

# Degree correlations in scale-free null models

Clara Stegehuis

Eindhoven University of Technology

October 16, 2018

## Abstract

We study the average nearest neighbor degree  $a(k)$  of vertices with degree  $k$ . In many real-world networks with power-law degree distribution  $a(k)$  falls off in  $k$ , a property ascribed to the constraint that any two vertices are connected by at most one edge. We show that  $a(k)$  indeed decays in  $k$  in three simple random graph null models with power-law degrees: the erased configuration model, the rank-1 inhomogeneous random graph and the hyperbolic random graph. We find that in the large-network limit for all three null models  $a(k)$  starts to decay beyond  $n^{(\tau-2)/(\tau-1)}$  and then settles on a power law  $a(k) \sim k^{\tau-3}$ , with  $\tau$  the degree exponent.

## 1 Introduction

Complex networks are often studied through mathematical analysis of null models that can match the network degree distribution. For scale-free networks, this degree distribution follows a power law. In many real-world networks, like the Internet, social networks and biological networks, the power-law exponent  $\tau$  is found to be between 2 and 3 [1, 19, 28, 42]. In such scale-free networks, high-degree vertices called hubs are likely present, and give rise to scale-free properties such as ultra-small distances and ultra-fast information spreading. The hubs also crucially influence local properties such as clustering [24, 39] and the occurrence of subgraphs [34]. Clustering can be measured in terms of the probability  $c(k)$  that a degree- $k$  vertex creates triangles. Both empirically [31, 36] and theoretically [17, 39] it was shown that  $c(k)$  falls off with  $k$ , and hence that hubs are less likely to take part in triangles.

Whereas triangles and even larger subgraphs require to study the correlation between at least three vertices, we study in this paper the degree correlation between pairs of two vertices in terms of  $a(k)$ , the average degree of a neighbor of a vertex of degree  $k$ . According to several studies [2, 15], this degree-degree correlation is an essential local network property, because it also falls off with  $k$  and can largely explain the fall-off of  $c(k)$  [9, 15, 38]. We provide support for this statement, by identifying an explicit relation between  $a(k)$  and  $c(k)$  for large  $k$ . But the main goal of this paper is to explain the full spectrum  $k \mapsto a(k)$  for all  $k$ , and to provide theoretical underpinning for the widely observed  $a(k)$  fall-off.

There exist a vast array of papers, empirical, non-rigorous and rigorous, on  $a(k)$  [2, 3, 9, 10, 15, 32, 35, 36, 41, 44]. The function  $k \mapsto a(k)$  describes the correlation between the degrees on the two sides of an edge, and classifies the network into one of the following three categories [33]. When  $a(k)$  increases with  $k$ , the network is said to be *assortative*: vertices with high degrees mostly connect to other vertices with high degrees. When  $a(k)$  decreases in  $k$ , the network is said to be *disassortative*. Then high-degree vertices typically connect to low-degree vertices. When  $a(k)$  is independent of  $k$ , the network is said to be *uncorrelated*. In this case, the degrees on the two different sides of an edge can be viewed as fully independent, a desirable property when studying the mathematical properties of networks. But the fact is that the majority of real-world networks with power-law degrees and unbounded degree fluctuations ( $\tau \in (2, 3)$ ) show a clear decay of

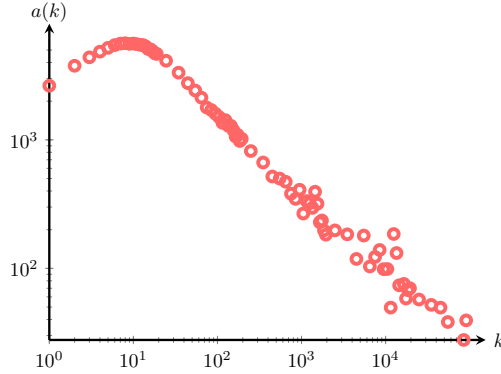


Figure 1:  $a(k)$  for the Youtube friendship network [30]

$a(k)$  as  $k$  grows large [31, 36]. Figure 1 illustrates this for the Youtube friendship network [30]. Hence, scale-free networks are inherently disassortative, and hubs are predominantly connected to small-degree nodes. In complex network theory, such a well established empirical fact then asks for a theoretical explanation. Typically, this explanation comes in the form of a null model that only matches the degree distribution and has the empirical observation as a property, in this case disassortivity, or more specifically, the essential features of the curve  $k \mapsto a(k)$ .

The popular configuration model [11] generates random networks with any prescribed degree distribution, but only results in uncorrelated networks when including self-loops and multi-edges. Hence, the configuration model can never explain the  $a(k)$  fall-off. We therefore resort to different null models that, contrary to the configuration model, generate random networks without self-loops and multi-edges. The resulting *simple* random networks are therefore prone to the structural correlations that come with the presence of hubs. We study  $a(k)$  for three widely used null models: the erased configuration model, the rank-1 inhomogeneous random graph (also called hidden variable model) and the hyperbolic random graph. We show that these models display universal  $a(k)$ -behavior: For  $k$  sufficiently small,  $a(k)$  is independent of  $k$ . Thus, in simple scale-free networks, small-degree vertices have similar neighbors. We then identify the value of  $k$  as of which  $a(k)$  starts decaying. An intuitive explanation for the  $a(k)$  fall-off is that in simple networks, high-degree vertices have so many neighbors that they must reach out to lower-degree vertices, because networks typically only contain a small amount of high-degree vertices. This causes the average degree of a neighbor of a high-degree vertex to be smaller. Thus, single-edge constraints may cause the decay of  $a(k)$ .

## 2 Main results

We first define the average nearest neighbor degree  $a(k, G)$  of a graph  $G$  in more detail. Let  $(D_i)_{i \in [n]}$  be the degree sequence of the graph, where  $[n] = 1, \dots, n$ . Furthermore, let  $N_k$  denote the total number of degree  $k$  vertices in the graph, and  $\mathcal{N}_i$  denote the neighborhood of vertex  $i$ . The average nearest neighbor degree of graph  $G$  is then defined as

$$a(k, G) = \frac{1}{kN_k} \sum_{i: D_i=k} \sum_{j \in \mathcal{N}_i} D_j. \quad (1)$$

Note that it is possible that no vertex of degree  $k$  exists in the graph. We therefore analyze

$$a_\varepsilon(k, G) = \frac{1}{k|M_\varepsilon(k)|} \sum_{i \in M_\varepsilon(k)} \sum_{j \in \mathcal{N}_i} D_j, \quad (2)$$

where  $M_\varepsilon(k) = \{i \in [n] : D_i \in [k(1 - \varepsilon), k(1 + \varepsilon)]\}$ . We will show that in the models we analyze,  $M_\varepsilon(k)$  is non-empty with high probability, so that  $a_\varepsilon(k)$  is well defined. Note that

$a(k, G) = a_0(k, G)$ . We now analyze  $a_\varepsilon(k, G)$ , first for the erased configuration model in Subsection 2.1 and then for the rank-1 inhomogeneous random graph and the hyperbolic random graph in Subsection 2.3.

## 2.1 The erased configuration model

Given a positive integer  $n$  and a degree sequence  $(D_1, D_2, \dots, D_n)$  such that the sum of the degrees is even, the configuration model is a (multi)graph where vertex  $i$  has degree  $D_i$  [11]. We start with  $D_j$  free half-edges adjacent to vertex  $j$ , for  $j = 1, \dots, n$ . The configuration model is then constructed by pairing free half-edges uniformly at random into edges, until no free half-edges remain. Conditionally on obtaining a simple graph, the resulting graph is a uniform graph with the prescribed degrees. This is why the configuration model is often used as a null model for real-world networks with given degrees. When the degree distribution has an infinite second moment however, the probability of obtaining a simple graph tends to zero as  $n$  grows large (see e.g., [22, Chapter 7]). In this setting the configuration model cannot be used as a null model for simple real-world networks anymore. The erased configuration model is the model where all multiple edges are merged and all self-loops are removed [13]. Where the configuration model has hard constraints on the degrees but does not create a simple graph, the erased configuration model generates a simple graph while putting soft constraints on the degrees. In particular, we take the original degree sequence to be an i.i.d. sample from the distribution

$$\mathbb{P}(D = k) = ck^{-\tau}(1 + o(1)), \quad \text{when } k \rightarrow \infty, \quad (3)$$

where  $\tau \in (2, 3)$  so that  $\mathbb{E}[D^2] = \infty$ . We denote  $\mathbb{E}[D] = \mu$ . When this sample constructs a degree sequence such that the sum of the degrees is odd, we add an extra half-edge to the last vertex. This does not affect our computations. We denote the actual degree sequence of the graph after merging the multiple edges and self-loops by  $(D^{(\text{er})})_{i \in [n]}$ , and we call these the erased degrees.

**Stable random variables.** The limit theorem of  $a(k, G_n)$  for the erased configuration model contains stable random variables. A random variable follows a stable distribution if for any positive numbers  $a_1$  and  $a_2$ , there exists a real number  $b_1 = b_1(a_1, a_2)$  and a positive number  $b_2 = b_2(a_1, a_2)$  such that

$$a_1 X_1 + a_2 X_2 \stackrel{d}{=} b_1 + b_2 X, \quad (4)$$

where  $X_1$  and  $X_2$  are independent copies of  $X$ . Stable random variables can be parametrized by four parameters, and are usually denoted by  $\mathcal{S}_\alpha(\sigma, \beta, \mu)$  (see for example [43, Chapter 4]). Throughout this paper, we will only use stable distributions with  $\sigma = 1, \beta = 1, \mu = 0$  and we denote  $\mathcal{S}_\alpha = \mathcal{S}_\alpha(1, 1, 0)$  to ease notation.

We now state the main result for the erased configuration model:

**Theorem 2.1** ( $a(k, G)$  in the erased configuration model). *Let  $(G_n)_{n \geq 1}$  be a sequence of erased configuration models on  $n$  vertices, where the degrees are an i.i.d. sample from (3). Take  $\varepsilon_n$  such that  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$  and  $\lim_{n \rightarrow \infty} n^{-1/(\tau-1)} k \varepsilon_n = \infty$  and let  $\Gamma$  denote the Gamma function.*

(i) For  $k \ll n^{(\tau-2)/(\tau-1)}$ ,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left( \frac{2c\Gamma(\frac{5}{2} - \frac{1}{2}\tau)}{(\tau-1)(3-\tau)} \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \quad (5)$$

where  $\mathcal{S}_{(\tau-1)/2}$  is a stable random variable.

(ii) For  $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$ ,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{3-\tau} k^{\tau-3}} \xrightarrow{\mathbb{P}} -c\mu^{2-\tau} \Gamma(2-\tau). \quad (6)$$

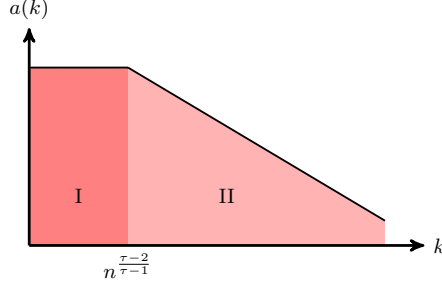


Figure 2: Illustration of the behavior of  $a(k, G_n)$  in the erased configuration model

**Remark 2.1.** The convergence in (5) also holds jointly in  $k$  and  $n$ , so that for  $m \geq 1$  and  $1 \leq k_1 < k_2 < \dots < k_m \ll n^{(\tau-2)/(\tau-1)}$ ,

$$\frac{(a_{\varepsilon_n}(k_1, G_n), \dots, a_{\varepsilon_n}(k_m, G_n))}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left( \frac{2c\Gamma(\frac{5}{2} - \frac{1}{2}\tau)}{(\tau-1)(3-\tau)} \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2} \mathbf{1}, \quad (7)$$

where  $\mathbf{1} \in \mathbb{R}^m$  is a vector with  $m$  entries equal to 1.

Figure 2 illustrates the behavior of  $a(k, G_n)$ . First,  $a(k, G_n)$  stays flat and does not depend on  $k$ . After that,  $a(k, G_n)$  starts decreasing in  $k$ , which shows that the erased configuration model indeed is a disassortative random graph. Theorem 2.1 shows that  $n^{(\tau-2)/(\tau-1)}$  serves as a threshold. Thus, the negative degree-degree correlations due to the single-edge constraint only affect vertices of degrees at least  $n^{(\tau-2)/(\tau-1)}$ . This can be understood as follows. In the erased configuration model the maximum contribution to  $a(k, G)$  (see Propositions 3.1 and 3.2) comes from vertices with degrees proportional to  $n/k$ . The maximal degree in an observation of  $n$  i.i.d. power-law distributed samples is proportional to  $n^{1/(\tau-1)}$  w.h.p. Therefore, if  $k \ll n^{(\tau-2)/(\tau-1)}$ , such vertices with degree proportional to  $n/k$  do not exist w.h.p. This explains the two regimes.

For  $k$  small,  $a(k, G_n)$  converges to a stable random variable, as was also shown in [44] for  $k$  fixed. Thus, for  $k$  small, different instances of the erased configuration model show wild fluctuations. The joint convergence in  $k$  of Remark 2.1 shows that  $a(k, G_n)$  still forms a flat curve in  $k$  for one realization of an erased configuration model when  $k$  is small. In contrast,  $a(k, G_n)$  converges to a constant for large  $k$ -values, so that different realizations of erased configuration models will result in similar  $a(k, G_n)$ -values.

## 2.2 Sketch of the proof

We now give a heuristic proof of Theorem 2.1. Conditionally on the degrees, the probability that vertices with degrees  $D_i$  and  $D_j$  are connected in the erased configuration model can be approximated by [23]

$$1 - e^{-D_i D_j / \mu n}. \quad (8)$$

Let  $v \in M_{\varepsilon_n}(k)$ , and let  $X_{ik}$  denote the indicator that vertex  $i$  is connected to  $v$ . The expected degree of a neighbor of  $v$  can then be approximated by

$$a_{\varepsilon_n}(k, G_n) \approx k^{-1} \sum_{i \in [n]} D_i \mathbb{P}(X_{iv} = 1) \approx k^{-1} \sum_{i \in [n]} D_i (1 - e^{-D_i k / (\mu n)}). \quad (9)$$

The maximum degree in an i.i.d. sample from (3) scales as  $n^{1/(\tau-1)}$  w.h.p.. Thus, as long as  $k \ll n^{(\tau-2)/(\tau-1)}$ , we can Taylor expand the exponential so that

$$a_{\varepsilon_n}(k, G_n) \approx \frac{1}{\mu n} \sum_{i \in [n]} D_i^2. \quad (10)$$

Because  $(D_i)_{i \in [n]}$  are samples from a power-law distribution with infinite second moment, the Stable Law Central Limit Theorem gives Theorem 2.1(i).

When  $k \gg n^{(\tau-2)/(\tau-1)}$ , we approximate the sum in (9) by the integral

$$a_{\varepsilon_n}(k, G_n) \approx cnk^{-1} \int_1^\infty x^{1-\tau} (1 - e^{-xk/(\mu n)}) dx = c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \int_{k/(\mu n)}^\infty y^{1-\tau} (1 - e^{-y}) dy,$$

using the degree distribution (3) and the change of variables  $y = xk/(\mu n)$ . When  $k \ll n$ , we can approximate this by

$$a_{\varepsilon_n}(k, G_n) \approx c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \int_0^\infty y^{1-\tau} (1 - e^{-y}) dy = -c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \Gamma(2-\tau). \quad (11)$$

The proof of Theorem 2.1(ii) then consists of showing that the above approximations are indeed valid. We prove Theorem 2.1 in detail in Sections 3.2 and 3.3.

### 2.3 Two more null models

We now turn to the rank-1 inhomogeneous random graph (or hidden variable model). This model constructs simple graphs with soft constraints on the degree sequence [9, 16]. The graph consists of  $n$  vertices with weights  $(h_i)_{i \in [n]}$ . These weights are an i.i.d. sample from the power-law distribution (3). We denote the average value of the weights by  $\mu$ . Then, every pair of vertices with weights  $(h, h')$  is connected with probability  $p(h, h')$ . In this paper, we take

$$p(h, h') = \min(hh'/(\mu n), 1), \quad (12)$$

which is the Chung-Lu version of the rank-1 inhomogeneous random graph [16]. This connection probability ensures that the degree of a vertex with weight  $h$  will be close to  $h$  [9]. We show the following result:

**Theorem 2.2** ( $a(k, G_n)$  in the rank-1 inhomogeneous random graph). *Let  $(G_n)_{n \geq 1}$  be a sequence of rank-1 inhomogeneous random graphs on  $n$  vertices, where the weights are an i.i.d. sample from (3). Take  $\varepsilon_n$  such that  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$  and  $\lim_{n \rightarrow \infty} n^{-1/(\tau-1)} k \varepsilon_n = \infty$  and let  $\Gamma$  denote the Gamma function.*

(i) For  $1 \ll k \ll n^{(\tau-2)/(\tau-1)}$ ,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left( \frac{2c\Gamma(\frac{5}{2} - \frac{1}{2}\tau)}{(\tau-1)(3-\tau)} \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \quad (13)$$

where  $\mathcal{S}_{(\tau-1)/2}$  is a stable random variable.

(ii) For  $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$ ,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{3-\tau} k^{\tau-3}} \xrightarrow{\mathbb{P}} \frac{c\mu^{2-\tau}}{(3-\tau)(\tau-2)}. \quad (14)$$

Theorem 2.2 is almost identical to Theorem 2.1. The proof of Theorem 2.2 exploits the deep connection between both models, and essentially carries over the results for the erased configuration model to the rank-1 inhomogeneous random graph. The similarity can be understood by noticing that in the erased configuration model the probability that vertices  $i$  and  $j$  with degrees  $D_i$  and  $D_j$  are connected can be approximated by (8) which is close to  $\min(1, \frac{D_i D_j}{\mu n})$ , the connection probability in the rank-1 inhomogeneous random graph. Similar arguments that lead to (9) show that  $a_{\varepsilon_n}(k, G_n)$  can be approximated by

$$a_{\varepsilon_n}(k, G_n) \approx k^{-1} \sum_{i \in [n]} h_i \min(h_i k / \mu n, 1) dx. \quad (15)$$

This sum behaves very similarly to the sum in (9), so that the only difference between Theorem 2.1 and 2.2 is the limiting constants in (6) and (14). The main difference between both models is that in the rank-1 inhomogeneous random graph the presence of all edges is independent as soon as the weights are sampled. This is not true in the erased configuration model, because we know that a vertex with sampled degree  $D_i$  cannot have more than  $D_i$  neighbors, creating dependence between the presence of edges incident to vertex  $i$ . We show that these correlations between the presence of different edges in the erased configuration model are small enough for  $a(k, G_n)$  to behave similarly in the erased configuration model and the rank-1 inhomogeneous random graph.

The third null model we consider is the hyperbolic random graph. This model was introduced in [29] and samples  $n$  vertices on a disk of radius  $R = 2 \log(n/\nu)$ , where the density of the radial coordinate  $r$  at a vertex  $p = (r, \phi)$  is

$$\rho(r) = \alpha \frac{\sinh(\alpha r)}{\cosh(\alpha R) - 1} \quad (16)$$

with  $\alpha = (\tau - 1)/2$ . The angle of  $p$  is sampled uniformly from  $[0, 2\pi]$ . Then, two vertices are connected if their hyperbolic distance is at most  $R$ . The hyperbolic distance of points  $u = (r_u, \phi_u)$  and  $v = (r_v, \phi_v)$  is defined by

$$\cosh(d(u, v)) = \cosh(r_u) \cosh(r_v) - \sinh(r_u) \sinh(r_v) \cos(\theta_{uv}), \quad (17)$$

where  $\theta_{uv}$  denotes the relative angle between  $\phi_u$  and  $\phi_v$ . This creates a simple random graph with power-law degrees with exponent  $\tau$  [29]. The parameter  $\nu$  fixes the average degree of the graph.

The hyperbolic random graph creates simple sparse random graphs with power-law degrees, but in contrast to the erased configuration model and the rank-1 inhomogeneous random graph, can at the same time create many triangles due to its geometric nature [14, 29]. In both the rank-1 inhomogeneous random graph and the erased configuration model, the connection probabilities of different pairs of vertices are (almost) independent. In the hyperbolic random graph, this is not true. When  $u$  is connected to  $v$  and  $u$  is connected to  $w$ , then  $v$  and  $w$  should also be close to one another by the triangle inequality. However, if we define

$$t(u) = e^{(R-r_u)/2} \quad (18)$$

then we show that we can approximate the probability that two randomly chosen vertices  $u$  and  $v$  are connected as

$$\mathbb{P}(X_{uv} = 1 \mid t(u), t(v)) = \min\left(\frac{1}{\pi} \cos^{-1}(1 - 2(\nu t(u)t(v)/n)^2), 1\right), \quad (19)$$

which behaves similarly as the connection probability in the rank-1 inhomogeneous random graph. Furthermore, by [7, Lemma 1.3], the density of  $2 \ln(t(u))$  can be written as

$$f_{2 \ln(t(u))}(x) = \frac{\tau-1}{2} e^{-(\tau-1)x/2} (1 + o(1)), \quad (20)$$

where the  $o(1)$  term is with respect to the network size  $n$ . Therefore,

$$\mathbb{P}(t(u) > x) = \mathbb{P}(2 \ln(t(u)) > 2 \ln(x)) = x^{-\tau+1} (1 + o(1)), \quad (21)$$

so that on a high level the hyperbolic random graph can be interpreted as a rank-1 inhomogeneous random graph with  $(t(u))_{u \in [n]}$  as weights (see [7, Section 1.1.1] for a more elaborate discussion).

The next theorem shows that indeed the behavior of  $a(k, G_n)$  in the hyperbolic random graph is similar as in the rank-1 inhomogeneous random graph:

**Theorem 2.3** ( $a(k, G_n)$  in the hyperbolic random graph). *Let  $(G_n)_{n \geq 1}$  be a sequence of hyperbolic random graphs on  $n$  vertices with power-law degrees with exponent  $\tau$  and parameter  $\nu$ . Take  $\varepsilon_n$  such that  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$  and  $\lim_{n \rightarrow \infty} n^{-1/(\tau-1)} k \varepsilon_n = \infty$  and let  $\Gamma$  denote the Gamma function.*

(i) For  $1 \ll k \ll n^{(\tau-2)/(\tau-1)}$ ,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{2\nu}{\pi} \left( \frac{2}{3-\tau} \Gamma\left(\frac{5}{2} - \frac{1}{2}\tau\right) \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \quad (22)$$

where  $\mathcal{S}_{(\tau-1)/2}$  is a stable random variable.

(ii) For  $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$ ,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{3-\tau} k^{\tau-3}} \xrightarrow{\mathbb{P}} \frac{\nu(\tau-1)^2}{(\tau-2)\pi} \left( \frac{\pi}{2\nu} \right)^{2-\tau} \int_0^\infty x^{1-\tau} \min\left(\frac{1}{\pi} \cos^{-1}(1-2x^2), 1\right) dx. \quad (23)$$

## 2.4 Discussion

**Universality.** The behavior of  $a(k)$  is universal across the three null models we consider. The erased configuration model and the rank-1 inhomogeneous random graph are closely related. They are known to behave similarly for example under critical percolation [4, 5], in terms of distances [40] when  $\tau > 3$ , and in terms of clustering when  $\tau \in (2, 3)$  [39]. The hyperbolic random graph typically shows different behavior, for example in terms of clustering [14, 21], or connectivity [7, 8]. Still, the behavior of  $a(k)$  is similar in the hyperbolic random graph and the other two null models. In all three null models, the main contribution to  $a(k)$  for  $k \gg n^{(\tau-2)/(\tau-1)}$  comes from vertices with degrees proportional to  $n/k$  (see Propositions 3.1 and 3.2). In the hyperbolic random graph, we can relate this maximum contribution to the geometry of the hyperbolic sphere. A vertex  $i$  of degree  $k$  has radius  $r_i \approx R - 2\log(k)$ . Similarly, a vertex  $j$  of degree  $n/(\nu k)$  has radius  $r_j \approx R - 2\log(n/(\nu k)) = 2\log(k)$ . Then,  $r_j \approx R - r_i$ , so that the major contributing vertices have radial coordinate proportional to  $R - r_i$ .

**Expected average nearest neighbor degree.** In Theorems 2.1-2.3 we show that  $a(k, G_n)$  converges in probability to a stable random variable when  $k$  is small. Thus, when we generate many samples of random graphs, we will see that for fixed  $k$ , the distribution of the values of  $a(k, G_n)$  across the different samples will look like a stable random variable. We can also study the expected value of  $a(k, G_n)$  across the different samples. For the erased configuration model for example, using similar techniques as in the proof of Theorem 2.1(ii), we can show that (see Section 3.4)

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[a(k, G_n)]}{(n/k)^{3-\tau}} = -c\mu^{2-\tau} \Gamma(2-\tau). \quad (24)$$

The difference between the scaling of the expected value of  $a(k, G_n)$  and the typical behavior of  $a(k, G_n)$  in Theorem 2.1(i) is caused by high-degree vertices. In typical degree sequences, the maximum degree is proportional to  $n^{1/(\tau-1)}$ . It is unlikely that vertices with higher degrees are present, but if they are, they have a high impact on the average nearest neighbor degree of low degree vertices, causing the difference between the expected average nearest neighbor degree and the typical average nearest neighbor degree. Thus, the expected value of  $a(k, G_n)$  is not very informative when  $k$  is small, since Theorem 2.1 shows that  $a(k, G_n)$  will almost always be smaller than its expected value when  $k$  is small. Also note that including  $\varepsilon_n$  in (24) is not necessary, since the expected value is not affected by the event that no vertex of degree  $k$  is present.

Figure 3 illustrates this difference in terms of the mean and median value of  $a(k, G_n)$  over many realizations of the erased configuration model, the rank-1 inhomogeneous random graph and the hyperbolic random graph. Here indeed we see that the expected average neighbor degree scales as a power of  $k$  over the entire range of  $k$ , where the median shows the straight part of the curve from Theorem 2.1. Thus, it is important to distinguish between mean and median of  $a(k, G_n)$  when simulating random graphs.

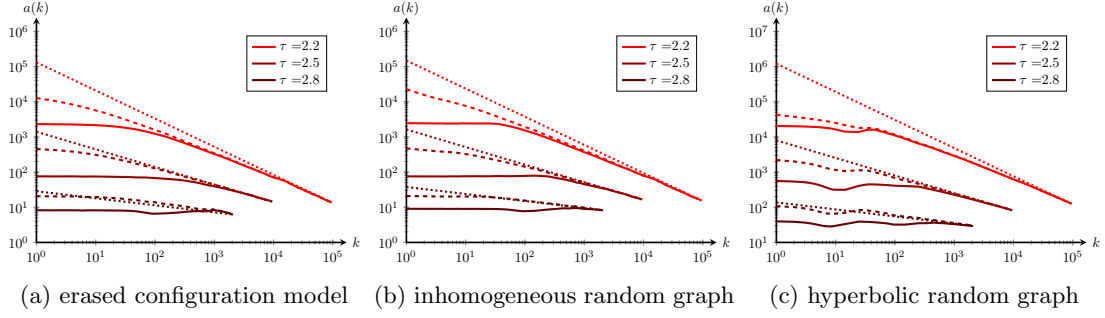


Figure 3:  $a(k, G_n)$  for different random graph models with  $n = 10^6$ . The solid line is the median of  $a(k, G_n)$  over  $10^4$  realizations of the random graph, and the dashed line is the average over these realizations. The dotted line is the asymptotic slope  $k^{\tau-3}$ .

**Vertices of degree  $k$**  The definition (1) assumes that a vertex of degree  $k$  is present. For large values of  $k$ , this is a rare event, by (3). Indeed, vertices of degree at most  $n^{1/\tau}$  are present with high probability in the erased configuration model, whereas the probability that a vertex of degree  $k \gg n^{1/\tau}$  is present tends to zero in the large network limit [44]. We avoid this problem by averaging  $a(k, G_n)$  over a small range of degrees. Another option is to condition on the event that a vertex of degree  $k$  is present. Our proofs for  $k \ll n^{(\tau-2)/(\tau-1)}$  for the erased configuration model can easily be adjusted to condition on this event. For  $k$  larger, we leave the behavior of  $a(k, G_n)$  conditionally on a vertex of degree  $k$  being present open for further research.

**Fixed degrees.** In the proof of Theorem 2.1 we show that the fluctuations that come with the stable laws for small  $k$  are not present when we condition on the degree sequence. Thus, the large fluctuations in  $a(k, G_n)$  for small  $k$  are only caused by fluctuations of the i.i.d. degrees, weights or radii. For a given real-world network, the network degrees are often preserved, and many samples of erased configuration models or inhomogeneous random graphs are created with the observed degree sequence. In this fixed-degree setting, the sample-to-sample fluctuations of  $a(k)$  are relatively small.

**Relation with local clustering.** The local clustering coefficient  $c(k)$  of vertices of degree  $k$  measures the probability that two randomly chosen neighbors of a randomly chosen vertex of degree  $k$  are connected. In many real-world networks as well as simple null models,  $c(k)$  decreases as a function of  $k$  [9, 24, 37, 39, 42]. The relation between the decay rate of  $c(k)$  and the decay rate of  $a(k)$  has been investigated for the rank-1 inhomogeneous random graph, where it was shown that  $c(k) < a(k)/k$  [38]. Using recent results for  $c(k)$  on the erased configuration model and the rank-1 inhomogeneous random graph, we can make the relation between  $c(k)$  and  $a(k)$  more precise. When  $k \gg \sqrt{n}$ ,  $c(k)$  in the erased configuration model satisfies [24]

$$c(k) = c^2 \Gamma(2 - \tau)^2 \mu^{3-2\tau} n^{5-2\tau} k^{2\tau-6} (1 + o_{\mathbb{P}}(1)). \quad (25)$$

Then, by Theorem 2.1, when  $k \gg \sqrt{n}$ ,

$$c(k) = \frac{a^2(k)}{\mu n} (1 + o_{\mathbb{P}}(1)). \quad (26)$$

Intuitively, we can see this relationship in the following way. Pick two neighbors of a vertex with degree  $k$ . By definition, these vertices have degree  $a(k)$  on average. Since  $k \gg \sqrt{n}$ , by Theorem 2.2  $a(k) \ll \sqrt{n}$ . Therefore, the probability of two vertices with weight  $a(k)$  to be connected is approximately  $1 - e^{-a(k)^2/\mu n} \approx a(k)^2/\mu n$ . Since the clustering coefficient can be interpreted as the probability that two randomly chosen neighbors are connected, the clustering coefficient should satisfy  $c(k) \sim a(k)^2/\mu n$  when  $k \gg \sqrt{n}$ . In particular, the decay of the clustering coefficient should



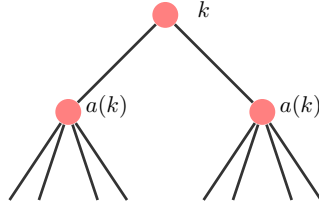


Figure 4: The neighbors of a vertex of degree  $k$  have average degree  $a(k)$

be twice as fast as the decay of the average neighbor degree. Analytical results on  $c(k)$  on the rank-1 inhomogeneous random graph show that (26) is also the correct relation between clustering and degree correlations in the rank-1 inhomogeneous random graph [39]. Future research might explore the relation between  $c(k)$  and  $a(k)$  in other null models, such as the hyperbolic random graph or the preferential attachment model. It would also be interesting to see if the difference between expectation and typical behavior that is present in  $a(k)$  also occurs for the local clustering coefficient  $c(k)$ .

**Correlations in the hyperbolic random graph.** The relation between  $a(k)$  and  $c(k)$  in the rank-1 inhomogeneous random graph and the erased configuration model is based on the fact that in these two models, the connection probabilities of pairs of vertices  $(i, j)$ ,  $(i, k)$  and  $(j, k)$  are (almost) independent. In the hyperbolic random graph, the geometry causes a strong dependence between these connection probabilities. If vertices  $j$  and  $k$  are neighbors of  $i$ , they are likely to be geometrically close to one another due to the triangle inequality. This makes the probability that  $j$  and  $k$  are connected larger than in the rank-1 inhomogeneous random graph or the erased configuration model. These correlations do not play a role when computing  $a(k)$ , since  $a(k)$  only involves the connection probability of two different vertices. When computing statistics of the hyperbolic random graph that include three-point correlations, the equivalence between the hyperbolic random graph and the rank-1 inhomogeneous random graph may fail to hold, as in the example of  $c(k)$ .

Interestingly, the number of cliques was also shown to be similar in the hyperbolic random graph, the rank-1 inhomogeneous random graph and the erased configuration model [20], even though cliques clearly involve three-point correlations. Cliques in the hyperbolic random graph are typically formed between vertices at radius proportional to  $R/2$  [20], so that their degrees are proportional to  $\sqrt{n}$  [7]. These vertices form a dense core, which is very similar to what happens in the erased configuration model and the rank-1 inhomogeneous random graph [27]. In the erased configuration model, many other small subgraphs typically occur between vertices of degrees proportional to  $\sqrt{n}$  [26]. It would be interesting to see if the number of these small subgraphs behaves similarly in the hyperbolic random graph.

### 3 Average nearest neighbor degree in the ECM

In this section, we prove Theorem 2.1. For  $k = o(n^{\frac{\tau-2}{\tau-1}})$ , we couple the degrees of neighbors of a uniformly chosen vertex of degree  $k$  to i.i.d. samples of the size-biased degree distribution in Section 3.2. When  $k \gg n^{\frac{\tau-2}{\tau-1}}$ , this coupling is no longer valid. We then show in Section 3.3 that a specific range of degrees contributes most to  $a_{\varepsilon_n}(k, G_n)$ .

#### 3.1 Preliminaries

We say that  $X_n = O_{\mathbb{P}}(b_n)$  for a sequence of random variables  $(X_n)_{n \geq 1}$  if  $|X_n|/b_n$  is a tight sequence of random variables, and  $X_n = o_{\mathbb{P}}(b_n)$  if  $X_n/b_n \xrightarrow{\mathbb{P}} 0$ . We often want to interchange the sampled

degree of a vertex  $i$ ,  $D_i$  and its erased degree  $D_i^{(\text{er})}$ . By [13, Eq. A(9)]

$$D_i^{(\text{er})} = D_i(1 + o_{\mathbb{P}}(1)), \quad (27)$$

when  $D_i = o(n)$ . Let  $L_n$  denote the total number of half-edges, so that  $L_n = \sum_i D_i$ . We define the events

$$J_n = \{|L_n - \mu n| \leq n^{2/\tau}\} \quad A_n = \{|M_{\varepsilon_n}| \geq 1\}. \quad (28)$$

By [25, Lemma 2.3],  $\mathbb{P}(J_n) \rightarrow 1$  as  $n \rightarrow \infty$ . By [13, Theorem 2.1]

$$|M_{\varepsilon_n}(k)| = cn \int_{k(1-\varepsilon_n)}^{k(1+\varepsilon_n)} x^{-\tau} dx (1 + o_{\mathbb{P}}(1)) = \tilde{C} n^{-1} k^{1-\tau} \varepsilon_n (1 + o_{\mathbb{P}}(1)), \quad (29)$$

for some  $\tilde{C} > 0$ , so that  $\mathbb{P}(A_n) \rightarrow 1$  for  $k \ll n^{1/(\tau-1)}$  by the choice of  $\varepsilon_n$  in Theorem 2.1.

In the rest of this section, we will often condition on the degree sequence. For some event  $\mathcal{E}$ , we use the notation  $\mathbb{P}_n(\mathcal{E}) = \mathbb{P}(\mathcal{E} \mid (D_i)_{i \in [n]})$ , and we define  $\mathbb{E}_n$  and  $\text{Var}_n$  similarly.

### 3.2 Small $k$ : Coupling to i.i.d. random variables

In this section we investigate the behavior of  $a_{\varepsilon_n}(k, G_n)$  when  $k = o(n^{(\tau-2)/(\tau-1)})$ . We first pick a random vertex  $v$  of degree  $k$ . We couple the degrees of the neighbors of  $v$  to i.i.d. copies of the size-biased degree distribution  $D_n^*$ , where

$$\mathbb{P}_n(D_n^* = k) = \frac{k}{L_n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = k\}}. \quad (30)$$

We then use this coupling to compute  $a_{\varepsilon_n}(k, G_n)$ .

*Proof of Theorem 2.1(i).* We first condition on the degree sequence  $(D_i)_{i \in [n]}$ . Let  $v$  be a vertex of degree  $k$ . In the erased configuration model, neighbors of  $v$  are constructed by pairing the half-edges of  $v$  uniformly to other half-edges. The distribution of the degree of a vertex attached to a uniformly chosen half-edge is given by  $D_n^*$ . However, the degrees of the neighbors of  $v$  are not an i.i.d. sample of  $D_n^*$  due to the fact that the half-edges should attach to distinct vertices, because the neighbors of  $v$  should be distinct vertices. We now show that we can still approximate the degrees of the neighbors of  $v$  by an i.i.d. sample of  $D_n^*$  by using a coupling argument. Denote the degrees of the neighbors of  $v$  by  $B_1, \dots, B_k$ , in the order in which we encounter them. Let  $Y_1, \dots, Y_k$  be i.i.d. samples of  $D_n^*$ . These samples can be obtained by sampling uniform half-edges with replacement and setting  $Y_i = D_{v'_i}$ , where  $v'_i$  denotes the vertex incident to the  $i$ th drawn half-edge. We use a similar coupling as in [6, Construction 4.2] to couple the  $B_i$  to  $Y_i$ . Let  $(v'_i)_{i \in [k]}$  denote vertices attached to  $k$  uniformly chosen half-edges (with replacement) and set  $V_0 = v$ . Then for  $i \in [k]$  the coupling is defined in the following way:

- If  $v'_i \notin V_{i-1}$ , then  $B_i = Y_i$  and  $v_i = v'_i$ . Set  $V_i = V_{i-1} \cup v'_i$ . We say that  $B_i$  and  $Y_i$  are successfully coupled.
- If  $v'_i \in V_{i-1}$ , we redraw a uniformly chosen half-edge from the set of half-edges not incident to  $V_{i-1}$ . Let  $v_i$  denote the vertex incident to the chosen half-edge. Set  $B_i = D_{v_i}$  and  $V_i = V_{i-1} \cup v_i$ . We then say that  $B_i$  and  $Y_i$  are miscoupled.

Thus, informally, we select  $k$  uniformly chosen half-edges, and look at the vertices they point to. If these vertices are all distinct, we have successfully coupled the neighbors of  $v$  to an i.i.d. sample of  $D_n^*$ . If not, we need to redraw some of these half-edges to ensure that all neighbors of  $v$  are distinct. We now show that the coupling is successful with high probability. By [6, Lemma 4.3], the probability of a miscoupling at step  $i$  can be bounded as

$$\mathbb{P}_n(B_i \neq Y_i \mid \mathcal{F}_{i-1}) \leq L_n^{-1} \left( k + \sum_{s=1}^{i-1} B_s \right), \quad (31)$$

where  $\mathcal{F}_i = \sigma(B_j, Y_j)_{j \in [i]}$  denotes the sigma-algebra containing all information about the  $Y_j$  and  $B_j$  variables encountered up to step  $i$ . Thus, the expected number of miscouplings up to time  $t$ ,  $N_{\text{mis}}(t)$ , satisfies

$$\mathbb{E}_n [N_{\text{mis}}(t)] \leq \frac{kt}{L_n} + \frac{1}{L_n} \sum_{i=1}^t \sum_{s=1}^{i-1} \mathbb{E}_n [B_s]. \quad (32)$$

When  $B_s$  is successfully coupled,  $\mathbb{E}_n [B_s \mid \text{successfully coupled}] = \mathbb{E}_n [D_n^*] = \sum_i D_i^2 / L_n$ . When  $B_s$  is not successfully coupled, it is drawn in a size-biased manner from the vertices that are not chosen yet. Then

$$\begin{aligned} \mathbb{E}_n [B_s \mid \mathcal{F}_{i-1}, \text{ miscoupled}] &= \frac{\sum_{i \notin V_s} D_i^2}{\sum_{i \notin V_s} D_i} \leq \frac{\sum_{i \in [n]} D_i^2}{\sum_{i \in [n]} D_i - \sum_{i \in V_s} D_i} \\ &= \frac{\sum_{i \in [n]} D_i^2}{\sum_{i \in [n]} D_i} \left( 1 + \frac{\sum_{i \in V_s} D_i}{\sum_{i \in [n]} D_i - \sum_{i \in V_s} D_i} \right). \end{aligned} \quad (33)$$

Since  $D_{\max} = O_{\mathbb{P}}(n^{1/(\tau-1)})$ ,  $\sum_{i \in V_s} D_i = O_{\mathbb{P}}(sn^{1/(\tau-1)})$  for all possible  $\mathcal{F}_{i-1}$ . For  $t$  large, we obtain from (3) that

$$\mathbb{P}(D^2 > t) = \mathbb{P}(D > \sqrt{t}) = \frac{c}{\tau-1} t^{(1-\tau)/2} (1 + o(1)). \quad (34)$$

Using (34) we can use the Stable Law Central Limit Theorem (see for example [43, Theorem 4.5.2]) to conclude that

$$\frac{\sum_{i \in [n]} D_i^2}{n^{2/(\tau-1)} \left( \frac{2c}{(\tau-1)(3-\tau)} \Gamma\left(\frac{5}{2} - \frac{1}{2}\tau\right) \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)}} \xrightarrow{d} \mathcal{S}_{(\tau-1)/2}, \quad (35)$$

where  $\mathcal{S}_{(\tau-1)/2}$  is a stable random variable. Thus, as long as  $s = o(n^{(\tau-2)/(\tau-1)})$ ,

$$\mathbb{E}_n [B_s] = O_{\mathbb{P}}\left(L_n^{-1} \sum_{i \in [n]} D_i^2\right) = O_{\mathbb{P}}\left(n^{(3-\tau)/(\tau-1)}\right). \quad (36)$$

Then, for  $k = o(n^{(\tau-2)/(\tau-1)})$

$$\mathbb{E}_n [N_{\text{mis}}(k)] = \frac{k^2}{L_n} + \frac{1}{L_n} O_{\mathbb{P}}\left(n^{(3-\tau)/(\tau-1)}\right) \sum_{i=1}^k (i-1) = O_{\mathbb{P}}\left(k^2 n^{\frac{2-\tau}{\tau-1}}\right). \quad (37)$$

Thus, as long as  $k = o(n^{\frac{\tau-2}{\tau-1}})$ ,

$$\mathbb{E}_n [N_{\text{mis}}(k)] = o_{\mathbb{P}}(1). \quad (38)$$

Then, by the Markov inequality

$$\mathbb{P}_n (N_{\text{mis}}(k) = 0) = 1 - \mathbb{P}_n (N_{\text{mis}}(k) \geq 1) \geq 1 - \mathbb{E}_n [N_{\text{mis}}(k)] = 1 - o_{\mathbb{P}}(1). \quad (39)$$

Thus, when  $k = o(n^{(\tau-2)/(\tau-1)})$ , we can approximate the sum of the degrees of the neighbors of a vertex with degree  $k$  by i.i.d. samples of the size-biased degree distribution. Because  $D_i^{(\text{er})} = D_i(1 + o_{\mathbb{P}}(1))$ , conditionally on the degree sequence

$$a_{\varepsilon_n}(k, G_n) = \frac{1}{k |M_{\varepsilon_n}(k)|} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{j \in \mathcal{N}_i} D_j^{(\text{er})} = \frac{1}{k} \mathbb{E}_n \left[ \sum_{j \in \mathcal{N}_{V_k}} D_j^{(\text{er})} \right] = (1 + o_{\mathbb{P}}(1)) \mathbb{E}_n [D_{\mathcal{N}_{V_k}(U)}],$$

where  $V_k$  denotes a uniformly chosen vertex in  $M_{\varepsilon_n}(k)$ , and  $\mathcal{N}_{V_k}(U)$  is a uniformly chosen neighbor of vertex  $V_k$ . Here the second equality holds because the average nearest neighbor degree averages over all neighbors of vertex  $j$ , and the third equality holds because it also averages over all vertices

in  $M_{\varepsilon_n}(k)$ , together with the fact that  $D_{V_k} = k(1 + o(1))$  and  $D_i^{(\text{er})} = D_i(1 + o_{\mathbb{P}}(1))$  uniformly in  $i$ . With high probability, we can couple the degrees of neighbors of a uniformly chosen vertex of degree in  $[k(1 - \varepsilon_n), k(1 + \varepsilon_n)]$  to i.i.d copies of  $D_n^*$ . Then, conditionally on the degree sequence,

$$a_{\varepsilon_n}(k, G_n) = (1 + o_{\mathbb{P}}(1)) \mathbb{E}_n[D_n^*] = (1 + o_{\mathbb{P}}(1)) L_n^{-1} \sum_{i \in [n]} D_i^2. \quad (40)$$

Note that this expression is independent of  $k$ . Combining this with (35) results in

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left( \frac{2c}{(\tau-1)(3-\tau)} \Gamma\left(\frac{5}{2} - \frac{1}{2}\tau\right) \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}. \quad (41)$$

The fact that (40) is independent of  $k$  proves the joint convergence of Remark 2.1.  $\square$

### 3.3 Large $k$

Now we study the value of  $a_{\varepsilon_n}(k, G_n)$  when  $k \gg n^{(\tau-2)/(\tau-1)}$ . We show that there exists a range of degrees  $W_n^k(\delta)$  which gives the largest contribution to  $a_{\varepsilon_n}(k, G_n)$ . For ease of notation, we write  $a_{\varepsilon_n}(k)$  for  $a_{\varepsilon_n}(k, G_n)$  in this section. We define

$$W_n^k(\delta) = \{u : D_u \in [\delta n/k, n/(\delta k)]\}, \quad (42)$$

and we write

$$\begin{aligned} a_{\varepsilon_n}(k) &= \frac{1}{k |M_{\varepsilon_n}(k)|} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{j \in W_n^k(\delta)} D_j^{(\text{er})} + \frac{1}{k |M_{\varepsilon_n}(k)|} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{j \notin W_n^k(\delta)} D_j^{(\text{er})} \\ &=: a_{\varepsilon_n}(k, W_n^k(\delta)) + a_{\varepsilon_n}(k, \bar{W}_n^k(\delta)), \end{aligned} \quad (43)$$

where  $a_{\varepsilon_n}(k, W_n^k(\delta))$  denotes the contribution to  $a_{\varepsilon_n}(k)$  from vertices in  $W_n^k(\delta)$ , and  $a_{\varepsilon_n}(k, \bar{W}_n^k(\delta))$  the contribution from vertices not in  $W_n^k(\delta)$ . In the rest of this section, we prove the following two propositions, which together show that the largest contribution to  $a_{\varepsilon_n}(k)$  indeed comes from vertices in  $W_n^k(\delta)$ .

**Proposition 3.1** (Minor contributions). *There exists  $\kappa > 0$  such that for  $k \gg n^{(\tau-2)/(\tau-1)}$ ,*

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}[a_{\varepsilon_n}(k, \bar{W}_n^k(\delta))]}{(n/k)^{3-\tau}} = O(\delta^\kappa). \quad (44)$$

**Proposition 3.2** (Major contributions).

$$\frac{a_{\varepsilon_n}(k, W_n^k(\delta))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx \quad (45)$$

We now show how these propositions prove part (ii) of Theorem 2.1.

*Proof of Theorem 2.1 (ii).* By the Markov inequality and Proposition 3.1,

$$\frac{a_{\varepsilon_n}(k, \bar{W}_n^k(\delta))}{(n/k)^{3-\tau}} = O_{\mathbb{P}}(\delta^\kappa). \quad (46)$$

Combining this with Proposition 3.2 results in

$$\frac{a_{\varepsilon_n}(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx + O_{\mathbb{P}}(\delta^\kappa). \quad (47)$$

Taking the limit of  $\delta \rightarrow 0$  then proves the theorem.  $\square$

The rest of this section is devoted to proving Propositions 3.1 and 3.2.

### 3.3.1 Conditional expectation

We first compute the expectation of  $a_{\varepsilon_n}(k, W_n^k(\delta))$  when we condition on the degree sequence.

**Lemma 3.3.** *When  $k \gg n^{(\tau-2)/(\tau-1)}$ ,*

$$\mathbb{E}_n [a_{\varepsilon_n}(k, W_n^k(\delta))] = \frac{1}{k} \sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k/L_n}) (1 + o_{\mathbb{P}}(1)). \quad (48)$$

*Proof.* It suffices to prove the lemma under the event  $J_n$  from (28), since  $\mathbb{P}(J_n) \rightarrow 1$ . Thus we may assume that  $L_n = \mu n(1 + o(1))$ . Let  $X_{ij}$  denote the indicator  $i$  and  $j$  are connected. By (1)

$$\mathbb{E}_n [a_{\varepsilon_n}(k, W_n^k(\delta))] = \frac{1}{k |M_{\varepsilon_n}(k)|} \sum_{v \in M_{\varepsilon_n}(k)} \sum_{u \in W_n^k(\delta)} D_u^{(\text{er})} \mathbb{P}_n(X_{uv} = 1) \quad (49)$$

By [23, Eq. (4.9)]

$$\mathbb{P}_n(X_{uv} = 1) = 1 - e^{D_u D_v/L_n} + O\left(\frac{D_u^2 D_v + D_u^2 D_v}{L_n^2}\right) = (1 - e^{D_u D_v/L_n})(1 + o_{\mathbb{P}}(1)), \quad (50)$$

where the last step follows because  $D_u \in n/k[\delta, 1/\delta]$  and by (27)  $D_v = k(1 + o_{\mathbb{P}}(1))$  when  $v \in M_{\varepsilon_n}(k)$ . Further using that  $D_u^{(\text{er})} = D_u(1 + o_{\mathbb{P}}(1))$ , we can write (49) as

$$\begin{aligned} \mathbb{E}_n [a_{\varepsilon_n}(k, W_n^k(\delta))] &= \frac{(1 + o_{\mathbb{P}}(1))}{k |M_{\varepsilon_n}(k)|} \sum_{v \in M_{\varepsilon_n}(k)} \sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k/L_n} e^{o_{\mathbb{P}}(D_u/L_n)}) \\ &= \frac{1}{k} \sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k/L_n} e^{o_{\mathbb{P}}(D_u/L_n)}) (1 + o_{\mathbb{P}}(1)) \\ &= \frac{1}{k} \sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k/L_n}) (1 + o_{\mathbb{P}}(1)) \end{aligned} \quad (51)$$

for  $k \ll n$ , which proves the lemma.  $\square$

### 3.3.2 Convergence of conditional expectation

We now show that  $\mathbb{E}_n [a_{\varepsilon_n}(k, W_n^k(\delta))]$  as computed in Lemma 3.3 converges to a constant when we take the i.i.d. degrees into account.

**Lemma 3.4.** *When  $k \gg n^{(\tau-2)/(\tau-1)}$ ,*

$$\frac{\mathbb{E}_n [a_{\varepsilon_n}(k, W_n^k(\delta))]}{n^{3-\tau} k^{\tau-3}} \xrightarrow{\mathbb{P}} c \mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx. \quad (52)$$

*Proof.* Define the random measure

$$M^{(n)}[a, b] = \frac{1}{\mu^{1-\tau} n^{2-\tau} k^{\tau-1}} \sum_{u \in [n]} \mathbb{1}_{\{D_u \in [a, b] \mu n/k\}}. \quad (53)$$

Since the degrees are i.i.d. samples from a power-law distribution, the number of vertices with degrees in interval  $[a, b]$  is binomially distributed. Then,

$$\begin{aligned} M^{(n)}[a, b] &= \frac{1}{\mu^{1-\tau} n^{2-\tau} k^{\tau-1}} |\{u : D_u \in [a, b] \mu n/k\}| \xrightarrow{\mathbb{P}} \frac{1}{(\mu n)^{1-\tau} k^{\tau-1}} \mathbb{P}(D \in [a, b] \mu n/k) \\ &= \frac{1}{(\mu n)^{1-\tau} k^{\tau-1}} \int_{a \mu n/k}^{b \mu n/k} c x^{-\tau} dx = \int_a^b c y^{-\tau} dy =: \lambda[a, b], \end{aligned} \quad (54)$$

where we used the change of variables  $y = xk/(\mu n)$ . By Lemma 3.3,

$$\begin{aligned}\mathbb{E}_n [a_{\varepsilon_n}(k, W_n^k(\delta))] &= \frac{\sum_{u \in W_n^k(\delta)} D_u(1 - e^{-D_u k/L_n})}{k} (1 + o_{\mathbb{P}}(1)) \\ &= \frac{\mu n}{k} \frac{\sum_{u \in W_n^k(\delta)} \frac{D_u k}{\mu n} (1 - e^{-D_u k/L_n})}{k} (1 + o_{\mathbb{P}}(1)) \\ &= \frac{\mu^{2-\tau} n^{3-\tau}}{k^{3-\tau}} \int_{\delta}^{1/\delta} t(1 - e^{-t}) dM^{(n)}(t) (1 + o_{\mathbb{P}}(1)).\end{aligned}\quad (55)$$

Fix  $\eta > 0$ . Since  $t(1 - e^{-t})$  is bounded and continuous on  $[\delta, 1/\delta]$ , we can find  $m < \infty$ , disjoint intervals  $(B_i)_{i \in [m]}$  and constants  $(b_i)_{i \in [m]}$  such that  $\cup B_i = [\delta, 1/\delta]$  and

$$\left| t(1 - e^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} \right| < \eta / \lambda([\delta, 1/\delta]), \quad (56)$$

for all  $t \in [\delta, 1/\delta]$ . Because  $M^{(n)}(B_i) \xrightarrow{\mathbb{P}} \lambda(B_i)$  for all  $i$ ,  $M^{(n)}(B_i) = O_{\mathbb{P}}(\lambda(B_i))$ . Then,

$$\begin{aligned}& \left| \int_{\delta}^{1/\delta} t(1 - e^{-t}) dM^{(n)}(t) - \int_{\delta}^{1/\delta} t(1 - e^{-t}) d\lambda(t) \right| \\ & \leq \left| \int_{\delta}^{1/\delta} t(1 - e^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} dM^{(n)}(t) \right| \\ & \quad + \left| \int_{\delta}^{1/\delta} t(1 - e^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} d\lambda(t) \right| \\ & \quad + \left| \sum_{i=1}^m b_i (M^{(n)}(B_i) - \lambda(B_i)) \right| \\ & \leq \eta M^{(n)}([\delta, 1/\delta]) / \lambda([\delta, 1/\delta]) + \eta + o_{\mathbb{P}}(\eta).\end{aligned}\quad (57)$$

Using that  $M^{(n)}([\delta, 1/\delta]) = O_{\mathbb{P}}(\lambda([\delta, 1/\delta]))$  proves that

$$\int_{\delta}^{1/\delta} t(1 - e^{-t}) dM^{(n)}(t) \xrightarrow{\mathbb{P}} \int_{\delta}^{1/\delta} t(1 - e^{-t}) d\lambda(t) = c \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx, \quad (58)$$

which proves the lemma.  $\square$

### 3.3.3 Conditional variance of $a(k)$

We now show that the variance of  $a_{\varepsilon_n}(k, W_n^k(\delta))$  is small when conditioning on the degree sequence, so that  $a_{\varepsilon_n}(k, W_n^k(\delta))$  concentrates around its expected value computed in Lemma 3.3.

**Lemma 3.5.** *When  $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$ ,*

$$\frac{\text{Var}_n(a_{\varepsilon_n}(k, W_n^k(\delta)))}{\mathbb{E}_n[a_{\varepsilon_n}(k, W_n^k(\delta))]^2} \xrightarrow{\mathbb{P}} 0. \quad (59)$$

*Proof.* Again, it suffices to prove the lemma under the event  $J_n$  and  $A_n$  from (28). We write the variance of  $a_{\varepsilon_n}(k, W_n^k(\delta))$  as

$$\begin{aligned}\text{Var}_n(a_{\varepsilon_n}(k, W_n^k(\delta))) &= \frac{1}{k^2 |M_{\varepsilon_n}(k)|^2} \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{u,v \in W_n^k(\delta)} D_u^{(\text{er})} D_v^{(\text{er})} \\ & \quad \times (\mathbb{P}_n(X_{iu} = X_{jv} = 1) - \mathbb{P}_n(X_{iu} = 1) \mathbb{P}_n(X_{jv} = 1)) \\ &= \frac{(1 + o_{\mathbb{P}}(1))}{k^2 |M_{\varepsilon_n}(k)|^2} \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{u,v \in W_n^k(\delta)} D_u D_v \\ & \quad \times (\mathbb{P}_n(X_{iu} = X_{jv} = 1) - \mathbb{P}_n(X_{iu} = 1) \mathbb{P}_n(X_{jv} = 1)).\end{aligned}\quad (60)$$

Equation (60) splits into various cases, depending on the size of  $\{i, j, u, v\}$ . We denote the contribution of  $|\{i, j, u, v\}| = r$  to the variance by  $V^{(r)}(k)$ . We first consider  $V^{(4)}(k)$ . We can write

$$\mathbb{P}_n(X_{iu} = X_{jv} = 0) = \mathbb{P}_n(X_{iu} = 0) \mathbb{P}_n(X_{jv} = 0 \mid X_{iu} = 0). \quad (61)$$

For the second term, we first pair all half-edges adjacent to vertex  $i$ , conditionally on not pairing to vertex  $u$ . Then the second term can be interpreted as the probability that vertex  $j$  does not pair to vertex  $v$  in a configuration model with  $L_n - D_i = L_n(1 + o(1))$  vertices, where the degree of vertex  $j$  is reduced by the amount of half-edges from vertex  $i$  that paired to  $j$ . Similarly, the new degree of vertex  $v$  is reduced by the amount of half-edges from vertex  $i$  that paired to  $v$ . Since the expected number of half-edges from  $i$  that pair to vertex  $j$  is  $O(D_i D_j / L_n) = D_j o(n^{-(\tau-1)/(\tau-1)})$  [18], the new degree of vertex  $j$  is  $D_j(1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)}))$ , and a similar statement holds for vertex  $v$ . Thus, by (50)

$$\mathbb{P}_n(X_{iu} = X_{jv} = 0) = e^{-D_i D_u / L_n} e^{-D_j D_v / L_n} (1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)})). \quad (62)$$

This results in

$$\begin{aligned} \mathbb{P}_n(X_{iu} = X_{jv} = 1) &= 1 - \mathbb{P}_n(X_{iu} = 0) - \mathbb{P}_n(X_{jv} = 0) + \mathbb{P}_n(X_{iu} = X_{jv} = 0) \\ &= 1 + (-e^{-\frac{D_u k}{L_n}} - e^{-\frac{D_v k}{L_n}} + e^{-\frac{D_u k}{L_n} - \frac{D_v k}{L_n}})(1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)})) \\ &= (1 - e^{-D_u k / L_n})(1 - e^{-D_v k / L_n})(1 + o_{\mathbb{P}}(1)), \end{aligned} \quad (63)$$

where the last equality holds because  $D_u k = \Theta(n)$  and  $D_v k = \Theta(n)$  for  $u, v \in W_n^k(\delta)$ . Therefore

$$\begin{aligned} V^{(4)}(k) &= \frac{1}{|M_{\varepsilon_n}(k)|^2 k^2} \sum_{i, j \in M_{\varepsilon_n}(k)} \sum_{u, v \in W_n^k(\delta)} D_u D_v (1 - e^{-D_u k / L_n})(1 - e^{-D_v k / L_n})(1 + o_{\mathbb{P}}(1)) \\ &\quad - D_u D_v (1 - e^{-D_u k / L_n})(1 - e^{-D_u k / L_n})(1 + o_{\mathbb{P}}(1)) \\ &= \sum_{u, v \in W_n^k(\delta)} o_{\mathbb{P}}\left(k^{-2} D_u D_v (1 - e^{-D_u k / L_n})(1 - e^{-D_v k / L_n})\right) \\ &= o_{\mathbb{P}}\left(\mathbb{E}_n[a_{\varepsilon_n}(k, W_n^k(\delta))]^2\right), \end{aligned} \quad (64)$$

where the last equality follows from Lemma 3.3. Since there are no overlapping edges when  $\{i, j, u, v\} = 3$ ,  $V^{(3)}(k)$  can be bounded similarly.

We then consider the contribution from  $V^{(2)}$ , which is the contribution where the two edges are the same. By Lemma 3.4, we have to show that this contribution is small compared to  $n^{6-2\tau} k^{2\tau-6}$ . We bound the summand in (60) as

$$D_u^2 \left( \mathbb{P}_n(X_{iu} = 1) - \mathbb{P}_n(X_{iu} = 1)^2 \right) \leq D_u^2. \quad (65)$$

Thus, using that on  $A_n$ ,  $|M_{\varepsilon_n}(k)| \geq 1$ ,  $V^{(2)}$ , can be bounded as

$$V^{(2)} \leq \frac{1}{k^2 |M_{\varepsilon_n}(k)|^2} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{u \in W_n^k(\delta)} D_u^2 = \frac{1}{k^2 |M_{\varepsilon_n}(k)|} \sum_{u \in W_n^k(\delta)} D_u^2 = O\left(\frac{n^2}{k^4}\right) |W_n^k(\delta)|. \quad (66)$$

Since the degrees are i.i.d. samples from (3),  $|W_n^k(\delta)|$  is distributed as a Binomial( $n, C(n/k)^{1-\tau}$ ) for some constant  $C$ . Therefore,  $|W_n^k(\delta)| = O_{\mathbb{P}}\left(n(n/k)^{1-\tau}\right)$ . This results in

$$V^{(2)} = O_{\mathbb{P}}\left(n^{4-\tau} k^{\tau-5}\right), \quad (67)$$

which is smaller than  $n^{6-2\tau} k^{2\tau-6}$  when  $k \gg n^{\frac{\tau-2}{\tau-1}}$ , as required.  $\square$

*Proof of Proposition 3.2.* Lemma 3.5 together with the Chebyshev inequality show that

$$\frac{a_{\varepsilon_n}(k, W_n^k(\delta))}{\mathbb{E}_n[a_{\varepsilon_n}(k, W_n^k(\delta))]} \xrightarrow{\mathbb{P}} 1. \quad (68)$$

Combining this with Lemmas 3.3 and 3.4 yields

$$\frac{a_{\varepsilon_n}(k, W_n^k(\delta))}{n^{3-\tau}k^{\tau-3}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau}(1 - e^{-x})dx. \quad (69)$$

□

### 3.3.4 Contributions outside $W_n^k(\delta)$

In this section, we prove Proposition 3.1 and show that the contribution to  $a_{\varepsilon_n}(k)$  outside of the major contributing regimes as described in (42) is negligible.

*Proof of Proposition 3.1.* We use that  $\mathbb{P}_n(X_{ij} = 1) \leq \min(1, \frac{D_i D_j}{L_n})$ . This yields

$$\begin{aligned} \mathbb{E}[a_{\varepsilon_n}(k, \bar{W}_n^k(\delta))] &= \mathbb{E}[\mathbb{E}_n[a_{\varepsilon_n}(k, \bar{W}_n^k(\delta))]] \leq \frac{n}{k} \mathbb{E}\left[D \min\left(1, \frac{kD}{L_n}\right) \mathbb{1}_{\{D \in \bar{W}_n^k(\delta)\}}\right] \\ &= \frac{n}{k} \int_0^{\delta\mu n/k} x^{1-\tau} \min\left(1, \frac{kx}{\mu n}\right) dx + \frac{n}{k} \int_{\mu n/(\delta k)}^{\infty} x^{1-\tau} \min\left(1, \frac{kx}{\mu n}\right) dx. \end{aligned} \quad (70)$$

For ease of notation, we assume that  $\mu = 1$  in the rest of this section. We have to show that the contribution to (70) from vertices  $u$  such that  $D_u < \delta n/k$  or  $D_u > n/(\delta k)$  is small. First, we study the contribution to (70) for  $D_u < \delta n/k$ . We can bound this contribution by taking the second term of the minimum, which bounds the contribution as

$$\int_0^{\delta n/k} x^{2-\tau} dx = \frac{\delta^{3-\tau}}{\tau-3} (k/n)^{\tau-3}. \quad (71)$$

Then, we study the contribution for  $D_u > n/(\delta k)$ . This contribution can be bounded very similarly by taking 1 for the minimum in (70)

$$\frac{n}{k} \int_{n/(\delta k)}^{\infty} x^{1-\tau} dx = \frac{\delta^{\tau-2}}{\tau-2} (k/n)^{\tau-3}. \quad (72)$$

Taking  $\kappa = \min(\tau-2, 3-\tau) > 0$  then proves the proposition. □

## 3.4 Expected average nearest neighbor degree

For  $k \gg n^{(\tau-2)/(\tau-1)}$ , it is easy to see that  $\mathbb{E}[a(k, G_n)]$  satisfies (24). Indeed, Proposition 3.1 together with Lemma 3.4 and taking the limit of  $\delta \rightarrow 0$  establish (24).

For  $k$  small, note that the limit in Theorem 2.1(i) has infinite mean, so that a similar scaling as in Theorem 2.1(i) cannot be expected to hold for  $\mathbb{E}[a(k, G_n)]$ . To prove Theorem 2.1(i), we used that the maximal degree scales as  $n^{1/(\tau-1)}$  with high probability. When computing the expected average nearest neighbor degree however, the rare event of the maximal degree being larger than  $n^{1/(\tau-1)}$  forms a major contribution to  $\mathbb{E}[a(k, G_n)]$ . In fact, we can follow the exact same lines as the proofs of Proposition 3.1 and Lemma 3.4, so that (24) also holds for  $k$  small.

## 4 Proofs of Theorem 2.2 and 2.3

We now briefly show how the proof of Theorem 2.1 can be adapted for the rank-1 inhomogeneous random graph and the hyperbolic random graph to prove Theorems 2.2 and 2.3. We denote by  $\mathbb{P}_n$  the probability conditioned on the weights in the rank-1 inhomogeneous random graph or conditioned on the radial coordinates in the hyperbolic model.



## 4.1 Inhomogeneous random graph

First, we show how to prove Theorem 2.2(i). Similar to (27), in the rank-1 inhomogeneous random graph the degree of a vertex with weight  $h$ ,  $D_h$ , satisfies  $D_h = h(1 + o_{\mathbb{P}}(1))$  when  $h \gg 1$  [39]. Furthermore, the largest weight is of order  $n^{1/(\tau-1)}$  with high probability. Thus, when  $h \ll n^{(\tau-2)/(\tau-1)}$ , w.h.p.  $p(h, h') = hh'/(\mu n)$  for all vertices. When  $u \in M_{\varepsilon_n}(k)$ ,  $h_u = k(1 + o_{\mathbb{P}}(1))$ , so that conditionally on the weight sequence

$$\begin{aligned} a_{\varepsilon_n}(k) &= \frac{1}{k|M_{\varepsilon_n}(k)|} \sum_{u \in M_{\varepsilon_n}(k)} \sum_{i \in [n]} D_i \mathbb{P}_n(X_{iu} = 1) \\ &= (1 + o_{\mathbb{P}}(1)) \frac{1}{k} \sum_{i \in [n]} h_i \frac{h_i k}{\mu n} = (1 + o_{\mathbb{P}}(1)) \sum_{i \in [n]} \frac{h_i^2}{\mu n}, \end{aligned} \quad (73)$$

which is equivalent to (40) because the weights are also sampled from (3). This proves Theorem 2.2(i).

Similarly to (42), we define for the rank-1 inhomogeneous random graph

$$W_n^{k, \text{HVM}}(\delta) = \{u : h_u \in [\delta \mu n/k, \mu n/(\delta k)]\}. \quad (74)$$

Then it is easy to show that Proposition 3.1 also holds for the rank-1 inhomogeneous random graph with (74) instead of  $W_n^k(\delta)$ . We use that  $\mathbb{P}_n(X_{ij} = 1) = \min(h_i h_j / (\mu n), 1)$ . Because the weights are sampled from (3), (70) also holds for the rank-1 inhomogeneous random graph, so that Proposition 3.1 indeed holds for the rank-1 inhomogeneous random graph.

We now sketch how to adjust the proof of Proposition 3.2 to prove an analogous version for the rank-1 inhomogeneous random graph, which states that

$$\frac{a_{\varepsilon_n}(k, W_n^{k, \text{HVM}}(\delta))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} \min(x, 1) dx. \quad (75)$$

Following the proofs of Lemmas 3.3-3.5, we see that these lemmas also hold for the rank-1 inhomogeneous random graph if we replace the connection probability of the erased configuration model of  $1 - e^{-D_i D_j / L_n}$  by  $\min(h_i h_j / (\mu n), 1)$ . Note that for the rank-1 inhomogeneous random graph the contribution to (60) from 3 or 4 different vertices is 0, because the edge probabilities in the rank-1 inhomogeneous random graph conditioned on the weights are independent. From these lemmas, (75) follows. This then shows similarly to (47) that

$$\frac{a_{\varepsilon_n}(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_0^{\infty} x^{1-\tau} \min(x, 1) dx = \frac{c\mu^{2-\tau}}{(3-\tau)(\tau-2)}. \quad (76)$$

which proves Theorem 2.2(ii).

## 4.2 Hyperbolic random graph

We first provide a lemma that gives the connection probabilities conditioned on the radial coordinates in the hyperbolic random graph.

**Lemma 4.1.** *For a hyperbolic random graph, the probability that  $u$  and  $v$  are connected conditionally on the radial coordinates can be written as*

$$\mathbb{P}_n(X_{uv} = 1) = \min\left(\frac{1}{\pi} \cos^{-1}(1 - 2(\nu t(u)t(v)/n)^2), 1\right) (1 + o_{\mathbb{P}}(1)). \quad (77)$$

*Proof.* Suppose  $\nu t(u)t(v)/n \geq 1$ . Then,

$$\frac{\nu t(u)t(v)}{n} = \frac{\nu e^R e^{-(r_u + r_v)/2}}{n} = \frac{n}{\nu} e^{-(r_u + r_v)/2}, \quad (78)$$

so that  $r_u + r_v \leq 2 \log(N/\nu) = R$ . Thus, by (17)

$$\begin{aligned} \cosh(d(u, v)) &= \cosh(r_u) \cosh(r_v) - \sinh(r_u) \sinh(r_v) \cos(\theta_{uv}) \\ &\leq \cosh(r_u + r_v) \leq \cosh(R), \end{aligned} \quad (79)$$

so that the distance between  $u$  and  $v$  is less than  $R$  and  $u$  and  $v$  are connected.

Now suppose that  $\nu t(u)t(v)/n < 1$ , so that  $r_u + r_v > R$ . We calculate the maximal value of  $\theta_{uv}$  such that  $u$  and  $v$  are connected, which we denote by  $\theta_{uv}^*$ . When the angle between  $u$  and  $v$  equals  $\theta_{uv}^*$ , the hyperbolic distance between  $u$  and  $v$  is precisely  $R$ . Thus, we obtain, using the definition of the hyperbolic sine and cosine

$$\frac{e^R - e^{-R}}{2} = \frac{e^{r_u} - e^{-r_u}}{2} \frac{e^{r_v} - e^{-r_v}}{2} - \frac{e^{r_u} + e^{-r_u}}{2} \frac{e^{r_v} + e^{-r_v}}{2} \cos(\theta_{uv}^*). \quad (80)$$

Because  $t(u)$  is distributed as (21), the maximal type scales is  $O_{\mathbb{P}}(n^{1/(\tau-1)})$ . Therefore,  $e^{r_u - r_v} = (t(v)/t(u))^2 = O_{\mathbb{P}}(n^{2/(\tau-1)})$ . Also,  $e^{-R} = O(n^{-2})$  so that (80) becomes

$$\frac{1}{2}e^R = \frac{1}{4}e^{r_u + r_v}(1 - \cos(\theta_{uv}^*)) + O_{\mathbb{P}}(n^{2/(\tau-1)}). \quad (81)$$

We then use that by the definitions of  $t(u), t(v)$  and  $R$

$$e^{r_u + r_v} = e^R \left( e^{(r_u + r_v - R)/2} \right)^2 = e^R \left( \frac{n}{\nu t(u)t(v)} \right)^2. \quad (82)$$

This yields for (81) that

$$1 - \cos(\theta_{uv}^*) = 2 \left( \frac{\nu t(u)t(v)}{n} \right)^2 + O_{\mathbb{P}} \left( n^{-2\frac{\tau-2}{\tau-1}} \frac{\nu t(u)t(v)}{n} \right), \quad (83)$$

so that

$$\theta_{uv}^* = \cos^{-1}(1 - 2(\nu t(u)t(v)/n)^2(1 + o_{\mathbb{P}}(1))). \quad (84)$$

Because  $u$  and  $v$  are connected if their angular coordinates are at most  $\theta_{uv}^*$  and the angular coordinates of  $u$  and  $v$  are sampled uniformly, we obtain that

$$\mathbb{P}_n(X_{uv} = 1) = \frac{1}{\pi} \cos^{-1}(1 - 2(\nu t(u)t(v)/n)^2(1 + o_{\mathbb{P}}(1))). \quad (85)$$

□

Using this lemma, we now prove Theorem 2.3.

*Proof of Theorem 2.2.* We first focus on  $k \ll n^{(\tau-2)/(\tau-1)}$ . By [29, Eq. (21)],  $\mathbb{E}_n[D_u] = \frac{\nu(\tau-1)}{\pi(\tau-2)}t(u)$ . Thus, by [12, Theorem 2.7 and Lemma 3.5]  $D_u = \frac{\nu(\tau-1)}{\pi(\tau-2)}t(u)(1 + o_{\mathbb{P}}(1))$  when  $t(u) \gg 1$ . Therefore, when  $u \in M_{\varepsilon_n}(k)$ ,  $t(u) = \frac{\pi(\tau-2)}{\nu(\tau-1)}k(1 + o_{\mathbb{P}}(1))$  when  $k \gg 1$ . Since the types are distributed as (21), the largest type is  $O_{\mathbb{P}}(n^{1/(\tau-1)})$ . Therefore, if  $u \in M_{\varepsilon_n}(k)$ , then  $t(u)t(v)/n = o_{\mathbb{P}}(1)$  for all  $v$ . Taylor expanding (77) shows that for  $t(u)t(v) \ll n$

$$\mathbb{P}_n(X_{uv} = 1) = \frac{2\nu t(u)t(v)}{\pi n}(1 + o_{\mathbb{P}}(1)). \quad (86)$$

Thus, similarly as in (73) we obtain with  $\zeta = \pi/(2\nu)$ ,

$$a_{\varepsilon_n}(k) = (1 + o_{\mathbb{P}}(1)) \frac{\nu(\tau-1)}{\pi(\tau-2)k} \sum_{i \in [n]} t(i) \frac{t(i)t(u)}{\zeta n} = (1 + o_{\mathbb{P}}(1)) \sum_{i \in [n]} \frac{t(i)^2}{\zeta n}. \quad (87)$$

Combining this with the distribution of the types (21) proves Theorem 2.3(i) (which is the same as Theorem 2.2(i) where  $\mu$  is replaced by  $\zeta$  and  $c/(\tau-1)$  by 1).

We now investigate the case  $k \gg n^{(\tau-2)/(\tau-1)}$ . Similarly to (42), we define for the hyperbolic random graph

$$W_n^{k, \text{HRG}}(\delta) = \{u : t(u) \in [\delta \zeta n/k, \zeta n/(\delta k)]\} \quad (88)$$

with  $\zeta = \pi/(2\nu)$ . Using that  $\cos^{-1}(1 - 2x^2)/\pi \leq x$  combined with Lemma 4.1, we obtain

$$\mathbb{P}_n(X_{uv} = 1) \leq \min(t(u)t(v)/(\zeta n), 1). \quad (89)$$

Combining this with the fact that the  $t(u)$ s are sampled from a distribution similar to (3) shows that (70) also holds for the hyperbolic random graph, apart from a multiplicative constant. From there we can follow the same lines as the proof of Proposition 3.1.

We can also prove an analogous proposition to Proposition 3.2 which states that

$$\frac{a_{\varepsilon_n}(k, W_n^{k, \text{HRG}}(\delta))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} \frac{\nu(\tau-1)^2}{(\tau-2)\pi} \left(\frac{\pi}{2\nu}\right)^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} \min\left(\frac{1}{\pi} \cos^{-1}(1 - 2x^2), 1\right) dx. \quad (90)$$

Because the connection probabilities conditioned on the radial coordinates in the hyperbolic model are given by Lemma 4.1, a variant of Lemma 3.3 holds which states that

$$\mathbb{E}_n[a_{\varepsilon_n}(k, W_n^k(\delta))] = \frac{\nu(\tau-1)}{k\pi(\tau-2)} \sum_{u \in W_n^k(\delta)} t(u) \min\left(\frac{1}{\pi} \cos^{-1}(1 - 2(\nu t(u)t(v)/n)^2), 1\right) (1 + o_{\mathbb{P}}(1))$$

Similarly, a variant of Lemma 3.4 holds for the hyperbolic random graph, replacing the connection probability  $1 - e^{-D_i D_j / (\mu n)}$  of the erased configuration model by (77) and replacing the constant  $c$  from (3) by its equivalent constant for the hyperbolic model of  $\tau - 1$  and  $\mu$  by  $\zeta$  (see (21)). Furthermore, because conditionally on the radial coordinates, the probabilities that two distinct edges are present are independent, Lemma 3.5 also holds for the hyperbolic random graph. Therefore, similar steps that lead to (47) then show that

$$\frac{a_{\varepsilon_n}(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} \frac{\nu(\tau-1)^2}{(\tau-2)\pi} \left(\frac{\pi}{2\nu}\right)^{2-\tau} \int_0^{\infty} x^{1-\tau} \min\left(\frac{1}{\pi} \cos^{-1}(1 - 2x^2), 1\right) dx \quad (91)$$

which proves Theorem 2.3(ii).  $\square$

**Acknowledgements.** This work was supported by NWO TOP grant 613.001.451. The author would like to thank Remco van der Hofstad and Johan S. H. van Leeuwen for their useful comments.

## References

- [1] ALBERT, R., JEONG, H. AND BARABÁSI, A.-L. (1999). Internet: Diameter of the world-wide web. *Nature* **401**, 130–131.
- [2] BARABÁSI, A.-L. (2016). *Network Science*. Cambridge University Press.
- [3] BARRAT, A., BARTHLEMY, M., PASTOR-SATORRAS, R. AND VESPIGNANI, A. (2004). The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752.
- [4] BHAMIDI, S., DHARA, S., VAN DER HOFSTAD, R. AND SEN, S. (2017). Universality for critical heavy-tailed network models: metric structure of maximal components. arXiv:1703.07145.
- [5] BHAMIDI, S., SEN, S. AND WANG, X. (2016). Continuum limit of critical inhomogeneous random graphs. *Probab. Theory Related Fields* 1–77.
- [6] BHAMIDI, S., VAN DER HOFSTAD, R. AND HOOGHIEMSTRA, G. (2017). Universality for first passage percolation on sparse random graphs. *Ann. Probab.* **45**, 2568–2630.
- [7] BODE, M., FOUNTOLAKIS, N. AND MLLER, T. (2015). On the largest component of a hyperbolic model of complex networks. *Electron. J. Combin.* **22**, P3–24.
- [8] BODE, M., FOUNTOLAKIS, N. AND MLLER, T. (2016). The probability of connectivity in a hyperbolic model of complex networks. *Random Structures & Algorithms* **49**, 65–94.

- [9] BOGUÑÁ, M. AND PASTOR-SATORRAS, R. (2003). Class of correlated random networks with hidden variables. *Phys. Rev. E* **68**, 036112.
- [10] BOGUÑÁ, M., PASTOR-SATORRAS, R. AND VESPIGNANI, A. (2003). Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.* **90**, 028701.
- [11] BOLLOBÁS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1**, 311–316.
- [12] BRINGMANN, K., KEUSCH, R. AND LENGELER, J. (2015). Sampling geometric inhomogeneous random graphs in linear time. *arXiv:1511.00576v3*.
- [13] BRITTON, T., DEIJFEN, M. AND MARTIN-LÖF, A. (2006). Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.* **124**, 1377–1397.
- [14] CANDELLERO, E. AND FOUNTOLAKIS, N. (2014). *Clustering and the hyperbolic geometry of complex networks*. Springer International Publishing. pp. 1–12.
- [15] CATANZARO, M., BOGUÑÁ, M. AND PASTOR-SATORRAS, R. (2005). Generation of uncorrelated random scale-free networks. *Phys. Rev. E* **71**, 027103.
- [16] CHUNG, F. AND LU, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99**, 15879–15882.
- [17] COLOMER-DE SIMON, P. AND BOGUÑÁ, M. (2012). Clustering of random scale-free networks. *Phys. Rev. E* **86**, 026120.
- [18] DOROGOVTSSEV, S. N. (2004). Clustering of correlated networks. *Phys. Rev. E* **69**, 027104.
- [19] FALOUTSOS, M., FALOUTSOS, P. AND FALOUTSOS, C. (1999). On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*. vol. 29 ACM. pp. 251–262.
- [20] FRIEDRICH, T. AND KROHMER, A. (2015). Cliques in hyperbolic random graphs. In *INFO-COM proceedings 2015*. IEEE. pp. 1544–1552.
- [21] GUGELMANN, L., PANAGIOTOU, K. AND PETER, U. (2012). Random hyperbolic graphs: degree sequence and clustering. In *ICALP proceedings 2012, Part II*. Springer, Berlin, Heidelberg. pp. 573–585.
- [22] VAN DER HOFSTAD, R. (2017). *Random Graphs and Complex Networks Vol. 1*. Cambridge University Press.
- [23] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. AND VAN MIEGHEM, P. (2005). Distances in random graphs with finite variance degrees. *Random Structures & Algorithms* **27**, 76–123.
- [24] VAN DER HOFSTAD, R., JANSSEN, A. J. E. M., VAN LEEUWAARDEN, J. S. H. AND STEGEHUIS, C. (2017). Local clustering in scale-free networks with hidden variables. *Phys. Rev. E* **95**, 022307.
- [25] VAN DER HOFSTAD, R., VAN DER HOORN, P., LITVAK, N. AND STEGEHUIS, C. (2017). Limit theorems for assortativity and clustering in the configuration model with scale-free degrees. *arXiv:1712.08097*.
- [26] VAN DER HOFSTAD, R., VAN LEEUWAARDEN, J. S. H. AND STEGEHUIS, C. (2017). Optimal subgraph structures in scale-free c. *arXiv:1709.03466*.
- [27] JANSON, S. AND LUCZAK, M. J. (2007). A simple solution to the k-core problem. *Random Structures & Algorithms* **30**, 50–62.
- [28] JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z. N. AND BARABÁSI, A.-L. (2000). The large-scale organization of metabolic networks. *Nature* **407**, 651–654.
- [29] KRIUKOV, D., PAPADOPOULOS, F., KITSAK, M., VAHDAT, A. AND BOGUNÁ, M. (2010). Hyperbolic geometry of complex networks. *Phys. Rev. E* **82**, 036106.
- [30] LESKOVEC, J. AND KREVL, A. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data> 2014. Date of access: 14/03/2017.
- [31] MASLOV, S., SNEPPEN, K. AND ZALIZNYAK, A. (2004). Detection of topological patterns in complex networks: correlation profile of the internet. *Phys. A* **333**, 529 – 540.
- [32] MAYO, M., ABDELZAHER, A. AND GHOSH, P. (2015). Long-range degree correlations in complex networks. *Computational Social Networks* **2**, 4.
- [33] NEWMAN, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701.
- [34] OSTILLI, M. (2014). Fluctuation analysis in complex networks modeled by hidden-variable models: Necessity of a large cutoff in hidden-variable models. *Phys. Rev. E* **89**, 022807.

- [35] PARK, J. AND NEWMAN, M. E. J. (2003). Origin of degree correlations in the internet and other networks. *Phys. Rev. E* **68**, 026112.
- [36] PASTOR-SATORRAS, R., VÁZQUEZ, A. AND VESPIGNANI, A. (2001). Dynamical and correlation properties of the internet. *Phys. Rev. Lett.* **87**, 258701.
- [37] RAVASZ, E. AND BARABÁSI, A.-L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112.
- [38] SERRANO, M. A. AND BOGUÑÁ, M. (2006). Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.* **97**, 088701.
- [39] STEGEHUIS, C., VAN DER HOFSTAD, R., VAN LEEUWAARDEN, J. S. H. AND JANSSEN, A. J. E. M. (2017). Clustering spectrum of scale-free networks. *Phys. Rev. E* **96**,
- [40] VAN DEN ESKEER, H., VAN DER HOFSTAD, R. AND HOOGHIEMSTRA, G. (2008). Universality for the distance in finite variance random graphs. *J. Stat. Phys.* **133**, 169–202.
- [41] VÁZQUEZ, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E* **67**, 056104.
- [42] VÁZQUEZ, A., PASTOR-SATORRAS, R. AND VESPIGNANI, A. (2002). Large-scale topological and dynamical properties of the internet. *Phys. Rev. E* **65**, 066130.
- [43] WHITT, W. (2006). *Stochastic-Process Limits*. Springer, New York.
- [44] YAO, D., VAN DER HOORN, P. AND LITVAK, N. (2017). Average nearest neighbor degrees in scale-free networks. arXiv:1704.05707v2.