and connect any two points within a distance of r_n (this is the *dense regime*). We call this graph $EG_{1,n}$. It has been defined by mimicking the construction of $HG_n(R_n; 1, 1)$.
2. Consider Poisson(n) points distributed uniformly in a sequence of Euclidean balls of radius r_n , and connect points within a distance s_n such that the expected number of edges grows linearly in n (this is the *thermodynamic regime*). We call this graph $EG_{2,n}$. The expected degree of this graph is of the same order as $HG_n(R_n; 1, 1)$.

We now give a sketch of the calculations for the subgraph counts in the above two models, referring the reader to [38, Chapter 3] for details. Denote by \mathcal{X}_n the collection of Poisson(n) points distributed uniformly in $B_{r_n}(0) \subset \mathbb{R}^d$, where $B_{r_n}(0)$ is the d -dimensional ball of radius r_n centred at the origin. Then $EG_n(r_n, s_n)$ denotes the graph with vertex set \mathcal{X}_n and edges between $X_i, X_j \in \mathcal{X}_n$ such that $|X_i - X_j| \leq s_n$, where $|\cdot|$ is the Euclidean metric. In this notation, $EG_{1,n} = EG_n(r_n, r_n)$ and $EG_{2,n} = EG_n(r_n, s_n)$, where s_n is chosen such that the expected number of edges grows linearly in n . Let Γ be a connected graph on k vertices, and let $J_{i,n}(\Gamma)$, $i = 1, 2$, denote the number of copies of Γ in $EG_{i,n}$, $i = 1, 2$. By the Mecke formula, we have that

$$\mathbb{E}(J_{1,n}(\Gamma)) = n^k \frac{C_1}{r_n^{dk}} \int_{B_{r_n}(0)^k} \prod_{(i,j) \in \Gamma} \mathbf{1}\{|x_i - x_j| \leq r_n\} dx_1 \dots dx_k = \Theta(n^k).$$

Since the order of $\mathbb{E}(J_{1,n}(\Gamma))$ is the same as that of the complete subgraph on \mathcal{X}_n , we call it the dense regime. We observe that the choice of r_n and the degree sequence of Γ do not contribute to the growth rate of $\mathbb{E}(J_{1,n}(\Gamma))$. This shows a striking contrast to the asymptotics for hyperbolic random geometric graphs in Corollary 2.

As in the above case, assuming $s_n = o(r_n)$, we can also derive that $\mathbb{E}(J_{2,n}(\Gamma)) = \Theta\left(n^k \left(\frac{s_n}{r_n}\right)^{d(k-1)}\right)$. In the special case of $k = 2$ and $s_n = n^{-1/d}r_n$, this implies that the expected number of edges is asymptotically equal to n , up to a constant factor. The resulting regime is called the thermodynamic regime (see [38, Chapter 3]), since the expected average degree (or empirical count of neighbours) is asymptotically constant. In contrast to the hyperbolic random geometric graph, the above calculation indicates that the asymptotics of $\mathbb{E}(J_{2,n}(\Gamma))$ is again independent of the degree sequence of Γ and the choice of r_n . Finally, we wish to emphasize that the broad message remains unchanged even for the second-order results; see [39, 8] for more detailed computations.

3. Proofs

We first prove the basic lemma on an approximation of the hyperbolic probability density. Subsequently, Section 3.1 presents a few more lemmas concerned with hyperbolic distance. Utilizing these lemmas, we prove expectation and variance results for the sub-tree counts in Section 3.2. Finally, Section 3.3 establishes the required CLT. Throughout this section, C^* denotes a generic positive constant, which may vary between lines and does not depend on n .

Proof of Lemma 1. We see that

$$\bar{\rho}_{n,\alpha}(t) = \frac{e^{\alpha(d-1)(R_n-t)} (1 - e^{-2\alpha(R_n-t)})^{d-1}}{\int_0^{R_n} e^{\alpha(d-1)s} (1 - e^{-2\alpha s})^{d-1} ds}. \tag{26}$$

Applying the binomial expansion

$$(1 - e^{-2\alpha s})^{d-1} = \sum_{k=0}^{d-1} \binom{d-1}{k} (-1)^k e^{-2\alpha ks}, \quad s > 0,$$

we find that $\int_0^{R_n} e^{\alpha(d-1)s} ds$ is asymptotically the leading term in the denominator in (26). Using the obvious inequality $(1 - e^{-2\alpha(R_n-t)})^{d-1} \leq 1$, we complete the proof of (i).

For the proof of (ii), we need to handle the numerator of (26) as well. By another application of the binomial expansion, we have that

$$e^{\alpha(d-1)(R_n-t)} \left[1 + \sum_{k=1}^{d-1} \binom{d-1}{k} (-1)^k e^{-2\alpha k(R_n-t)} \right] = e^{\alpha(d-1)(R_n-t)} (1 + o(1))$$

uniformly for $0 \leq t \leq \lambda R_n$, where $0 < \lambda < 1$. □

3.1. Lemmas on hyperbolic distance

We first present a crucial lemma that explains why we consider (not necessarily induced) sub-trees, which are counted as injective homomorphisms in the hyperbolic random geometric graph. In particular, the conditional independence proven in Lemma 2 below does not hold for more general subgraph counts. The asymptotics of general subgraph counts is highly non-trivial not only for ‘induced’ sub-tree counts, but also for even seemingly simple triangle counts. For the latter case, [14] introduced the notion of a global clustering coefficient defined as three times the ratio of the number of triangles to the number of paths of length two.

Lemma 2. *Let Γ_k be a tree on k vertices with edge set E . Let X_1, \dots, X_k be i.i.d. random points with common density $\rho_{n,\alpha} \otimes \pi$. Define $T_i = R_n - d_\zeta(0, X_i)$, $i = 1, \dots, k$, and write $\mathbf{T} = (T_1, \dots, T_k)$. Then it holds that*

$$\mathbb{P}(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E \mid \mathbf{T}) = \prod_{(i,j) \in E} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{T}) \text{ almost surely (a.s.)}$$

Proof. Let Θ_i be the angular part of X_i , $i = 1, \dots, k$. For the proof, it suffices to show that

$$\mathbb{P}(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E \mid \Theta_1, \mathbf{T}) = \prod_{(i,j) \in E} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{T}) \text{ a.s.} \tag{27}$$

Let us denote by Θ_{12} the relative angle between Θ_1 and Θ_2 . Then, because of the uniformity of the angular density of Θ_i in (5), we can derive that the density π_{rel} of Θ_{12} , which is also the same as the conditional density of $\Theta_{12} \mid \Theta_1$, is given by

$$\pi_{rel}(\theta) := (\kappa_{d-2})^{-1} \sin^{d-2} \theta, \quad \theta \in [0, \pi]. \tag{28}$$

Note that $d_\zeta(X_1, X_2)$ depends only on T_1, T_2, Θ_{12} , and so we obtain that

$$\mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \Theta_1, \mathbf{T}) = \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{T}) \text{ a.s.} \tag{29}$$

Suppose that Γ_k is of depth 1 rooted at vertex 1. In this case, since X_2, \dots, X_k are i.i.d.,

$$\begin{aligned} \mathbb{P}(d_\zeta(X_1, X_j) \leq R_n, j = 2, \dots, k \mid \Theta_1, \mathbf{T}) &= \prod_{j=2}^k \mathbb{P}(d_\zeta(X_1, X_j) \leq R_n \mid \Theta_1, \mathbf{T}) \\ &= \prod_{j=2}^k \mathbb{P}(d_\zeta(X_1, X_j) \leq R_n \mid \mathbf{T}) \text{ a.s.} \end{aligned}$$

Suppose, for induction, that (27) holds for any tree Γ_k rooted at vertex 1 with depth $M \geq 1$. Assume subsequently that Γ_k is rooted at vertex 1 with depth $M + 1$. Let $2, \dots, m$ be the vertices connected to 1, and S_2, \dots, S_m the corresponding trees rooted at $2, \dots, m$. Let $E(S_\ell)$ be the edge set of S_ℓ . Then, from the disjointness of the trees and independence of the points X_i , we have

$$\mathbb{P}(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E \mid \Theta_1, \mathbf{T}) = \prod_{\ell=2}^m \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E(S_\ell) \mid \Theta_1, \mathbf{T}) \text{ a.s.}$$

Now, an application of conditional expectations for each $\ell = 2, \dots, m$ gives us the desired result as follows:

$$\begin{aligned} &\mathbb{P}(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E(S_\ell) \mid \Theta_1, \mathbf{T}) \\ &= \mathbb{E} \left[\mathbf{1}\{d_\zeta(X_1, X_\ell) \leq R_n\} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E(S_\ell) \setminus \{(1, \ell)\} \mid \Theta_\ell, \mathbf{T}) \mid \Theta_1, \mathbf{T} \right] \\ &= \mathbb{E} \left[\mathbf{1}\{d_\zeta(X_1, X_\ell) \leq R_n\} \prod_{(i,j) \in E(S_\ell) \setminus \{(1,\ell)\}} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{T}) \mid \Theta_1, \mathbf{T} \right] \\ &= \prod_{(i,j) \in E(S_\ell)} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{T}) \text{ a.s.,} \end{aligned}$$

where the induction hypothesis is used for the second equality, and (29) for the third equality. □

We now mention some lemmas that help us to approximate the Poincaré metric. Given $u_1, u_2 \in B(0, R_n)$, let $\theta_{12} \in [0, \pi]$ be the relative angle between the two vectors $\overrightarrow{Ou_1}$ and $\overrightarrow{Ou_2}$, where O denotes the origin of $B_d^{(\zeta)}$. We also define $t_i = R_n - d_\zeta(0, u_i), i = 1, 2$.

Lemma 3. *Set $\hat{\theta}_{12} = (e^{-2\zeta(R_n-t_1)} + e^{-2\zeta(R_n-t_2)})^{1/2}$. If $\hat{\theta}_{12}/\theta_{12}$ vanishes as $n \rightarrow \infty$, then*

$$d_\zeta(u_1, u_2) = 2R_n - (t_1 + t_2) + \frac{2}{\zeta} \log \sin \left(\frac{\theta_{12}}{2} \right) + O \left(\left(\frac{\hat{\theta}_{12}}{\theta_{12}} \right)^2 \right), \quad n \rightarrow \infty,$$

uniformly for all u_1, u_2 , with $t_1 + t_2 \leq R_n - \omega_n$, where $\omega_n = \log \log R_n$.

Proof. Fix a great circle of $B(0, R_n)$ spanned by $\overrightarrow{Ou_1}$ and $\overrightarrow{Ou_2}$. Then the hyperbolic law of cosines yields

$$\cosh(\zeta d_\zeta(u_1, u_2)) = \cosh \zeta(R_n - t_1) \cosh \zeta(R_n - t_2) - \sinh \zeta(R_n - t_1) \sinh \zeta(R_n - t_2) \cos(\theta_{12}).$$

Since this great circle is a two-dimensional subspace of $B(0, R_n)$, the rest of the argument is exactly the same as that for Lemma 2.3 in [19]. □

Lemma 4. *In terms of the hyperbolic polar coordinates, let $X_1 = (t_1, \Theta_1), X_2 = (t_2, \Theta_2)$, where Θ_1, Θ_2 are i.i.d. random vectors on C_d with density π , and t_1, t_2 are deterministic, representing the hyperbolic distance from the boundary. With the setup in Lemma 3,*

$$\mathbb{P}(d_\zeta(X_1, X_2) \leq R_n) \sim \frac{2^{d-1}}{(d-1)\kappa_{d-2}} e^{-\zeta(d-1)(R_n-t_1-t_2)/2}, \quad n \rightarrow \infty,$$

uniformly on $\{(t_1, t_2) : t_1 + t_2 \leq R_n - \omega_n\}$, where $\kappa_{d-2} = \int_0^\pi \sin^{d-2} \theta \, d\theta$ and $\omega_n = \log \log R_n$.

Proof. We again denote the relative angle between Θ_1 and Θ_2 by Θ_{12} . Recall that $\Theta_{12} \stackrel{d}{=} |\Theta_2 - \Theta_1|$. From the hyperbolic law of cosines, we know that the distance between X_1 and X_2 is determined by t_1, t_2 and their relative angle Θ_{12} . Since t_1, t_2 are fixed, by (28) we can write

$$\begin{aligned} \mathbb{P}(d_\zeta(X_1, X_2) \leq R_n) &= \int_{C_d^2} \mathbf{1}\{d_\zeta(u_1, u_2) \leq R_n\} \pi(\theta_1)\pi(\theta_2) \, d\theta_1 d\theta_2 \\ &= \frac{1}{\kappa_{d-2}} \int_0^\pi \mathbf{1}\{d_\zeta(u_1, u_2) \leq R_n\} \sin^{d-2} \theta_{12} \, d\theta_{12}, \end{aligned}$$

where we write $u_i = (t_i, \theta_i), i = 1, 2$, and θ_{12} is the relative angle between u_1 and u_2 . Now, we shall approximate the above integral. Let $A_{12} = e^{\zeta(R_n-t_1-t_2)/2}$. Claim 2.5 in [19] proves that $A_{12}^{-1}/(\omega_n \hat{\theta}_{12}) \rightarrow \infty$ as $n \rightarrow \infty$. By virtue of Lemma 3, along with $A_{12}^{-1} \rightarrow 0$ as $n \rightarrow \infty$, we have, on the set $\{(t_1, t_2) : t_1 + t_2 \leq R_n - \omega_n\}$,

$$\begin{aligned} \int_{A_{12}^{-1}/\omega_n}^\pi \mathbf{1}\{d_\zeta(u_1, u_2) \leq R_n\} \sin^{d-2} \theta_{12} \, d\theta_{12} &\sim \int_{A_{12}^{-1}/\omega_n}^\pi \mathbf{1}\left\{\sin\left(\frac{\theta_{12}}{2}\right) \leq A_{12}^{-1}\right\} \sin^{d-2} \theta_{12} \, d\theta_{12} \\ &\sim \int_{A_{12}^{-1}/\omega_n}^\pi \mathbf{1}\{\theta_{12} \leq 2A_{12}^{-1}\} (\theta_{12})^{d-2} \, d\theta_{12} \sim \frac{(2A_{12}^{-1})^{d-1}}{d-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\int_0^\pi \mathbf{1}\{d_\zeta(u_1, u_2) \leq R_n\} \sin^{d-2} \theta_{12} \, d\theta_{12} \\ &= o(A_{12}^{-(d-1)}) + \int_{A_{12}^{-1}/\omega_n}^\pi \mathbf{1}\{d_\zeta(u_1, u_2) \leq R_n\} \sin^{d-2} \theta_{12} \, d\theta_{12} \sim \frac{(2A_{12}^{-1})^{d-1}}{d-1}, \quad n \rightarrow \infty, \end{aligned}$$

as required. □

3.2. Proofs of the expectation and variance results

In what follows, we calculate the moments of $S_n^{(\gamma)}$. We first introduce some notation to save space. For $u_i \in B(0, R_n), i = 1, \dots, k$, we write

$$g_{n,\gamma}(u_1, \dots, u_k) := \mathbf{1}\{0 < d_\zeta(u_i, u_j) \leq R_n, (i, j) \in E, t_i \leq \gamma R_n, i = 1, \dots, k\} \tag{30}$$

and

$$h_{n,\gamma}(u_1, \dots, u_k) := \mathbf{1}\{0 < d_\zeta(u_i, u_j) \leq R_n, (i, j) \in E, t_i > \gamma R_n \text{ for some } i = 1, \dots, k\}, \tag{31}$$

where $t_i = R_n - d_\zeta(0, u_i)$. Note that $g_{n,\gamma}$ represents the product of multiple indicator functions in the right-hand side of (10), while $h_{n,\gamma}$ is a similar indicator for $S_n - S_n^{(\gamma)}$.

Proof of Theorem 1. Let X_1, \dots, X_k be i.i.d. random points with density $\rho_{n,\alpha} \otimes \pi$, and set $T_i = R_n - d_\zeta(0, X_i)$ as before. By an application of the multivariate Mecke formula for Poisson processes (see, e.g., [33, Chapter 4]),

$$\begin{aligned} \mathbb{E}(S_n^{(\gamma)}) &= n^k \mathbb{E}(g_{n,\gamma}(X_1, \dots, X_k)) \\ &= n^k \mathbb{E}[g_{n,\gamma}(X_1, \dots, X_k) \prod_{(i,j) \in E} \mathbf{1}\{T_i + T_j \leq R_n - \omega_n\}] \\ &\quad + n^k \mathbb{E}[g_{n,\gamma}(X_1, \dots, X_k) \mathbf{1}\{\cup_{(i,j) \in E} \{T_i + T_j > R_n - \omega_n\}\}] \\ &=: A_n + B_n, \end{aligned} \tag{32}$$

where $\omega_n = \log \log R_n$. From now on, the argument aims to show that A_n coincides asymptotically with the right-hand side of (11). It follows from the conditioning on the radial distances of X_1, \dots, X_k from the boundary and Lemma 2 that

$$\begin{aligned} A_n &= n^k \mathbb{E} \left[\mathbf{1}\{T_i + T_j \leq R_n - \omega_n, (i, j) \in E, T_i \leq \gamma R_n, i = 1, \dots, k\} \prod_{(i,j) \in E} \right. \\ &\quad \left. \times \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{T}) \right], \end{aligned}$$

where $\mathbf{T} = (T_1, \dots, T_k)$. Setting $\mathbf{t} = (t_1, \dots, t_k)$, we may write

$$\begin{aligned} A_n &= n^k \int_0^{\gamma R_n} dt_1 \cdots \int_0^{\gamma R_n} dt_k \mathbf{1}\{t_i + t_j \leq R_n - \omega_n, (i, j) \in E\} \\ &\quad \times \prod_{(i,j) \in E} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{t}) \bar{\rho}_{n,\alpha}(\mathbf{t}), \end{aligned} \tag{33}$$

where $\bar{\rho}_{n,\alpha}(\mathbf{t})$ denotes the product of densities in (7):

$$\bar{\rho}_{n,\alpha}(\mathbf{t}) := \prod_{i=1}^k \bar{\rho}_{n,\alpha}(t_i).$$

Applying Lemma 4 on the set $\{t_i + t_j \leq R_n - \omega_n, (i, j) \in E\}$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \prod_{(i,j) \in E} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{t}) &\sim \prod_{(i,j) \in E} \frac{2^{d-1}}{(d-1)\kappa_{d-2}} e^{-\zeta(d-1)(R_n - t_i - t_j)/2} \\ &= \left(\frac{2^{d-1}}{(d-1)\kappa_{d-2}} \right)^{k-1} e^{-[(k-1)R_n - \sum_{i=1}^k d_i t_i] \zeta(d-1)/2}. \end{aligned}$$

Furthermore, it follows from Lemma 1(ii) that

$$\bar{\rho}_{n,\alpha}(\mathbf{t}) \sim \alpha^k (d-1)^k e^{-\alpha(d-1) \sum_{i=1}^k t_i}, \quad n \rightarrow \infty,$$

uniformly for $t_i \leq \gamma R_n, i = 1, \dots, k$. Substituting these results into (33), we obtain

$$A_n \sim \left(\frac{2^{d-1}}{\kappa^{d-2}}\right)^{k-1} \alpha^k (d-1) n^k e^{-\zeta(d-1)(k-1)R_n/2} \prod_{i=1}^k a_n^{(\gamma)}(d_i). \tag{34}$$

We now show that $B_n = o(A_n)$ as $n \rightarrow \infty$, unless $\gamma = 1/2$. We check only the case in which there is an edge joining X_1 and X_2 such that $T_1 + T_2 > R_n - \omega_n$, while all the other edges satisfy $T_i + T_j \leq R_n - \omega_n$. However, the argument below can be applied in an obvious way even when multiple edges satisfy $T_i + T_j > R_n - \omega_n$. Specifically, we shall verify

$$C_n := n^k \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E, T_i \leq \gamma R_n, i = 1, \dots, k, T_1 + T_2 > R_n - \omega_n, T_i + T_j \leq R_n - \omega_n, (i, j) \in E \setminus \{(1, 2)\}) = o(A_n). \tag{35}$$

Proceeding in the same way as in the computation for A_n , while applying the obvious bound

$$\mathbf{1}\{d_\zeta(X_1, X_2) \leq R_n\} \leq 1,$$

we see that

$$C_n \leq C^* n^k e^{-\zeta(d-1)(k-2)R_n/2} \prod_{i=3}^n a_n^{(\gamma)}(d_i) \times \int_0^{\gamma R_n} \int_0^{\gamma R_n} e^{2^{-1}\zeta(d-1)\sum_{i=1}^2 (d_i-1-2\alpha/\zeta)t_i} \mathbf{1}\{t_1 + t_2 > R_n - \omega_n\} dt_1 dt_2.$$

By comparing this upper bound with the right-hand side of (11), we see that it suffices to show that

$$D_n := e^{\zeta(d-1)R_n/2} \int_0^{\gamma R_n} \int_0^{\gamma R_n} e^{2^{-1}\zeta(d-1)\sum_{i=1}^2 (d_i-1-2\alpha/\zeta)t_i} \mathbf{1}\{t_1 + t_2 > R_n - \omega_n\} dt_1 dt_2 = o(a_n^{(\gamma)}(d_1)a_n^{(\gamma)}(d_2)), \quad n \rightarrow \infty. \tag{36}$$

If $0 < \gamma < 1/2$, then D_n is identically 0; hence, we may restrict ourselves to the case $1/2 < \gamma < 1$. Without loss of generality, we may assume that $d_1 \geq d_2$ and rewrite D_n as

$$D_n = e^{\zeta(d-1)R_n/2} \int_{(1-\gamma)R_n-\omega_n}^{\gamma R_n} e^{\zeta(d-1)(d_1-1-2\alpha/\zeta)t_1/2} \int_{R_n-t_1-\omega_n}^{\gamma R_n} e^{\zeta(d-1)(d_2-1-2\alpha/\zeta)t_2/2} dt_2 dt_1.$$

If $0 < 2\alpha/\zeta < d_2 - 1$, then

$$D_n \leq C^* e^{\zeta(d-1)[R_n+\sum_{i=1}^2 (d_i-1-2\alpha/\zeta)\gamma R_n]/2} = e^{(1/2-\gamma)\zeta(d-1)R_n} O(a_n^{(\gamma)}(d_1)a_n^{(\gamma)}(d_2)) = o(a_n^{(\gamma)}(d_1)a_n^{(\gamma)}(d_2)).$$

On the other hand, let $2\alpha/\zeta > d_2 - 1$. Then

$$D_n \leq C^* e^{\zeta(d-1)[(d_2-2\alpha/\zeta)R_n+(d_1-d_2)\gamma R_n]/2} = o(a_n^{(\gamma)}(d_1)a_n^{(\gamma)}(d_2)).$$

The same result can also be obtained in the boundary case $2\alpha/\zeta = d_2 - 1$, and thus we have proven (36). Finally, if $\gamma = 1/2$, the above calculations imply

$$D_n = O\left(a_n^{(\gamma)}(d_1)a_n^{(\gamma)}(d_2)\right), \quad n \rightarrow \infty,$$

and thus (12) follows.

We now proceed to proving (13). For $0 < \gamma < 1$, define $U_n^{(\gamma)} = S_n - S_n^{(\gamma)}$. Suppose $2\alpha/\zeta > d_{(k)}$. Appealing to (11), we obtain

$$\mathbb{E}(S_n^{(\gamma)}) \sim \left(\frac{2^{d-1}}{(d-1)\kappa_{d-2}}\right)^{k-1} \alpha^k \prod_{i=1}^k (\alpha - \zeta d_i/2)^{-1} n^k e^{-\zeta(d-1)(k-1)R_n/2},$$

because, for every $i = 1, \dots, k$, $a_n^{(\gamma)}(d_i)$ converges to $(d-1)^{-1}(\alpha - \zeta d_i/2)^{-1}$. Therefore, to complete our proof, it suffices to verify that

$$\mathbb{E}(U_n^{(\gamma)}) = o\left(n^k e^{-\zeta(d-1)(k-1)R_n/2}\right). \tag{37}$$

Recalling the notation (31), we apply the Mecke formula to obtain that

$$\begin{aligned} \mathbb{E}\left(U_n^{(\gamma)}\right) &= n^k \mathbb{E}(h_{n,\gamma}(X_1, \dots, X_k)) \\ &= n^k \mathbb{E}\left[h_{n,\gamma}(X_1, \dots, X_k) \prod_{(i,j) \in E} \mathbf{1}\{T_i + T_j \leq R_n - \omega_n\}\right] \\ &\quad + n^k \mathbb{E}\left[h_{n,\gamma}(X_1, \dots, X_k) \mathbf{1}\{\cup_{(i,j) \in E} \{T_i + T_j > R_n - \omega_n\}\}\right] =: A'_n + B'_n. \end{aligned}$$

We can calculate A'_n in almost the same manner as A_n ; the only difference is that when handling $\bar{\rho}_{n,\alpha}(\mathbf{t})$, we apply the inequality in Lemma 1(i) instead of Lemma 1(ii). We then obtain

$$\begin{aligned} A'_n &\leq C^* n^k e^{-\zeta(d-1)(k-1)R_n/2} \int_0^{R_n} dt_1 \dots \int_0^{R_n} dt_k \\ &\quad \times \mathbf{1}\{t_i > \gamma R_n \text{ for some } i = 1, \dots, k\} e^{2^{-1}\zeta(d-1)\sum_{i=1}^k (d_i - 2\alpha/\zeta)t_i} \\ &= o\left(n^k e^{-\zeta(d-1)(k-1)R_n/2}\right). \end{aligned}$$

Here, the second equality follows from the assumption that $d_i - 2\alpha/\zeta < 0$ for all $i = 1, \dots, k$. Furthermore, we can show that $B'_n = o\left(n^k e^{-\zeta(d-1)(k-1)R_n/2}\right)$; the proof of this is similar to that of the corresponding result for the derivation of (35) and (36), so we omit it. Thus, we have (37) as needed to complete the proof of the theorem. \square

Proof of Corollary 1. Define $S_n^{(\gamma)}$ as the sub-tree counts relating to a k -star with $d_{(k)} = k - 1$, $d_{(k-1)} = \dots = d_{(1)} = 1$:

$$S_n^{(\gamma)} = \sum_{(X_1, \dots, X_k) \in \mathcal{P}_{n, \neq}^k} \mathbf{1}\{0 < d_\zeta(X_1, X_i) \leq R_n, \quad i = 2, \dots, k, \quad T_i \leq \gamma R_n, \quad i = 1, \dots, k\}.$$

We also define $U_n^{(\gamma)} = S_n - S_n^{(\gamma)}$. First, (15) is a direct consequence of (13), so we shall prove only (ii) and (iii).

For the proof of (ii), we start by deriving a suitable upper bound for $\mathbb{E}(U_n^{(\gamma)})$. By the Mecke formula and Lemma 1(i),

$$\begin{aligned} \mathbb{E}(U_n^{(\gamma)}) &= n^k \mathbb{P}(d_\zeta(X_1, X_i) \leq R_n, i = 2, \dots, k, T_i > \gamma R_n \text{ for some } i = 1, \dots, k) \\ &\leq n^k k \mathbb{P}(T_1 > \gamma R_n) \leq C^* n^k e^{-\alpha(d-1)\gamma R_n} R_n. \end{aligned}$$

Taking the logarithm on both sides, we get

$$\limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}(U_n^{(\gamma)})}{R_n \vee \log n} \leq k(c \vee 1)^{-1} - \alpha(d-1)\gamma(1 \vee c^{-1})^{-1}.$$

On the other hand, it follows from Theorem 1 that

$$\mathbb{E}(S_n^{(\gamma)}) \sim C^* n^k e^{-\zeta(d-1)(k-1)R_n/2} a_n^{(\gamma)}(k-1) \sim C^* n^k e^{-\alpha(d-1)\gamma R_n - \zeta(d-1)(k-1)(1-\gamma)R_n/2},$$

which implies that, as $n \rightarrow \infty$,

$$\frac{\log \mathbb{E}(S_n^{(\gamma)})}{R_n \vee \log n} \rightarrow k(c \vee 1)^{-1} - \alpha(d-1)\gamma(1 \vee c^{-1})^{-1} - \zeta(d-1)(k-1)(1-\gamma)(1 \vee c^{-1})^{-1}/2.$$

Moreover, if $\limsup_{n \rightarrow \infty} \mathbb{E}(U_n^{(\gamma)}) / \mathbb{E}(S_n^{(\gamma)}) < \infty$,

$$\limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}(S_n)}{R_n \vee \log n} = \limsup_{n \rightarrow \infty} \frac{\log E(S_n^{(\gamma)})}{R_n \vee \log n},$$

and if $\limsup_{n \rightarrow \infty} \mathbb{E}(U_n^{(\gamma)}) / \mathbb{E}(S_n^{(\gamma)}) = \infty$,

$$\limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}(S_n)}{R_n \vee \log n} = \limsup_{n \rightarrow \infty} \frac{\log E(U_n^{(\gamma)})}{R_n \vee \log n}.$$

Therefore, using the obvious inequalities

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{\log \mathbb{E}(S_n^{(\gamma)})}{R_n \vee \log n} &\leq \liminf_{n \rightarrow \infty} \frac{\log \mathbb{E}(S_n)}{R_n \vee \log n} \leq \limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}(S_n)}{R_n \vee \log n} \\ &\leq \limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}(S_n^{(\gamma)})}{R_n \vee \log n} \vee \limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}(U_n^{(\gamma)})}{R_n \vee \log n}, \end{aligned}$$

and letting $\gamma \nearrow 1$, we obtain

$$\frac{\log \mathbb{E}(S_n)}{R_n \vee \log n} \rightarrow k(c \vee 1)^{-1} - \alpha(d-1)(1 \vee c^{-1})^{-1}, \text{ as } n \rightarrow \infty.$$

The proof of (iii) is very similar to that of (ii), so we omit it. □

Proof of Theorem 2. We start by writing

$$\begin{aligned} \mathbb{E}[(S_n^{(\gamma)})^2] &= \sum_{\ell=0}^k \mathbb{E} \left[\sum_{(X_1, \dots, X_{2k-\ell}) \in \mathcal{P}_{n, \neq}^{2k-\ell}} g_{n, \gamma}(X_1, \dots, X_k) \right. \\ &\quad \left. \times g_{n, \gamma}(X_1, \dots, X_\ell, X_{k+1}, \dots, X_{2k-\ell}) \right] =: \sum_{\ell=0}^k \mathbb{E}(\mathcal{I}_\ell), \end{aligned}$$

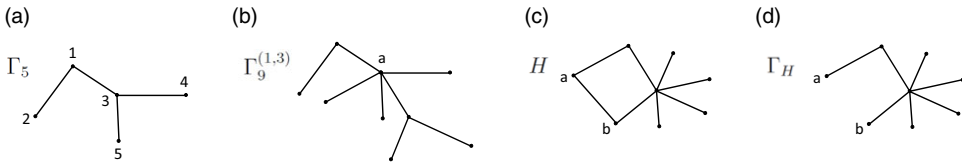


FIGURE 4. (a) Let $k = 5$, $i = 1$, and $j = 3$. Take two copies of Γ_5 . (b) Identify vertex 1 in one copy and vertex 3 in the other copy, and glue them together. The vertex ‘a’ denotes the glued vertex. The degree sequence of $\Gamma_9^{(1,3)}$ is $\{5, 2, 3, 1, 1, 1, 1, 1, 1\}$. (c) Consider the same Γ_5 as in (a) and set $\ell = 2$. Glue vertex 2 in one copy of Γ_5 to vertex 2 in the other copy. Do the same for vertex 3 for the two copies of Γ_5 ; this yields the subgraph H shown here. (d) Removing an edge (a, b) from H , we get one of the corresponding spanning trees.

where $g_{n,\gamma}$ is defined at (30). For $\ell = 0$, the Mecke formula yields

$$\mathbb{E}(I_0) = n^{2k} \left[\mathbb{E}(g_{n,\gamma}(\mathcal{X})) \right]^2 = \left[\mathbb{E}(S_n^{(\gamma)}) \right]^2.$$

Let $\Gamma_{2k-1}^{(i,j)}$ be a tree on $[2k - 1]$ (meaning $|E| = 2k - 2$) formed by taking two copies of Γ_k and identifying the vertex of degree d_i in one copy with the vertex of degree d_j in the other copy. In other words, the degree sequence of $\Gamma_{2k-1}^{(i,j)}$ is $d_i + d_j, d_i, d_j$ and a pair of d_ℓ ’s for $\ell \in [k] \setminus \{i, j\}$; see Figure 4(a)–(b).

We first note that $I_1 = \sum_{i,j=1}^k \mathcal{C}(\Gamma_{2k-1}^{(i,j)}, HG_n^{(\gamma)}(R_n; \alpha, \zeta))$, and so, from the identity for $\mathbb{E}(I_0)$ above, we have that

$$\text{Var}(S_n^{(\gamma)}) = \sum_{\ell=1}^k \mathbb{E}(I_\ell) \geq \mathbb{E}(I_1) \geq E \left[\mathcal{C}(\Gamma_{2k-1}^{(i',j')}, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right], \tag{38}$$

where $i' = j' = (k)$, that is, $d_{i'} = d_{j'} = d_{(k)}$. Therefore, applying Theorem 1 to (38), we derive that

$$\text{Var}(S_n^{(\gamma)}) = \Omega \left(n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} a_n^{(\gamma)}(2d_{(k)}) \prod_{i=1}^{k-1} a_n^{(\gamma)}(d_{(i)})^2 \right). \tag{39}$$

Note that owing to (32) and (34), the assumption $\gamma \neq 1/2$ is not required for the lower bound. Similarly, by using the bound $I_k \geq \mathcal{C}(\Gamma_k, HG_n^{(\gamma)}(R_n; \alpha, \zeta))$ and Theorem 1, we get

$$\text{Var}(S_n^{(\gamma)}) \geq \mathbb{E}(I_k) \geq \mathbb{E} \left[\mathcal{C}(\Gamma_k, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right] = O \left(n^k e^{-\zeta(d-1)(k-1)R_n/2} \prod_{i=1}^k a_n^{(\gamma)}(d_{(i)}) \right). \tag{40}$$

Finally, combining (39) and (40) establishes (16).

We now proceed to showing (17). Assume that $\alpha/\zeta > d_{(k)}$, and for $0 < \gamma < 1$, write $\text{Var}(S_n^{(\gamma)}) = \sum_{\ell=1}^k \mathbb{E}(I_\ell)$ as in (38). To derive exact asymptotics for $\text{Var}(S_n^{(\gamma)})$, we need to derive exact asymptotics for $\mathbb{E}(I_1)$ and $\mathbb{E}(I_k)$ and also show that $\mathbb{E}(I_\ell) = o(\mathbb{E}(I_1) \vee \mathbb{E}(I_k))$ for $2 \leq \ell \leq k - 1$.

From the representation of I_1 before (38) and Theorem 1, we have that

$$\begin{aligned} \mathbb{E}(I_1) &= \sum_{i,j=1}^k \mathbb{E} \left[\mathcal{C}(\Gamma_{2k-1}^{(i,j)}, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right] \\ &\sim C^* n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} \sum_{i,j=1}^k \prod_{\ell=1, \ell \neq i,j}^k a_n^{(\gamma)}(d_\ell)^2 a_n^{(\gamma)}(d_i + d_j) a_n^{(\gamma)}(d_i) a_n^{(\gamma)}(d_j). \end{aligned}$$

However, because of the constraint $\alpha/\zeta > d_{(k)}$, we see that

$$\prod_{\ell=1, \ell \neq i,j}^k a_n^{(\gamma)}(d_\ell)^2 a_n^{(\gamma)}(d_i + d_j) a_n^{(\gamma)}(d_i) a_n^{(\gamma)}(d_j)$$

tends to a positive constant for all $i, j = 1, \dots, k$. Thus, we conclude that

$$\mathbb{E}(I_1) \sim C^* n^{2k-1} e^{-\zeta(d-1)(k-1)R_n}, \text{ as } n \rightarrow \infty.$$

Similarly, we can verify that $\mathbb{E}(I_k) \sim C^* n^k e^{-\zeta(d-1)(k-1)R_n/2}$ as $n \rightarrow \infty$.

Next we investigate the rate of $\mathbb{E}(I_\ell)$ for $\ell = 2, \dots, k - 1$. Similarly to I_1 , we can express I_ℓ as a sum of $\mathcal{C}(H, HG_n^{(\gamma)}(R_n; \alpha, \zeta))$ over subgraphs H on $[2k - \ell]$ formed by identifying ℓ vertices on two copies of Γ_k . Choosing a spanning tree Γ_H of H , and using the monotonicity of \mathcal{C} , we get $\mathcal{C}(H, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \leq \mathcal{C}(\Gamma_H, HG_n^{(\gamma)}(R_n; \alpha, \zeta))$; see Figure 4(c)–(d). By the choice of H and Γ_H , the vertex degrees are necessarily smaller than α/ζ . Thus, from Theorem 1, we get that for every Γ_H ,

$$\mathbb{E} \left[\mathcal{C}(\Gamma_H, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right] \sim C^* n^{2k-\ell} e^{-\zeta(d-1)(2k-\ell-1)R_n/2}.$$

Now, we derive that

$$\mathbb{E}(I_\ell) = O(n^{2k-\ell} e^{-\zeta(d-1)(2k-\ell-1)R_n/2}), \quad \ell = 2, \dots, k - 1.$$

Note that for every $n \geq 1$, $g(m) = n^m e^{-\zeta(d-1)(m-1)R_n/2}$ is monotonic in $m \in \mathbb{N}_+$, so either $\mathbb{E}(I_1)$ or $\mathbb{E}(I_k)$ determines the actual growth rate of $\mathbb{V}ar(S_n^{(\gamma)})$. Thus, we have that $\mathbb{E}(I_\ell) = o(\mathbb{E}(I_1) \vee \mathbb{E}(I_k))$ for all $2 \leq \ell \leq k - 1$, and we can conclude that

$$\mathbb{V}ar(S_n^{(\gamma)}) \sim C^* \left[n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} \vee n^k e^{-\zeta(d-1)(k-1)R_n/2} \right], \quad n \rightarrow \infty.$$

Next, let $U_n^{(\gamma)} = S_n - S_n^{(\gamma)}$ for $0 < \gamma < 1$. We can finish the proof provided that

$$\mathbb{V}ar(U_n^{(\gamma)}) = o(n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} \vee n^k e^{-\zeta(d-1)(k-1)R_n/2}). \tag{41}$$

As in the proof for $\mathbb{V}ar(S_n^{(\gamma)})$, we write

$$\begin{aligned} \mathbb{V}ar(U_n^{(\gamma)}) &= \sum_{\ell=1}^k \mathbb{E} \left[\sum_{(X_1, \dots, X_{2k-\ell}) \in \mathcal{P}_{n, \neq}^{2k-\ell}} h_{n, \gamma}(X_1, \dots, X_k) \right. \\ &\quad \left. \times h_{n, \gamma}(X_1, \dots, X_\ell, X_{k+1}, \dots, X_{2k-\ell}) \right] =: \sum_{\ell=1}^k \mathbb{E}(J_\ell). \end{aligned}$$

Repeating the same argument as in the proof of (37) in Theorem 1, we obtain that for all $\ell = 1, \dots, k$,

$$\mathbb{E}(J_\ell) = o(n^{2k-\ell} e^{-\zeta(d-1)(2k-\ell-1)R_n/2}).$$

Thus, (41) follows. Moreover, by the Cauchy–Schwarz inequality,

$$|\text{Cov}(S_n^{(\gamma)}, U_n^{(\gamma)})| \leq \sqrt{\text{Var}(S_n^{(\gamma)})\text{Var}(U_n^{(\gamma)})} = o(n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} \vee n^k e^{-\zeta(d-1)(k-1)R_n/2}),$$

from which we have

$$\begin{aligned} \text{Var}(S_n) &= \text{Var}(S_n^{(\gamma)}) + \text{Var}(U_n^{(\gamma)}) + 2\text{Cov}(S_n^{(\gamma)}, U_n^{(\gamma)}) \\ &\sim C^* \left[n^{2k-1} e^{-\zeta(d-1)(k-1)R_n} \vee n^k e^{-\zeta(d-1)(k-1)R_n/2} \right], \quad n \rightarrow \infty. \quad \square \end{aligned}$$

3.3. Proof of the central limit theorem

The proof relies on the normal approximation bound derived from the Malliavin–Stein method in [32] (see Theorem 4 below). The Malliavin–Stein method has emerged as a crucial tool for the approximation of Gaussian, Poisson, and Rademacher functionals. An important aspect of the method is to express the terms in Stein’s equation in terms of the Malliavin operators and derive suitable estimates for the latter. The bound in [32] reduces estimates involving the Malliavin operators to those expressed in terms of more tractable difference operators. For a more detailed introduction to this subject, see [33, 37]. In our proof, we use only the first- and second-order difference operators. Our sub-tree counts are U-statistics, and one may express these difference operators in terms of $g_{n,\gamma}$ as defined in (30). The bulk of our work (see Lemmas 5–7) will be in deducing good estimates for these difference operators. By substituting the obtained estimates and the already derived variance estimates into the second-order Poincaré inequality, we obtain the required normal approximation.

This section is organized as follows. In Section 3.3.1 we introduce the normal approximation bound, as well as simplifications for our case. Next, in Section 3.3.2 we derive necessary estimates for difference operators, which we use in Section 3.3.3 to complete the proof of the CLT.

3.3.1. *Malliavin–Stein bound for Poisson functionals.* Let \mathcal{P} be a Poisson point process on a finite measure space \mathbf{X} with intensity measure $\lambda(\cdot)$. For a functional F of Radon counting measures (i.e., locally finite collection of points), define the *Malliavin difference operators* as follows. For $x \in \mathbf{X}$, the *first-order difference operator* is $D_x F := F(\mathcal{P} \cup \{x\}) - F(\mathcal{P})$. The *higher-order difference operators* are defined inductively as $D_{x_1, \dots, x_\ell}^\ell F := D_{x_\ell}(D_{x_1, \dots, x_{\ell-1}}^{\ell-1} F)$. We require only the first- and second-order difference operators. The latter is easily seen to be

$$D_{x,y}^2 F = F(\mathcal{P} \cup \{x, y\}) - F(\mathcal{P} \cup \{y\}) - F(\mathcal{P} \cup \{x\}) + F(\mathcal{P}),$$

for $x, y \in \mathbf{X}$. We say that $F \in \text{dom } D$ if

$$\mathbb{E}(F(\mathcal{P})^2) < \infty, \quad \mathbb{E} \left[\int_{\mathbf{X}} (D_x F(\mathcal{P}))^2 \lambda(dx) \right] < \infty.$$

The normal approximation bounds in [32, Theorems 1.1 and 1.2] are given in a compact form but involve the second moments of the product of second-order difference operators. A simplified version of [32, Theorems 1.1 and 1.2] is given in [32, Proposition 1.4], but it does not suffice for our purposes. Therefore, the bounds we shall use below are those obtained by combining [32, Theorems 1.1 and 1.2] and [32, Theorem 6.1].

Theorem 4. ([32, Theorems 1.1, 1.2, and 6.1]) *Let $F \in \text{dom } D$ and let N be a standard normal random variable. Define*

$$c_1 := \sup_{x \in \mathbf{X}} \mathbb{E}(|D_x F|^5), \quad c_2 := \sup_{x, y \in \mathbf{X}} \mathbb{E}(|D_{x,y}^2 F|^5),$$

where these supremums are essential supremums with respect to λ and λ^2 , respectively. Then

$$d_W \left(\frac{F - \mathbb{E}(F)}{\sqrt{\text{Var}(F)}}, N \right) \leq W_1 + W_2 + W_3,$$

$$d_K \left(\frac{F - \mathbb{E}(F)}{\sqrt{\text{Var}(F)}}, N \right) \leq W_1 + W_2 + W_3 + W_4 + W_5 + W_6,$$

where W_1, \dots, W_6 are defined as follows:

$$W_1 = \frac{2(c_1 c_2)^{1/5}}{\text{Var}(F)} \left[\int_{\mathbf{X}^3} \left[\mathbb{P}(D_{x_1, x_3}^2 F \neq 0) \mathbb{P}(D_{x_2, x_3}^2 F \neq 0) \right]^{1/20} \lambda^3(d(x_1, x_2, x_3)) \right]^{1/2},$$

$$W_2 = \frac{c_2^{2/5}}{\text{Var}(F)} \left[\int_{\mathbf{X}^3} \left[\mathbb{P}(D_{x_1, x_3}^2 F \neq 0) \mathbb{P}(D_{x_2, x_3}^2 F \neq 0) \right]^{1/10} \lambda^3(d(x_1, x_2, x_3)) \right]^{1/2},$$

$$W_3 = \frac{1}{\text{Var}(F)^{3/2}} \int_{\mathbf{X}} \mathbb{E}(|D_x F|^3) \lambda(dx),$$

$$W_4 = \frac{c_1^{3/5} \lambda(\mathbf{X})}{\text{Var}(F)^{3/2}} + \frac{c_1^{4/5} \lambda(\mathbf{X})^{5/4} + 2c_1^{4/5} \lambda(\mathbf{X})^{3/2}}{\text{Var}(F)^2},$$

$$W_5 = \frac{c_1^{2/5} \lambda(\mathbf{X})^{1/2}}{\text{Var}(F)},$$

$$W_6 = \frac{\sqrt{6}(c_1 c_2)^{1/5} + \sqrt{3}c_2^{2/5}}{\text{Var}(F)} \left[\int_{\mathbf{X}^2} \mathbb{P}(D_{x_1, x_2}^2 F \neq 0) \right]^{1/10} \lambda^2(d(x_1, x_2)) \right]^{1/2}.$$

For a self-contained proof, we state [32, Theorems 1.1 and 1.2] in the appendix, explaining how the above theorem can be deduced using the proof strategy of [32, Theorem 6.1].

In order to apply Theorem 4 to our setup, we take

$$F = S_n^{(\gamma)}, \quad \lambda(dx) = n \bar{\rho}_{n, \alpha}(t) \pi(\theta) \mathbf{1}\{t \leq \gamma R_n\} dt d\theta,$$

where x is represented by its hyperbolic polar coordinates (t, θ) .

In what follows, we write $g = g_{n, \gamma}$ (see (30)). Observe that $S_n^{(\gamma)}$ is a U -statistic and its Malliavin derivatives have a neat form as follows: for two distinct points $x, y \in B_d^{(c)}$,

$$D_x S_n^{(\gamma)} = \sum_{\ell=1}^k \sum_{(X_1, \dots, X_{k-1}) \in \mathcal{D}_{n, \neq}^{k-1}} g(X_1, \dots, x, \dots, X_{k-1}), \tag{42}$$

$$D_{x,y}^2 S_n^{(\gamma)} = \sum_{1 \leq \ell_1 < \ell_2 \leq k} \sum_{(X_1, \dots, X_{k-2}) \in \mathcal{D}_{n, \neq}^{k-2}} g(X_1, \dots, x, \dots, y, \dots, X_{k-2}) \tag{43}$$

$$+ \sum_{1 \leq \ell_2 < \ell_1 \leq k} \sum_{(X_1, \dots, X_{k-2}) \in \mathcal{D}_{n, \neq}^{k-2}} g(X_1, \dots, y, \dots, x, \dots, X_{k-2}),$$

where ℓ and the ℓ_i denote the positions of the corresponding coordinates. In particular, (42) represents the number of injective homomorphisms from Γ_k to $HG_n^{(\gamma)}(R_n; \alpha, \zeta)$, with one of its vertices fixed at x and the remaining $k - 1$ points taken from \mathcal{P}_n . Similarly, (43) represents the number of injective homomorphisms, but here, two of the vertices are fixed at x and y , respectively.

From (42) and (43), the constants c_1, c_2 in Theorem 4 have become

$$c_{1,n} := \sup_{x \in D_\gamma(R_n)} \mathbb{E}[|D_x S_n^{(\gamma)}|^5], \quad c_{2,n} := \sup_{x,y \in D_\gamma(R_n)} \mathbb{E}[|D_{x,y}^2 S_n^{(\gamma)}|^5].$$

For notational convenience later on, we also define

$$c_{3,n} := \int_0^{\gamma R_n} \int_{C_d} \mathbb{E}[|D_x S_n^{(\gamma)}|^3] \bar{\rho}_{n,\alpha}(t) \pi(\theta) dt d\theta,$$

where x is again represented by its hyperbolic polar coordinates (t, θ) . With this notation and the preliminary facts now available, we can rephrase $W_i, i = 1, \dots, 6$, in Theorem 4 as follows:

$$\begin{aligned} W_1 &= \frac{2n^{3/2}(c_{1,n}c_{2,n})^{1/5}}{\text{Var}(S_n^{(\gamma)})} \\ &\quad \times \left[\int_{[0,\gamma R_n]^3 \times C_d^3} [\mathbb{P}(D_{x_1,x_3}^2 S_n^{(\gamma)} \neq 0) \mathbb{P}(D_{x_2,x_3}^2 S_n^{(\gamma)} \neq 0)]^{1/20} \prod_{i=1}^3 \bar{\rho}_{n,\alpha}(t_i) \pi(\theta_i) dt_i d\theta_i \right]^{1/2}, \\ W_2 &= \frac{n^{3/2}(c_{2,n})^{2/5}}{\text{Var}(S_n^{(\gamma)})} \\ &\quad \times \left[\int_{[0,\gamma R_n]^3 \times C_d^3} [\mathbb{P}(D_{x_1,x_3}^2 S_n^{(\gamma)} \neq 0) \mathbb{P}(D_{x_2,x_3}^2 S_n^{(\gamma)} \neq 0)]^{1/10} \prod_{i=1}^3 \bar{\rho}_{n,\alpha}(t_i) \pi(\theta_i) dt_i d\theta_i \right]^{1/2}, \\ W_3 &= \frac{nc_{3,n}}{\text{Var}(S_n^{(\gamma)})^{3/2}}, \\ W_4 &= \frac{n(c_{1,n})^{3/5}}{\text{Var}(S_n^{(\gamma)})^{3/2}} + \frac{n^{5/4}(c_{1,n})^{4/5} + 2n^{3/2}(c_{1,n})^{4/5}}{\text{Var}(S_n^{(\gamma)})^2}, \\ W_5 &= \frac{n^{1/2}(c_{1,n})^{2/5}}{\text{Var}(S_n^{(\gamma)})}, \\ W_6 &= \frac{n\{\sqrt{6}(c_{1,n}c_{2,n})^{1/5} + \sqrt{3}(c_{2,n})^{2/5}\}}{\text{Var}(S_n^{(\gamma)})} \\ &\quad \times \left[\int_{[0,\gamma R_n]^2 \times C_d^2} [\mathbb{P}(D_{x_1,x_2}^2 S_n^{(\gamma)} \neq 0)]^{1/10} \prod_{i=1}^2 \bar{\rho}_{n,\alpha}(t_i) \pi(\theta_i) dt_i d\theta_i \right]^{1/2}, \end{aligned} \tag{44}$$

where we identify the terms x_i with their hyperbolic polar coordinates (t_i, θ_i) .

3.3.2. *Some auxiliary lemmas for the proof of the central limit theorem.* Having already derived variance bounds in Theorem 2, we now deduce necessary estimates for the W_i terms in (44). More specifically, Lemmas 5 and 6 below are used to estimate $c_{1,n}, c_{2,n}$, and $c_{3,n}$, while Lemma

7 calculates the rate of the remaining integral terms. These estimates are then substituted into the W_i terms in (44). Recall that C^* is a positive generic constant, which is independent of n .

Lemma 5. *Let the assumptions of Theorem 3 hold, and let $\gamma \in (0, 1/2)$. Set $\rho_n := ne^{-\zeta(d-1)R_n/2}$. For $p \geq 1$, define $\mathcal{G}_p^{(1)}$ and $\mathcal{G}_p^{(2)}$ respectively to be the set of all trees on $\{0\} \cup [p]$ and $\{-1, 0\} \cup [p]$. For a tree $T \in \mathcal{G}_p^{(1)}$, let d'_0, d'_1, \dots, d'_p be the degree sequence, and similarly, for a tree $T \in \mathcal{G}_p^{(2)}$, let $d'_{-1}, d'_0, \dots, d'_p$ be the degree sequence. Then we have*

$$c_{1,n} = O\left(\rho_n^{5(k-1)} c'_{1,n}\right), \tag{45}$$

$$c_{2,n} = O\left(\rho_n^{5(k-2)} c'_{2,n}\right), \tag{46}$$

where

$$c'_{1,n} := \max_{\substack{p=k-1, \dots, 5(k-1), \\ T \in \mathcal{G}_p^{(1)}}} e^{\zeta(d-1)d'_0 \gamma R_n/2} \prod_{i=1}^p a_n^{(\gamma)}(d'_i),$$

$$c'_{2,n} := \max_{\substack{p=k-2, \dots, 5(k-2), \\ T \in \mathcal{G}_p^{(2)}}} e^{\zeta(d-1)(d'_{-1}+d'_0) \gamma R_n/2} \prod_{i=1}^p a_n^{(\gamma)}(d'_i).$$

Lemma 6. *In the notation of Lemma 5, for $0 < \gamma < 1/2$,*

$$c_{3,n} = O\left(\rho_n^{3(k-1)} c'_{3,n}\right),$$

where

$$c'_{3,n} := \max_{\substack{p=k-1, \dots, 3(k-1), \\ T \in \mathcal{G}_p^{(1)}}} \prod_{i=0}^p a_n^{(\gamma)}(d'_i).$$

Lemma 7. *For $0 < \gamma < 1/2$ and $0 < a < 1$, we have*

$$\int_{[0, \gamma R_n]^3 \times C_d^3} \left[\mathbb{P}(D_{x_1, x_3}^2 S_n^{(\gamma)} \neq 0) \mathbb{P}(D_{x_2, x_3}^2 S_n^{(\gamma)} \neq 0) \right]^a$$

$$\times \prod_{i=1}^3 \bar{\rho}_{n, \alpha}(t_i) \pi(\theta_i) dt_i d\theta_i = O(e^{-\zeta(d-1)(1-2\gamma)R_n}),$$

$$\int_{[0, \gamma R_n]^2 \times C_d^2} \left[\mathbb{P}(D_{x_1, x_2}^2 S_n^{(\gamma)} \neq 0) \right]^a \prod_{i=1}^2 \bar{\rho}_{n, \alpha}(t_i) \pi(\theta_i) dt_i d\theta_i = O(e^{-\zeta(d-1)(1-2\gamma)R_n/2}).$$

Remark 3.

- (i) One can make the growth rate of the $c'_{i,n}$ as slow as one likes by choosing γ small enough. To make this a little more clear, assume, for simplicity, that $2\alpha/\zeta$ is not an integer. Then, as $n \rightarrow \infty$,

$$c'_{1,n} \sim \max_{\substack{p=k-1, \dots, 5(k-1), \\ T \in \mathcal{G}_p^{(1)}}} \prod_{j=1}^p \left| (d-1)(\alpha - \zeta d'_j/2) \right|^{-1} e^{\zeta(d-1)[d'_0 + \sum_{i=1}^p (d'_i - 2\alpha/\zeta)_+]} \gamma R_n/2.$$

Note also that $\rho_n \rightarrow c \in (0, \infty]$ implies $R_n = O(\log n)$. Therefore, for any $\epsilon > 0$, there exists $\gamma_0 > 0$ such that for all $0 < \gamma < \gamma_0$, we have that $c'_{1,n} = o(n^\epsilon)$. The same claim holds for $c'_{2,n}, c'_{3,n}$ as well.

- (ii) The separation of each $c_{i,n}$ into $c'_{i,n}$ and ρ_n terms is because $c'_{i,n}$ depends on γ , whereas ρ_n does not.
- (iii) In many applications to Euclidean stochastic geometric functionals (see [32, Section 7]), $c_{1,n}, c_{2,n}$ are shown to be bounded in n , whereas in our case, $c_{1,n}, c_{2,n}$ are unbounded in n . This is why it is challenging to obtain optimal Berry–Esseen bounds in our normal approximation result.

Proof of Lemma 5. Fix $x \in B_d^{(\zeta)}$. For $p = k - 1, \dots, 5(k - 1)$, let $\Sigma_{5(k-1),p}$ denote the set of all surjective maps from $[5(k - 1)]$ to $[p]$. From (42), we have that

$$\begin{aligned} |D_x S_n^{(\gamma)}|^5 &\leq C^* \sum_{\ell=1}^k \left(\sum_{(X_1, \dots, X_{k-1}) \in \mathcal{P}_{n, \neq}^{k-1}} g(X_1, \dots, x, \dots, X_{k-1}) \right)^5 \\ &= C^{*k} \sum_{\ell=1}^k \sum_{p=k-1}^{5(k-1)} \frac{1}{p!} \sum_{\sigma \in \Sigma_{5(k-1),p}} \sum_{(X_1, \dots, X_p) \in \mathcal{P}_{n, \neq}^p} \\ &\quad \times \prod_{i=1}^5 g(X_{\sigma((i-1)(k-1)+1)}, \dots, x, \dots, X_{\sigma(i(k-1))}). \end{aligned}$$

It is possible that under some surjections σ , the coordinates in $(\sigma((i - 1)(k - 1) + 1), \dots, \sigma(i(k - 1)))$ may repeat for some i , but in such cases, $g = 0$ by definition; thus, $\Sigma_{5(k-1),p}$ in the last expression can be replaced with

$$\begin{aligned} \Sigma_{5(k-1),p}^* &= \{ \sigma \in \Sigma_{5(k-1),p} : \sigma((i - 1)(k - 1) + 1), \dots, \sigma(i(k - 1)) \\ &\quad \text{are distinct for all } i = 1, \dots, 5 \}. \end{aligned}$$

Now, let us fix $\ell = 1$, without loss of generality, and $p \in \{k - 1, \dots, 5(k - 1)\}$, $\sigma \in \Sigma_{5(k-1),p}^*$; then we shall bound

$$A_{n,p,\sigma} := \mathbb{E} \left[\sum_{(X_1, \dots, X_p) \in \mathcal{P}_{n, \neq}^p} \prod_{i=1}^5 g(x, X_{\sigma((i-1)(k-1)+1)}, \dots, X_{\sigma(i(k-1))}) \right]. \tag{47}$$

Let G_σ be a simple graph on $\{0\} \cup [p]$ with the edge set defined as follows: for every $i = 1, \dots, 5$, define $(\sigma((i - 1)(k - 1) + j_1), \sigma((i - 1)(k - 1) + j_2)) \in E_\sigma$ if $(j_1 + 1, j_2 + 1)$ is an edge in Γ_k . Similarly we say that $(0, \sigma((i - 1)(k - 1) + j)) \in E_\sigma$ if $(1, j + 1)$ is an edge in Γ_k . Then, setting $X_0 = x$, we have that G_σ is the graph counted by the summand in (47), i.e.,

$$\prod_{i=1}^5 g(x, X_{\sigma((i-1)(k-1)+1)}, \dots, X_{\sigma(i(k-1))}) = \prod_{(i,j) \in G_\sigma} \mathbf{1}\{d_\zeta(X_i, X_j) \leq R_n\}. \tag{48}$$

Now, the surjectivity of σ implies that G_σ is connected, and thus we can always find a spanning tree G'_σ of G_σ on $\{0\} \cup [p]$. Let d'_0, \dots, d'_p be the degree sequence of G'_σ , and let E'_σ

be the edge set of G'_σ . By the definitions of $A_{n,p,\sigma}$, G_σ , G'_σ , together with the monotonicity of \mathcal{C} and the above identity, we have that

$$A_{n,p,\sigma} = \mathbb{E} \left[\mathcal{C}(G_\sigma, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right] \leq \mathbb{E} \left[\mathcal{C}(G'_\sigma, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right] \\ = n^p \mathbb{P} \left(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E'_\sigma, T_i \leq \gamma R_n, i = 1, \dots, p, t_0 \leq \gamma R_n \right), \tag{49}$$

where $t_0 = R_n - d_\zeta(0, x)$ is deterministic, and the Mecke formula is applied at the last equality. Note that $|E'_\sigma| = p$.

Proceeding as in the derivation of (34) and noting that $T_i + T_j \leq R_n - \omega_n$ with $\omega_n = \log \log R_n$ always holds as we are taking $\gamma < 1/2$, we derive that

$$A_{n,p,\sigma} \leq n^p \int_0^{\gamma R_n} dt_1 \cdots \int_0^{\gamma R_n} dt_p \mathbf{1}_{\{t_0 \leq \gamma R_n\}} \prod_{(i,j) \in E'_\sigma} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{t}) \bar{\rho}_{n,\alpha}(\mathbf{t}) \\ \sim \left(\frac{2^{d-1}}{\kappa_{d-2}} \right)^p \alpha^p n^p e^{-\zeta(d-1)pR_n/2} e^{\zeta(d-1)d'_0 t_0/2} \mathbf{1}_{\{t_0 \leq \gamma R_n\}} \prod_{i=1}^p a_n^{(\gamma)}(d'_i) \\ = O(\rho_n^p e^{\zeta(d-1)d'_0 \gamma R_n/2} \prod_{i=1}^p a_n^{(\gamma)}(d'_i)). \tag{50}$$

Now, let us take the maximum of $A_{n,p,\sigma}$ over all $p \in \{k-1, \dots, 5(k-1)\}$, $\sigma \in \Sigma_{5(k-1),p}^*$, and the corresponding spanning tree G'_σ of a degree sequence d'_0, d'_1, \dots, d'_p . Equivalently, we take the maximum of $A_{n,p,\sigma}$ over all $p \in \{k-1, \dots, 5(k-1)\}$ and $T \in \mathcal{G}_p^{(1)}$, to conclude that

$$c_{1,n} = O \left(\max_{\substack{p=k-1, \dots, 5(k-1), \\ T \in \mathcal{G}_p^{(1)}}} \rho_n^p e^{\zeta(d-1)d'_0 \gamma R_n/2} \prod_{i=1}^p a_n^{(\gamma)}(d'_i) \right). \tag{51}$$

If $\rho_n \rightarrow \infty$, then clearly $\rho_n^p = O(\rho_n^{5(k-1)})$ for all $k-1 \leq p \leq 5(k-1)$, and hence the bound (45) holds trivially. Otherwise, $\rho_n \rightarrow c \in (0, \infty)$, but we can still easily get (45).

Now we shall derive the bound for $c_{2,n}$ in (46). For $p \in \{k-2, \dots, 5(k-2)\}$, let

$$\Sigma_{5(k-2),p}^* = \left\{ \sigma \in \Sigma_{5(k-2),p} : \sigma((i-1)(k-2)+1), \dots, \sigma(i(k-2)) \text{ are distinct for all } i = 1, \dots, 5 \right\}.$$

Then, arguing as in the derivation of the bounds for $c_{1,n}$, we find that the task of bounding $\mathbb{E}[|D_{x,y} S_n^{(\gamma)}|^5]$ is again reduced to that of bounding

$$B_{n,p,\sigma} := \mathbb{E} \left[\sum_{(X_1, \dots, X_p) \in \mathcal{P}_{n,\neq}^p} \prod_{i=1}^5 g(x, y, X_{\sigma((i-1)(k-2)+1)}, \dots, X_{\sigma(i(k-2))}) \right]$$

for all $p \in \{k-2, \dots, 5(k-2)\}$ and $\sigma \in \Sigma_{5(k-2),p}^*$.

Setting $X_{-1} = x, X_0 = y$ respectively, we can define a simple graph G_σ on $\{-1, 0\} \cup [p]$ as in (48) such that

$$\prod_{i=1}^5 g(x, y, X_{\sigma((i-1)(k-2)+1)}, \dots, X_{\sigma(i(k-2))}) = \prod_{(i,j) \in G_\sigma} \mathbf{1}_{\{d_\zeta(X_i, X_j) \leq R_n\}}.$$

Let G'_σ be a spanning tree of G_σ , which exists as G_σ is connected by the surjectivity of σ . Let $d'_{-1}, d'_0, \dots, d'_p$ be the respective degrees of vertices $\{-1, 0\} \cup [p]$ in G'_σ , and let E'_σ be its edge set. Note that $|E'_\sigma| = p + 1$.

Then, setting $t_{-1} = R_n - d_\zeta(0, x)$ and $t_0 = R_n - d_\zeta(0, y)$ (which are deterministic), we again derive that

$$B_{n,p,\sigma} \leq \mathbb{E} \left[\mathcal{C}(G'_\sigma, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right] \\ = n^p \mathbb{P} \left(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E'_\sigma, T_i \leq \gamma R_n, i = 1, \dots, p, t_{-1}, t_0 \leq \gamma R_n \right). \tag{52}$$

We basically proceed again as in the derivation of (34) by conditioning on the T_i , but here, we need to account for one extra complication, i.e., whether $(-1, 0) \in E'_\sigma$ or not. If $(-1, 0) \in E'_\sigma$, then whether $X_{-1} = x$ and $X_0 = y$ are connected is purely deterministic and the randomness arises only in the remaining p edges. In that case, the rightmost term in (52) is bounded by

$$n^p \int_0^{\gamma R_n} dt_1 \cdots \int_0^{\gamma R_n} dt_p \mathbf{1}_{\{t_{-1}, t_0 \leq \gamma R_n\}} \prod_{(i,j) \in E'_\sigma \setminus \{(-1,0)\}} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{t}) \bar{\rho}_{n,\alpha}(\mathbf{t}) \\ \sim C^* n^p e^{-\zeta(d-1)pR_n/2+2^{-1}\zeta(d-1)(d'_{-1}-1)t_{-1}+(d'_0-1)t_0} \mathbf{1}_{\{t_{-1}, t_0 \leq \gamma R_n\}} \prod_{i=1}^p a_n^{(\gamma)}(d'_i) \\ = O \left(\rho_n^p e^{\zeta(d-1)(d'_{-1}+d'_0)\gamma R_n/2-\zeta(d-1)\gamma R_n} \prod_{i=1}^p a_n^{(\gamma)}(d'_i) \right). \tag{53}$$

On the other hand, if $(-1, 0) \notin E'_\sigma$, the randomness arises in all $p + 1$ edges of E'_σ . Then the rightmost term in (52) is bounded by

$$n^p \int_0^{\gamma R_n} dt_1 \cdots \int_0^{\gamma R_n} dt_p \mathbf{1}_{\{t_{-1}, t_0 \leq \gamma R_n\}} \prod_{(i,j) \in E'_\sigma} \mathbb{P}(d_\zeta(X_i, X_j) \leq R_n \mid \mathbf{t}) \bar{\rho}_{n,\alpha}(\mathbf{t}) \\ \sim C^* n^p e^{-\zeta(d-1)(p+1)R_n/2+2^{-1}\zeta(d-1)(d'_{-1}t_{-1}+d'_0t_0)} \mathbf{1}_{\{t_{-1}, t_0 \leq \gamma R_n\}} \prod_{i=1}^p a_n^{(\gamma)}(d'_i) \\ = O \left(\rho_n^p e^{-\zeta(d-1)R_n/2+\zeta(d-1)(d'_{-1}+d'_0)\gamma R_n/2} \prod_{i=1}^p a_n^{(\gamma)}(d'_i) \right). \tag{54}$$

Combining (53) and (54), we conclude that

$$B_{n,p,\sigma} = O \left(\rho_n^p e^{\zeta(d-1)(d'_{-1}+d'_0)\gamma R_n/2} \prod_{i=1}^p a_n^{(\gamma)}(d'_i) \right).$$

Now, proceeding as in the derivation of (45) (see below (51)), we get (46). □

Proof of Lemma 6. From the same reasoning as in Lemma 5, it suffices to bound

$$C_{n,p,\sigma} := \int_0^{\gamma R_n} \int_{C_d} \mathbb{E} \left[\sum_{(X_1, \dots, X_p) \in \mathcal{P}_{n,\neq}^p} \prod_{i=1}^3 g(x, X_{\sigma((i-1)(k-1)+1)}, \dots, X_{\sigma(i(k-1))}) \right] \bar{\rho}_{n,\alpha}(t) \pi(\theta) d\theta dt$$

for every $p \in \{k - 1, \dots, 3(k - 1)\}$ and $\sigma \in \Sigma_{3(k-1),p}^*$.

Let G_σ be the same simple graph on $\{0\} \cup [p]$ as that constructed in the proof of Lemma 5(i) (see (48)), for which x is identified as a vertex ‘0’. Once again, let G'_σ be a spanning tree of G_σ , with d'_0, \dots, d'_p the degree sequence and E'_σ the edge set of G'_σ . It follows from the monotonicity of \mathcal{C} and the Mecke formula that

$$C_{n,p,\sigma} \leq \mathbb{E} \left[\mathcal{C}(G'_\sigma, HG_n^{(\gamma)}(R_n; \alpha, \zeta)) \right] \\ = n^p \mathbb{P} \left(d_\zeta(X_i, X_j) \leq R_n, (i, j) \in E'_\sigma, T_i \leq \gamma R_n, i = 0, \dots, p \right). \tag{55}$$

For further calculations, we note that X_0 in (55) is random, whereas $X_0 = x$ in (49) was purely deterministic. Taking into consideration this difference and using Theorem 1 with $k = p + 1$, we have

$$C_{n,p,\sigma} \leq C^* \rho_n^p \prod_{i=0}^p a_n^{(\gamma)}(d'_i) = O \left(\rho_n^{3(k-1)} \prod_{i=0}^p a_n^{(\gamma)}(d'_i) \right).$$

Finally, taking the maximum completes the proof. □

Proof of Lemma 7. We first need the following fact. For $0 < \gamma < 1/2$, let $HG(\mathcal{X})$ be a hyperbolic geometric graph on a point set $\mathcal{X} \subset D_\gamma(R_n)$, connecting any two points within a distance R_n . Suppose that $y_1, y_2 \in \mathcal{X}$ are connected by a path of length $\ell \geq 1$ in $HG(\mathcal{X})$, and their hyperbolic distances from the boundary, given by $t_i = R_n - d_\zeta(0, y_i)$, $i = 1, 2$, satisfy $t_1, t_2 \leq \gamma R_n$. For the relative angle θ_{12} between y_1 and y_2 , we claim that

$$\theta_{12} \leq (1 + o(1)) 2\ell e^{-\zeta(1-2\gamma)R_n/2}, \quad n \rightarrow \infty, \tag{56}$$

uniformly for $t_1, t_2 \leq \gamma R_n$.

The proof of (56) can be done inductively. For $\ell = 1$, set $\hat{\theta}_{12} = (e^{-2\zeta(R_n-t_1)} + e^{-2\zeta(R_n-t_2)})^{1/2}$ as in Lemma 3. Since $t_1 + t_2 \leq 2\gamma R_n < R_n - \omega_n$, we get

$$\hat{\theta}_{12} = o(e^{-\zeta(R_n-t_1-t_2)/2}) \rightarrow 0, \quad n \rightarrow \infty, \tag{57}$$

as in the proof of Lemma 4.

If $\theta_{12} \leq \hat{\theta}_{12}$, then $\theta_{12} \leq (1 + o(1))e^{-\zeta(1-2\gamma)R_n/2}$ by (57) with $t_1, t_2 \leq \gamma R_n$, and so (56) holds for $\ell = 1$. If $\theta_{12} \gg \hat{\theta}_{12}$, Lemma 3 yields the following: uniformly for $t_i \leq \gamma R_n$, $i = 1, 2$, we have that

$$R_n \geq d_\zeta(y_1, y_2) = 2R_n - (t_1 + t_2) + \frac{2}{\zeta} \log \sin \left(\frac{\theta_{12}}{2} \right) + o(1) \\ \geq 2(1 - \gamma)R_n + \frac{2}{\zeta} \log \sin \left(\frac{\theta_{12}}{2} \right) + o(1), \quad n \rightarrow \infty.$$

Equivalently, we get that $\sin(\theta_{12}/2) \leq (1 + o(1))e^{-\zeta(1-2\gamma)R_n/2}$. As the right-hand side tends to 0 as $n \rightarrow \infty$, we have, uniformly for $t_i \leq \gamma R_n$, $i = 1, 2$,

$$\theta_{12} \leq (1 + o(1)) 2e^{-\zeta(1-2\gamma)R_n/2}, \quad n \rightarrow \infty.$$

Hence, in either case, the claim for $\ell = 1$ follows.

Now, suppose that the claim holds for $\ell - 1$. Then, if y_1, y_2 have a path of length ℓ , there exists a y_0 such that y_1 and y_0 have a path of length $\ell - 1$, and y_0 and y_2 have a path of length 1.

Denoting the corresponding relative angles by θ_{10} and θ_{02} , we see that $\theta_{12} \leq \theta_{10} + \theta_{02}$; hence, the proof of (56) can be completed by the induction hypothesis.

Fix x_1, x_2, x_3 such that $t_i \leq \gamma R_n$ for $i = 1, 2, 3$. From (43), $D_{x_1, x_3}^2 S_n^{(\gamma)} \neq 0$ implies that there is a path of length at most $\text{diam}(\Gamma_k)$ (i.e., the diameter of the graph) from x_1 to x_3 in the hyperbolic random geometric graph on $(\mathcal{P}_n \cap D_\gamma(R_n)) \cup \{x_1, x_3\}$ with radius of connectivity R_n . Thus, from (56) and $\text{diam}(\Gamma_k) \leq k$, we have that

$$\theta_{13} \leq (1 + o(1))2ke^{-\zeta(1-2\gamma)R_n/2}, \quad n \rightarrow \infty.$$

Therefore,

$$\begin{aligned} & \int_{[0, \gamma R_n]^3 \times C_d^3} [\mathbb{P}(D_{x_1, x_3}^2 S_n^{(\gamma)} \neq 0) \mathbb{P}(D_{x_2, x_3}^2 S_n^{(\gamma)} \neq 0)]^a \prod_{i=1}^3 \bar{\rho}_{n, \alpha}(t_i) \pi(\theta_i) dt_i d\theta_i \\ & \leq C^* \int_{[0, \gamma R_n]^3 \times C_d^3} \mathbf{1}\{\theta_{j3} \leq 2ke^{-\zeta(1-2\gamma)R_n/2}, j = 1, 2\} \prod_{i=1}^3 \bar{\rho}_{n, \alpha}(t_i) \pi(\theta_i) dt_i d\theta_i \\ & = C^* \mathbb{P}(\Theta_{j3} \leq 2ke^{-\zeta(1-2\gamma)R_n/2}, j = 1, 2, T_i \leq \gamma R_n, i = 1, 2, 3), \end{aligned}$$

where Θ_{j3} denotes the relative angle between X_j and X_3 , $j = 1, 2$, and $T_i = R_n - d_\zeta(0, X_i)$ for $i = 1, 2, 3$.

Now, the probability of the last term equals

$$\int_{[0, \gamma R_n]^3} \prod_{j=1}^2 \mathbb{P}(\Theta_{j3} \leq 2ke^{-\zeta(1-2\gamma)R_n/2} \mid t_1, t_2, t_3) \prod_{i=1}^3 \bar{\rho}_{n, \alpha}(t_i) dt_i. \tag{58}$$

Using the density (28) of a relative angle, it is easy to see that

$$\prod_{j=1}^2 \mathbb{P}(\Theta_{j3} \leq 2ke^{-\zeta(1-2\gamma)R_n/2} \mid t_1, t_2, t_3) \sim \left(\frac{(2k)^{d-1}}{(d-1)\kappa_{d-2}} \right)^2 e^{-\zeta(d-1)(1-2\gamma)R_n}, \quad n \rightarrow \infty,$$

uniformly for $t_i \leq \gamma R_n$, $i = 1, 2, 3$. It now follows from Lemma 1(ii) that (58) is asymptotically equal to

$$C^* e^{-\zeta(d-1)(1-2\gamma)R_n} \left(\int_0^{\gamma R_n} e^{-\alpha(d-1)t} dt \right)^3 = O(e^{-\zeta(d-1)(1-2\gamma)R_n}).$$

This proves the first statement in the lemma. By exactly the same argument, we obtain the second statement in the lemma. □

3.3.3. *Proof of the central limit theorem in Theorem 3.* We now put together all the bounds and prove our main CLT.

Proof of Theorem 3. From (16), we have

$$\text{Var}(S_n^{(\gamma)}) = \Omega \left(n \rho_n^{2(k-1)} a_n^{(\gamma)} (2d(k)) \prod_{i=1}^{k-1} a_n^{(\gamma)}(d(i))^2 \right).$$

Substituting the lower bound for variance above and the bounds in Lemmas 5–7 into (44), and using the definition of ρ_n , we obtain

$$\begin{aligned}
 W_1 &\leq C^* n^{-1/2} \frac{(c'_{1,n} c'_{2,n})^{1/5} e^{\gamma \zeta (d-1) R_n}}{a_n^{(\gamma)} (2d(k)) \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^2}, \\
 W_2 &\leq C^* n^{-1/2} \rho_n^{-1} \frac{(c'_{2,n})^{2/5} e^{\gamma \zeta (d-1) R_n}}{a_n^{(\gamma)} (2d(k)) \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^2}, \\
 W_3 &\leq C^* n^{-1/2} \frac{c'_{3,n}}{a_n^{(\gamma)} (2d(k))^{3/2} \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^3}.
 \end{aligned}$$

From Theorem 4, we know that

$$d_W \left(\frac{S_n^{(\gamma)} - \mathbb{E} \left(S_n^{(\gamma)} \right)}{\sqrt{\text{Var}(S_n^{(\gamma)})}}, N \right) \leq W_1 + W_2 + W_3.$$

By the definition of $a_n^{(\gamma)}$, one can see that

$$\begin{aligned}
 c'_{1,n} &\leq C^* \max_{\substack{p=k-1, \dots, 5(k-1), \\ T \in \mathcal{G}_p^{(1)}}} e^{2^{-1} \zeta (d-1) [d'_0 + \sum_{i=1}^p (d'_i - 2\alpha/\zeta)_+] \gamma R_n}, \\
 c'_{2,n} &\leq C^* \max_{\substack{p=k-2, \dots, 5(k-2), \\ T \in \mathcal{G}_p^{(2)}}} e^{2^{-1} \zeta (d-1) [d'_{-1} + d'_0 + \sum_{i=1}^p (d'_i - 2\alpha/\zeta)_+] \gamma R_n}, \\
 c'_{3,n} &\leq C^* \max_{\substack{p=k-1, \dots, 3(k-1), \\ T \in \mathcal{G}_p^{(1)}}} e^{2^{-1} \zeta (d-1) \sum_{i=0}^p (d'_i - 2\alpha/\zeta)_+ \gamma R_n},
 \end{aligned}$$

$$a_n^{(\gamma)} (2d(k)) \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^2 \geq C^* e^{2^{-1} \zeta (d-1) [2(d_{(k)} - \alpha/\zeta)_+ + \sum_{i=1}^{k-1} (d_{(i)} - 2\alpha/\zeta)_+] \gamma R_n}.$$

Note also that (18) ensures $R_n \leq C^* \log n$. Substituting all of these bounds back into $W_1, W_2,$ and W_3 above, we find that

$$W_1 \leq C^* n^{-1/2+c_1\gamma}, \quad W_2 \leq C^* n^{-1/2+c_2\gamma} \rho_n^{-1}, \quad W_3 \leq C^* n^{-1/2+c_3\gamma},$$

for some positive constants $c_i, i = 1, 2, 3$. Hence, for any $a < 1/2$, we can choose γ_0 so small that $W_1 + W_2 + W_3 = O(n^{-a})$ as $n \rightarrow \infty$, for all $0 < \gamma < \gamma_0$. This proves the Wasserstein bound in (19).

To prove the Kolmogorov bound in (19), again using the bounds in Theorem 4, along with Lemmas 5, 6, and 7 and the variance lower bound above, we derive using (44) that

$$\begin{aligned}
 W_4 &\leq C^* n^{-1/2} \left[\frac{(c'_{1,n})^{3/5}}{a_n^{(\gamma)} (2d(k))^{3/2} \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^3} + \frac{(c'_{1,n})^{4/5}}{a_n^{(\gamma)} (2d(k))^2 \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^4} \right], \\
 W_5 &\leq C^* n^{-1/2} \frac{(c'_{1,n})^{2/5}}{a_n^{(\gamma)} (2d(k)) \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^2}, \\
 W_6 &\leq C^* (n\rho_n)^{-1/2} \frac{e^{\gamma \zeta (d-1) R_n / 2}}{a_n^{(\gamma)} (2d(k)) \prod_{i=1}^{k-1} a_n^{(\gamma)} (d_{(i)})^2} \left((c'_{1,n} c'_{2,n})^{1/5} + \rho_n^{-1} (c'_{2,n})^{2/5} \right).
 \end{aligned}$$

From Theorem 4, we have

$$d_K \left(\frac{S_n^{(\gamma)} - \mathbb{E}(S_n^{(\gamma)})}{\sqrt{\text{Var}(S_n^{(\gamma)})}}, N \right) \leq W_1 + W_2 + W_3 + W_4 + W_5 + W_6.$$

Substituting the bounds for $c'_{1,n}$, $c'_{2,n}$, and $c'_{3,n}$ into W_4 , W_5 , and W_6 , as well as $R_n \leq C^* \log n$, one can see that

$$W_4 \leq C^* n^{-1/2+c_4\gamma}; \quad W_5 \leq C^* n^{-1/2+c_5\gamma}; \quad W_6 \leq C^* n^{-1/2+c_6\gamma} \rho_n^{-1/2},$$

for some $c_i > 0$, $i = 4, 5, 6$. Now, for any $a < 1/2$, we can choose γ_0 small enough so that the Kolmogorov bound in (19) holds for all $0 < \gamma < \gamma_0$.

In order to show (20), let us assume $\alpha/\zeta > d_{(k)}$. First, choose $0 < \gamma < 1/2$ such that

$$d_W \left(\frac{S_n^{(\gamma)} - \mathbb{E}(S_n^{(\gamma)})}{\sqrt{\text{Var}(S_n^{(\gamma)})}}, N \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Setting $U_n^{(\gamma)} = S_n - S_n^{(\gamma)}$, we write

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} = \sqrt{\frac{\text{Var}(S_n^{(\gamma)})}{\text{Var}(S_n)}} \times \frac{S_n^{(\gamma)} - \mathbb{E}(S_n^{(\gamma)})}{\sqrt{\text{Var}(S_n^{(\gamma)})}} + \frac{U_n^{(\gamma)} - \mathbb{E}(U_n^{(\gamma)})}{\sqrt{\text{Var}(S_n)}}.$$

From (17), we have that $\text{Var}(S_n^{(\gamma)}) \sim \text{Var}(S_n)$ as $n \rightarrow \infty$. Since the CLT holds for $S_n^{(\gamma)}$, the first term converges in distribution to N as $n \rightarrow \infty$. Now, from (41), we know that $\text{Var}(U_n^{(\gamma)})/\text{Var}(S_n) \rightarrow 0$ as $n \rightarrow \infty$, and hence, by Chebyshev’s inequality, the second term converges to 0 in probability. Thus, applying Slutsky’s theorem, we obtain the CLT for S_n as required. □

Remark 4. If $2\alpha/\zeta > 5d_{(k)}$, one can deduce more precise estimates for W_i , $i = 1, \dots, 6$. In fact, by the choice of the d_i , we have that each $a_n^{(\gamma)}(d'_i)$ converges to a constant, and so we derive that $c'_{i,n} \leq C^*(n/\rho_n)^{5(k-1)\gamma}$ for $i = 1, 2$. Thus we have that

$$\begin{aligned} W_1 &\leq C^* n^{-1/2+2k\gamma} \rho_n^{-2k\gamma}, & W_2 &\leq C^* n^{-1/2+2k\gamma} \rho_n^{-1-2k\gamma}, \\ W_3 &\leq C^* n^{-1/2}, & W_4 &\leq C^* n^{-1/2+4(k-1)\gamma} \rho_n^{-4(k-1)\gamma}, \\ W_5 &\leq C^* n^{-1/2+2(k-1)\gamma} \rho_n^{-2(k-1)\gamma}, & W_6 &\leq C^* n^{-1/2+(2k-1)\gamma} \rho_n^{-1/2-(2k-1)\gamma}. \end{aligned}$$

In particular, the above inequalities imply that $W_i \rightarrow 0$ for $i = 1, \dots, 6$, whenever $\gamma < 1/8(k - 1)$.

Appendix A

In this appendix, we state the main normal approximation result in [32] and explain how Theorem 4 can be deduced from the same. Recall the notation from Section 3.3.1.

Theorem 5. ([32, Theorems 1.1 and 1.2]) *Let F be as in Theorem 4. Then*

$$d_W \left(\frac{F - \mathbb{E}(F)}{\sqrt{\text{Var}(F)}}, N \right) \leq \gamma_1(F) + \gamma_2(F) + \gamma_3(F),$$

$$d_K \left(\frac{F - \mathbb{E}(F)}{\sqrt{\text{Var}(F)}}, N \right) \leq \gamma_1(F) + \gamma_2(F) + \gamma_3(F) + \gamma_4(F) + \gamma_5(F) + \gamma_6(F),$$

where $\gamma_i(F)$, $i = 1, \dots, 6$, are given by

$$\gamma_1(F) := \frac{2}{\text{Var}(F)} \left[\int_{\mathbf{X}^3} \left(\mathbb{E}([D_{x_1}F]^2 [D_{x_2}F]^2) \mathbb{E}([D_{x_1,x_3}^2 F]^2 [D_{x_2,x_3}^2 F]^2) \right)^{1/2} \lambda^3(d(x_1, x_2, x_3)) \right]^{1/2},$$

$$\gamma_2(F) := \frac{1}{\text{Var}(F)} \left[\int_{\mathbf{X}^3} \mathbb{E}([D_{x_1,x_3}^2 F]^2 [D_{x_2,x_3}^2 F]^2) \lambda^3(d(x_1, x_2, x_3)) \right]^{1/2},$$

$$\gamma_3(F) := \frac{1}{\text{Var}(F)^{3/2}} \int_{\mathbf{X}} \mathbb{E}(|D_x F|^3) \lambda(dx),$$

$$\gamma_4(F) := \frac{1}{2\text{Var}(F)^2} \left[\mathbb{E}([F - \mathbb{E}(F)]^4) \right]^{1/4} \int_{\mathbf{X}} \left[\mathbb{E}([D_x F]^4) \right]^{3/4} \lambda(dx),$$

$$\gamma_5(F) := \frac{1}{\text{Var}(F)} \left[\int_{\mathbf{X}} \mathbb{E}([D_x F]^4) \lambda(dx) \right]^{1/2},$$

$$\gamma_6(F) := \frac{1}{\text{Var}(F)} \left[\int_{\mathbf{X}^2} 6 \left(\mathbb{E}([D_{x_1}F]^4) \mathbb{E}([D_{x_1,x_2}^2 F]^4) \right)^{1/2} + 3 \mathbb{E}([D_{x_1,x_2}^2 F]^4) \lambda^2(d(x_1, x_2)) \right]^{1/2}.$$

Proof of Theorem 4. We aim to deduce the bounds for $\gamma_i(F)$ above via the proof strategy for [32, Theorem 6.1]. Using the definitions of c_1, c_2 , as well as Hölder’s inequality, we have that

$$\mathbb{E}([D_x F]^4) \leq c_1^{4/5}; \quad \mathbb{E}(|D_x F|^3) \leq c_1^{3/5}; \quad \mathbb{E}([D_{x_1,x_2}^2 F]^4) \leq c_2^{4/5} \mathbb{P}(D_{x_1,x_2}^2 F \neq 0)^{1/5}. \tag{59}$$

Applying these bounds, together with further applications of Hölder’s inequality, gives that $\gamma_1(F) \leq W_1$ and $\gamma_2(F) \leq W_2$. Note that $\gamma_3(F) = W_3$. This completes the proof for the Wasserstein bound in Theorem 4.

Subsequently, the bounds in (59) also yield that $\gamma_5(F) \leq W_5$ and $\gamma_6(F) \leq W_6$. Using the equation at the end of the proof of [32, Theorem 6.1], we obtain that

$$\gamma_4(F) \leq \frac{c_1^{3/5} \Gamma_F}{\text{Var}(F)^{3/2}} + \frac{c_1^{4/5} \Gamma_F^{5/4} + 2c_1^{4/5} \Gamma_F^{3/2}}{\text{Var}(F)^2},$$

where

$$\Gamma_F := \int_{\mathbf{X}} \mathbb{P}(D_x F \neq 0)^{1/10} \lambda(dx).$$

Using the trivial inequality $\mathbb{P}(D_x F \neq 0) \leq 1$, we get that $\gamma_4(F) \leq W_4$. Now the Kolmogorov bound in Theorem 4 has been obtained. □

Acknowledgements

The authors are very grateful for the detailed and useful comments received from two anonymous referees. These comments helped the authors to introduce a number of improvements to the paper. The work benefited from the visit of both the authors to Technion, Israel, and the authors are thankful to their host Robert Adler for the same. D. Y. is also thankful to the Department of Statistics at Purdue University for hosting him. D. Y. wishes to thank Subhojoy Gupta for some helpful discussions on hyperbolic geometry.

Funding information

T. O.'s research is partially supported by the NSF grant DMS-1811428. D. Y.'s research was supported in part by DST-INSPIRE Faculty fellowship and CPDA from the Indian Statistical Institute.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process for this article.

References

- [1] ABDULLAH, M. A., BODE, M. AND FOUNTOLAKIS, N. (2017). Typical distances in a geometric model for complex networks. *Internet Math.* 1.
- [2] ANDERSON, J. W. (2008). *Hyperbolic Geometry*, 2nd edn. Springer, London.
- [3] BACCELLI, F. AND BLASZCZYSZYN, B. (2009). *Stochastic Geometry and Wireless Networks*, Vol. I, Theory. Now Publishers, Boston.
- [4] BACCELLI, F. AND BLASZCZYSZYN, B. (2010). *Stochastic Geometry and Wireless Networks*, Vol. II, Applications. Now Publishers, Boston.
- [5] BENJAMINI, I. (2013). *Coarse Geometry and Randomness: école d'été de Probabilités de Saint-Flour XLI—2011*. Springer, Cham.
- [6] BENJAMINI, I. AND SCHRAMM, O. (2001). Percolation in the hyperbolic plane. *J. Amer. Math. Soc.* 4, 487–507.
- [7] BLÄSIUS, T., FRIEDRICH, T. AND KROHMER, A. (2018). Cliques in hyperbolic random graphs. *Algorithmica* 80, 2324–2344.
- [8] BOBROWSKI, O. AND MUKHERJEE, S. (2015). The topology of probability distributions on manifolds. *Prob. Theory Relat. Fields* 161, 651–686.
- [9] BOBROWSKI, O. AND OLIVEIRA, G. (2019). Random Čech complexes on Riemannian manifolds. *Random Structures Algorithms* 54, 373–412.
- [10] BODE, M., FOUNTOLAKIS, N. AND MÜLLER, T. (2015). On the largest component of a hyperbolic model of complex networks. *Electron. J. Combinatorics* 22, 3–24.
- [11] BODE, M., FOUNTOLAKIS, N. AND MÜLLER, T. (2016). The probability of connectivity in a hyperbolic model of complex networks. *Random Structures Algorithms* 49, 65–94.
- [12] BROOKS, R. AND MAKOVER, E. (2004). Random construction of Riemann surfaces. *J. Differential Geom.* 68, 121–157.
- [13] CANDELLERO, E. AND FOUNTOLAKIS, N. (2016). Bootstrap percolation and the geometry of complex networks. *Stoch. Process. Appl.* 126, 234–264.
- [14] CANDELLERO, E. AND FOUNTOLAKIS, N. (2016). Clustering and the hyperbolic geometry of complex networks. *Internet Math.* 12, 2–53.
- [15] CANNON, J. W., FLOYD, W. J., KENYON, R. AND PARRY, W. R. (1997). Hyperbolic geometry. In *Flavors of Geometry*, ed. S. LEVY, Cambridge University Press, pp. 59–115.
- [16] CASTELLANOS, J. A. J. AND DARNELL, E. (2007). NonEuclid 2007.04. Available at <http://cs.unm.edu/joel/NonEuclid/NonEuclid.html>.
- [17] CUNNINGHAM, W., ZUEV, K. AND KRIOUKOV, D. (2017). Navigability of random geometric graphs in the universe and other spacetimes. *Sci. Reports* 7, paper no. 8699.
- [18] DE KERGORLAY, H. L., TILLMANN, U. AND VIPOND, O. (2019). *Random Čech complexes on manifolds with boundary*. To appear in *Random Structures Algorithms*.

- [19] FOUNTOLAKIS, N. (2012). *On the evolution of random graphs on spaces of negative curvature*. Preprint. Available at <https://arxiv.org/abs/1205.2923>.
- [20] FOUNTOLAKIS, N. (2015). On a geometrization of the Chung–Lu model for complex networks. *J. Complex Networks* **3**, 361–387.
- [21] FOUNTOLAKIS, N., VAN DER HOORN, P., MÜLLER, T. AND SCHEPERS, M. (2021). Clustering in a hyperbolic model of complex networks. *Electron. J. Probab.* **26**, 132 pp.
- [22] FOUNTOLAKIS, N. AND MÜLLER, T. (2018). Law of large numbers for the largest component in a hyperbolic model of complex networks. *Ann. Appl. Probab.* **28**, 607–650.
- [23] FOUNTOLAKIS, N. AND YUKICH, J. (2020). Limit theory for isolated and extreme points in hyperbolic random geometric graphs. *Electron. J. Probab.* **25**, 51 pp.
- [24] GILBERT, E. N. (1961). Random plane networks. *J. SIAM* **9**, 533–543.
- [25] GUGELMANN, L., PANAGIOTOU, K. AND PETER, U. (2012). *Random hyperbolic graphs: degree sequence and clustering*. In *Automata, Languages, and Programming*, Springer, Berlin, Heidelberg, pp. 573–585.
- [26] HAENGGI, M. (2012). *Stochastic Geometry for Wireless Networks*. Cambridge University Press.
- [27] KIWI, M. AND MITSCHKE, D. (2014). A bound for the diameter of random hyperbolic graphs. In *2015 Proceedings of the Twelfth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, Society for Industrial and Applied Mathematics, Philadelphia, pp. 26–39.
- [28] KIWI, M. AND MITSCHKE, D. (2018). Spectral gap of random hyperbolic graphs and related parameters. *Ann. Appl. Probab.* **28**, 941–989.
- [29] KIWI, M. AND MITSCHKE, D. (2019). On the second largest component of random hyperbolic graphs. *SIAM J. Discrete Math.* **33**, 2200–2217.
- [30] KRIOUKOV, D. *et al.* (2010). Hyperbolic geometry of complex networks. *Phys. Rev. E* **82**, paper no. 036106.
- [31] LALLEY, S. P. (2014). Statistical regularities of self-intersection counts for geodesics on negatively curved surfaces. *Duke Math. J.* **163**, 1191–1261.
- [32] LAST, G., PECCATI, G. AND SCHULTE, M. (2016). Normal approximation on Poisson spaces: Mehler’s formula, second order Poincaré inequalities and stabilization. *Prob. Theory Relat. Fields* **165**, 667–723.
- [33] LAST, G. AND PENROSE, M. (2017). *Lectures on the Poisson Process*. Cambridge University Press.
- [34] LYONS, R. AND PERES, Y. (2016). *Probability on Trees and Networks*. Cambridge University Press.
- [35] MEESTER, R. AND ROY, R. (1996). *Continuum Percolation*. Cambridge University Press.
- [36] MÜLLER, T. AND STAPS, M. (2019). The diameter of KPKVB random graphs. *Adv. Appl. Probab.* **51**, 358–377.
- [37] PECCATI, G. AND REITZNER, M. (eds) (2016). *Stochastic Analysis for Poisson Point Processes: Malliavin Calculus, Wiener–Itô Chaos Expansions and Stochastic Geometry*. Springer, Cham.
- [38] PENROSE, M. (2003). *Random Geometric Graphs*. Oxford University Press.
- [39] PENROSE, M. D. AND YUKICH, J. E. (2013). Limit theory for point processes in manifolds. *Ann. Appl. Probab.* **23**, 2161–2211.
- [40] PETRI, B. AND THAELE, C. (2018). Poisson approximation of the length spectrum of random surfaces. *Indiana Univ. Math. J.* **67**, 1115–1141.
- [41] RATCLIFFE, J. G. (2006). *Foundations of Hyperbolic Manifolds*. Springer, New York.
- [42] YUKICH, J. E. (2006). *Probability Theory of Classical Euclidean Optimization Problems*. Springer, Berlin, Heidelberg.