# Degree correlations in scale-free null models

# Degree correlations in scale-free null models

Clara Stegehuis

Department of Mathematics and Computer Science
Eindhoven University of Technology

September 6, 2017

**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 consider the large-network limit when the number of nodes $n$ tends to infinity. We find for all three null models that $a(k)$ 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 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 [1,18,27,41]. 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 [23, 38] and the occurrence of subgraphs [33]. Clustering can be measured in terms of the probability $c(k)$ that a degree-$k$ vertex creates triangles. Both empirically [30, 35] and theoretically [16, 38] it was shown that $c(k)$ falls off with $k$, and hence that hubs are less likely to take part in triadic closures. This phenomenon can be explained by viewing the scale-free networks as an hierarchical structure, in which the hubs are not part of communities, but instead connect several communities of small dense collections of vertices. Triangles then occur within and not between the communities.

Where 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 nodes in terms of $a(k)$, the average nearest neighbor degree of vertices of degree $k$. According to several studies [2,14], 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, 14, 37]. We also provide support for his 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 providing 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, 14, 31, 34, 35, 40, 43]. 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 [32]. 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

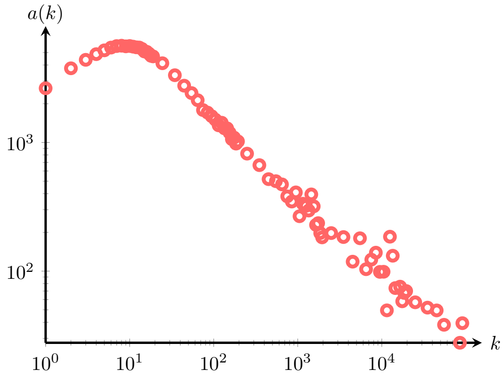



Figure 1: $a(k)$ for the Youtube friendship network [29]

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 $a(k)$ as $k$ grows large [30, 35]. Figure 1 illustrates this for the Youtube friendship network [29]. Hence, such 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 and hierarchical features 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, the uncorrelated structure is still visible for small-degree vertices. We then identify the value of $k$ as of which $a(k)$ starts decaying, and the degree correlations start playing a role. 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. Thus, single-edge constraints may cause the decaying $a(k)$. We show that $k \mapsto a(k)$, for all three null models with $n$ vertices, remains constant until $k = n^{(\tau-1)/(\tau-2)}$ and then settles on the power law $a(k) \sim k^{\tau-3}$ with an exponent depending on $\tau$.

## 2 Main results

We first define the average nearest neighbor degree $a(k)$ 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 is then defined as

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

While it is possible that no vertex with degree equal to $k$ exists, definition (1) should be understood as that at least one vertex of degree $k$ is present. We will now analyze $a(k)$, 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 $(d_1, d_2, \ldots, d_n)$ 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 $j = 1, \ldots, n$. 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., [21, 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 [12]. 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 degrees to be an i.i.d. sample from the distribution

$$\mathbb{P}(D = k) = ck^{-\tau}(1 + o(1)), \quad \text{when } k \to \infty, \tag{2}$$

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.

Here is the main result for the erased configuration model:

**Theorem 2.1** ($a(k)$ in the erased configuration model)**.** *Let $G$ be an erased configuration model, where the degrees are an i.i.d. sample from* (2) *and let $\Gamma$ denote the Gamma function.*

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

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

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

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

$$\frac{a(k)}{n^{3-\tau}k^{\tau-3}} \xrightarrow{d} -c\mu^{2-\tau}\Gamma(2-\tau). \tag{4}$$

**Remark 2.1.** *The convergence in* (3) *also holds jointly in $k$ and $n$, so that for $m \geq 1$ and $1 \leq k_1 < k_2 < \cdots < k_m \ll n^{(\tau-2)/(\tau-1)}$,*

$$\frac{(a(k_1), \ldots, a(k_m))}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu}\left(\frac{2c\Gamma(\frac{5}{2}-\frac{1}{2}\tau)}{(\tau-1)(3-\tau)}\cos\left(\frac{\pi(\tau-1)}{4}\right)\right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}\mathbf{1}, \tag{5}$$

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

Figure 2 illustrates the behavior of $a(k)$. First, $a(k)$ stays flat and does not depend on $k$. After that, $a(k)$ 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 $a(k)$ (see Propositions 3.1 and 3.2) comes from vertices with

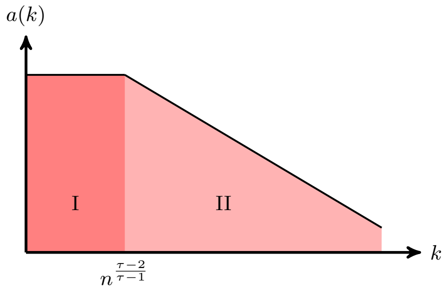



Figure 2: Illustration of the behavior of $a(k)$ in the erased configuration model

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, $a(k)$ converges to a stable random variable, as was also shown in [43] 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 $a(k)$ still forms a flat curve in $k$ for one realization of an erased configuration model when $k$ is small. In contrast, $a(k)$ converges to a constant for large $k$-values, so that different realizations of erased configuration models will result in similar $a(k)$-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 [22]

$$1 - \mathrm{e}^{-D_i D_j/\mu n}. \tag{6}$$

Let $\mathbb{1}_{\{i \leftrightarrow k\}}$ denote the indicator that vertex $i$ is connected to a randomly chosen vertex of degree $k$. The expected degree of a neighbor of a vertex with degree $k$ can then be approximated by

$$a(k) \approx k^{-1} \sum_{i\in[n]} D_i \mathbb{P}\left(i \leftrightarrow k\right) \approx k^{-1} \sum_{i\in[n]} D_i (1 - \mathrm{e}^{-D_i k/(\mu n)}). \tag{7}$$

The maximum degree in an i.i.d. sample from (2) 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

$$a(k) \approx \frac{1}{\mu n} \sum_{i\in[n]} D_i^2. \tag{8}$$

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 (7) by the integral

$$a(k) \approx cnk^{-1} \int_1^\infty x^{1-\tau}(1 - \mathrm{e}^{-xk/(\mu n)})\mathrm{d}x = c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \int_{k/(\mu n)}^\infty y^{1-\tau}(1 - \mathrm{e}^{-y})\mathrm{d}y, \tag{9}$$

using the degree distribution (2) and the change of variables $y = xk/(\mu n)$. When $k \ll n$, we can approximate this by

$$a(k) \approx c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \int_0^\infty y^{1-\tau}(1 - \mathrm{e}^{-y})\mathrm{d}y = -c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \Gamma(2-\tau). \tag{10}$$

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, 15]. 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 (2). 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

$$p(h,h') = \min\left(\frac{hh'}{\mu n}, 1\right), \tag{11}$$

which is the Chung-Lu version of the rank-1 inhomogeneous random graph [15]. 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** ($a(k)$ in the rank-1 inhomogeneous random graph)**.** *Let $G$ be a rank-1 inhomogeneous random graph, where the weights are an i.i.d. sample from* (2) *and $\Gamma$ denotes the Gamma function.*

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

$$\frac{a(k)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu}\left(\frac{2c\Gamma(\frac{5}{2}-\frac{1}{2}\tau)}{(\tau-1)(3-\tau)}\cos\left(\frac{\pi(\tau-1)}{4}\right)\right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \tag{12}$$

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

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

$$\frac{a(k)}{n^{3-\tau}k^{\tau-3}} \xrightarrow{d} \frac{c\mu^{2-\tau}}{(3-\tau)(\tau-2)}. \tag{13}$$

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. Why $a(k)$ is highly similar in both models 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 (6) which is close to $\min(1, \frac{D_iD_j}{\mu n})$, the connection probability in the rank-1 inhomogeneous random graph. Similar arguments that lead to (7) show that $a(k)$ can be approximated by

$$a(k) \approx k^{-1}\sum_{i\in[n]} h_i \min(h_i k/\mu n, 1)\mathrm{d}x. \tag{14}$$

This sum behaves very similarly to the sum in (7), so that the only difference between Theorem 2.1 and 2.2 is the limiting constants in (4) and (13). 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 $a(k)$ 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 [28] and samples $n$ vertices on a disk of radius $R = 2\log(n/\nu)$, where the density of the radial coordinate $r$ a vertex $p = (r, \phi)$ is

$$\rho(r) = \alpha\frac{\sinh(\alpha r)}{\cosh(\alpha R) - 1} \tag{15}$$

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 $v = (r_v, \phi_v)$ satisfies

$$\cosh(\mathrm{d}(u, v)) = \cosh(r_u)\cosh(r_v) - \sinh(r_u)\sinh(r_v)\cos(\Delta\theta), \tag{16}$$

where $\Delta\theta$ denotes the relative angle between $\phi_u$ and $\phi_v$. This creates a simple random graph with power-law degrees with exponent $\tau$ [28]. The parameter $\nu$ fixes the average degree $\bar{k}$ 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 [13,28]. In both the rank-1 inhomogeneous random graph and the erased configuration model, the connection probabilities of different pairs 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 geometric connection probabilities. However, if we define

$$t(u) = \mathrm{e}^{(R - r_u)/2} \tag{17}$$

then we can write the probability that two randomly chosen vertices $u$ and $v$ are connected as [7]

$$\mathbb{P}\left(u \leftrightarrow v \mid t(u), t(v)\right) = \min\left(\frac{2\nu t(u)t(v)}{n}, 1\right)(1 + o(1)), \tag{18}$$

which is very similar to the connection probability in the rank-1 inhomogeneous random graph. Furthermore, by [7, Lemma 1.3],

$$\mathbb{P}\left(t(u) > x\right) = \tfrac{\tau-1}{2} x^{-\tau+1}(1 + o(1)), \tag{19}$$

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.

The next theorem shows that this high level equivalence to the rank-1 inhomogeneous random graph makes $a(k)$ in the hyperbolic random graph is similar to $a(k)$ in the rank-1 inhomogeneous random graph:

**Theorem 2.3** ($a(k)$ in the hyperbolic random graph)**.** *Let $G$ be a hyperbolic random graph with power-law degrees with exponent $\tau$ and parameter $\nu$.*

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

$$\frac{a(k)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} = \frac{2\nu}{\pi}\left(\frac{\tau-1}{3-\tau}\Gamma(\tfrac{5}{2} - \tfrac{1}{2}\tau)\cos\left(\frac{\pi(\tau-1)}{4}\right)\right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \tag{20}$$

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

2. *For $k \gg n^{(\tau-2)/(\tau-1)}$*

$$\frac{a(k)}{n^{3-\tau}k^{\tau-3}} \xrightarrow{d} \frac{(\tau-1)^2}{2(3-\tau)(\tau-2)}\left(\frac{\pi}{2\nu}\right)^{2-\tau}. \tag{21}$$

## 2.4 Discussion

**Universality.** The behavior of $a(k)$ is universal across the three null models we considered. 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] or in terms of distances [39] when $\tau > 3$, and in terms of clustering when $\tau \in (2, 3)$ [38]. The hyperbolic random graph typically shows different behavior, for example in terms of clustering [13, 20], 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 $r_j \approx R - 2\log(n/(k\nu)) = 2\log(k)$. Then, $r_j \approx R - r_i$, so that the major contributing vertices all have radial coordinate proportional to $R - r_i$.

**Expected average nearest neighbor degree.** In Theorems 2.1-2.3 we show that $a(k)$ converges in probability to a stable distribution 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 $a(k)$ across the different samples will look like a stable distribution. We can also study the expected value of $a(k)$ across the different samples. For the rank-1 inhomogeneous random graph for example, using similar techniques as in the proof of Theorem 2.2(ii), we can show that

$$\frac{\mathbb{E}\left[a(k)\right]}{(n/k)^{3-\tau}} \to \frac{c\mu^{2-\tau}}{(3-\tau)(\tau-2)} \tag{22}$$

for $k \gg 1$ as $n \to \infty$. The difference between the scaling of the expected value of $a(k)$ and the typical behavior of $a(k)$ in Theorem 2.2(i) is caused by high-degree vertices. In typical degree sequences, the maximum degree will be 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. This shows that the expected value of $a(k)$ is not very informative when $k$ is small, since Theorem 2.2 shows that $a(k)$ will almost always be smaller than its expected value when $k$ is small.

Figure 3 illustrates this difference in terms of the mean and median value of $a(k)$ 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 is a linear function of $k$ over the entire range of $k$, where the median shows the straight part of the curve from Theorem 2.2. Thus, it is important to distinguish between mean and median when simulating random graphs.

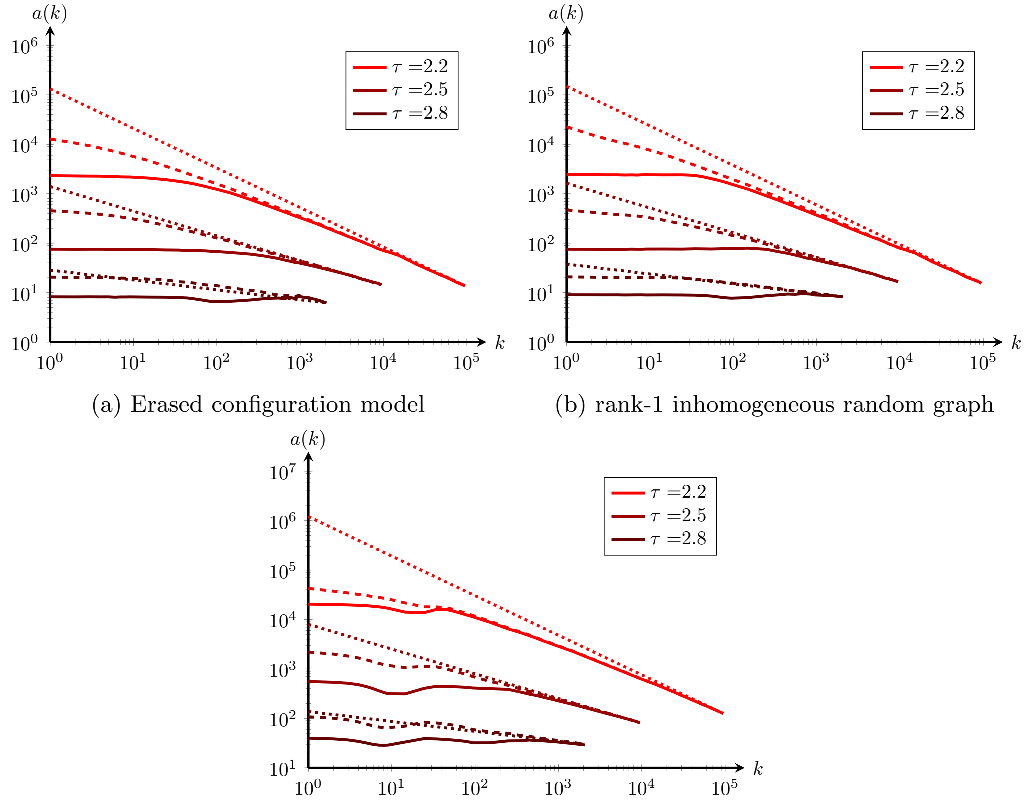
**Fixed degrees.** In the proof of Theorem 2.1 we show that the fluctuations that come with the stable laws for small $k$ are not present when we condition on the degree sequence. Thus, the large fluctuations in $a(k)$ for small $k$ are only caused by fluctuations of the i.i.d. degrees, weights or radii. For a given real-world network, the degrees of the network are often preserved, and many samples of erased configuration models or rank-1 inhomogeneous random graphs for the observed degree sequence are created. In this setting, the degrees are fixed, so that the sample-to-sample fluctuations of $a(k)$ will be 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 af 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 $k$ [9,23,36,38,41]. 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 $c(k) < a(k)/k$ [37]. 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 [23]

$$c(k) = c^2\Gamma(2-\tau)^2\mu^{3-2\tau}n^{5-2\tau}k^{2\tau-6}(1 + o_{\mathbb{P}}(1)). \tag{23}$$

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

$$c(k) = \frac{a^2(k)}{\mu n}(1 + o_{\mathbb{P}}(1)). \tag{24}$$




(a) Erased configuration model

(b) rank-1 inhomogeneous random graph

(c) hyperbolic random graph

Figure 3: $a(k)$ for different random graph models with $n = 10^6$. The solid line is the median of $a(k)$ 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 $k^{\tau-3}$.

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 $1 - \mathrm{e}^{-a(k)^2/\mu n} \approx a(k)^2/\mu n$. 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 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 (24) is also the correct relation between clustering and degree correlations in the rank-1 inhomogeneous random graph [38]. 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 vertices $i, j$, vertices $i, k$ and vertices $j, k$ are (almost) independent. In the hyperbolic random graph, the geometry causes a strong dependence between these connection probabilities. If we know that vertices $j$ and $k$ are neighbors of $i$, they are likely to be geometrically close to one another. This makes the probability that $j$ and $k$ larger than in the rank-1 inhomogeneous random graph or the erased configuration model. These correlations

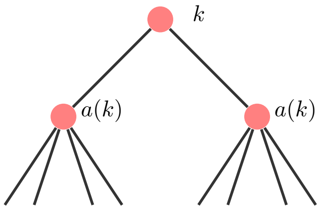



Figure 4: The neighbors of a vertex of degree $k$ have average degree $a(k)$

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 [19], even though cliques clearly involve three-point correlations. Cliques in the hyperbolic random graph are typically formed between vertices at radius proportional to $R/2$ [19], 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 [26]. In the erased configuration model, many other small subgraphs typically occur between vertices of degrees proportional to $\sqrt{n}$ [24]. 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. When $k = o(n^{\frac{\tau-2}{\tau-1}})$, we couple the degrees of a uniformly chosen neighbor 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 there is a specific range of degrees that contributes to $a(k)$.

## 3.1 Preliminaries

We often want to interchange the sampled degree of a vertex $i$, $D_i$ and its erased degree $D_i^{(\mathrm{er})}$. 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 $X_n/b_n \xrightarrow{\mathbb{P}} 0$. Then, by [12, Eq. A(9)]

$$D_i^{(\mathrm{er})} = D_i(1 + o_{\mathbb{P}}(1)), \tag{25}$$

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 event

$$J_n = \{|L_n - \mu n| \leq n^{-1/(\tau-1)}\}. \tag{26}$$

By [25], $\mathbb{P}(J_n) \to 1$ as $n \to \infty$. In the rest of this section, we will often condition on the degree sequence. For some event $\mathcal{E}$, we use the notation $\mathbb{P}_n(\mathcal{E}) = \mathbb{P}\left(\mathcal{E} \mid (D_i)_{i\in[n]}\right)$, and we define $\mathbb{E}_n$ and $\mathrm{Var}_n$ similarly.

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

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

$$\mathbb{P}_n\left(D_n^*=k\right)=\frac{k}{L_n}\sum_{i\in[n]}\mathbb{1}_{\{D_i=k\}}. \tag{27}$$

We then use this coupling to compute $a(k)$.

*Proof of Theorem 2.1(i).* We first condition on the degree sequence $(D_i)_{i\in[n]}$. Let $v$ be a uniformly chosen 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. We use a similar coupling as in [6, Construction 4.2] to couple the degrees of the neighbors of $v$, $B_i$ to i.i.d. samples of $D_n^*$, $Y_i$. 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. Then, $Y_i=d_{v_i'}$, where $v_i'$ denotes the vertex incident to the $i$th drawn half-edge. Then for $i\in[k]$ the coupling is defined in the following way:

- If $v_i'\notin\{v,v_1,\dots,v_{i-1}\}$, then $B_i=Y_i$ and $v_i=v_i'$. We say that $B_i$ and $Y_i$ are successfully coupled.
- If $v_i'\in\{v,v_1,\dots,v_{i-1}\}$, we redraw a uniformly chosen half-edge from the set of half-edges incident to $\{v,v_1,\dots,v_{i-1}\}$. Let $v_i$ denote the vertex incident to the chosen half-edge, and $B_i=d_{v_i}$. We then say that $B_i$ and $Y_i$ are miscoupled.

By [6, Lemma 4.3], the probability of a miscoupling at step $i$ can be bounded as

$$\mathbb{P}_n\left(B_i\neq Y_i\mid\mathcal{F}_{i-1}\right)\leq {L_n}^{-1}\Big(k+\sum_{s=1}^{i-1}B_s\Big). \tag{28}$$

Thus, the expected number of miscouplings up to time $t$, $N_{\text{mis}}(t)$, satisfies

$$\mathbb{E}_n\left[N_{\text{mis}}(t)\right]\leq\frac{kt}{L_n}+\frac{1}{L_n}\sum_{i=1}^{t}\sum_{s=1}^{i-1}\mathbb{E}_n\left[B_s\right]. \tag{29}$$

When $B_s$ is successfully coupled, $\mathbb{E}_n\left[B_s\right]=\mathbb{E}_n\left[D_n^*\right]=\sum_i D_i^2/L_n$. When $B_s$ is not successfully coupled, it is drawn in a size-biased manner from the vertices that are not chosen yet. Let $V_s=\{v,v_1,\dots,v_{s-1}\}$. Then

$$\mathbb{E}_n\left[B_s\right]=\frac{\sum_{i\notin V_s}D_i^2}{\sum_{i\notin s}D_i}\leq\frac{\sum_{i\in[n]}D_i^2}{\sum_{i\in[n]}D_i-\sum_{i\in V_s}D_i}=\frac{\sum_{i\in[n]}D_i^2}{\sum_{i\in[n]}D_i}\left(1+\frac{\sum_{i\in V_s}D_i}{\sum_{i\in[n]}D_i-\sum_{i\in V_s}D_i}\right). \tag{30}$$

Since $D_{\max}=O_{\mathbb{P}}\left(n^{1/(\tau-1)}\right)$, $\sum_{i\in V_s}D_i=O_{\mathbb{P}}\left(sn^{1/(\tau-1)}\right)$. Thus, as long as $s=o(n^{(\tau-2)/(\tau-1)})$,

$$\mathbb{E}_n\left[B_s\right]=O_{\mathbb{P}}\Big(L_n^{-1}\sum_{i\in[n]}D_i^2\Big)=O_{\mathbb{P}}\left(n^{(3-\tau)/(\tau-1)}\right), \tag{31}$$

where the last step follows from the Stable Law Central Limit Theorem (see for example [42, Theorem 4.5.2]). Then, for $k=o(n^{(\tau-2)/(\tau-1)})$

$$\mathbb{E}_n\left[N_{\text{mis}}(k)\right]=\frac{k^2}{L_n}+\frac{1}{L_n}O_{\mathbb{P}}\left(n^{(3-\tau)/(\tau-1)}\right)\sum_{i=1}^{k}(i-1)=O_{\mathbb{P}}\left(k^2n^{2\frac{2-\tau}{\tau-1}}\right). \tag{32}$$

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

$$\mathbb{E}_n\left[N_{\text{mis}}(k)\right]=o_{\mathbb{P}}(1). \tag{33}$$

Then, by the Markov inequality

$$\mathbb{P}_n\left(N_{\rm mis}(k)=0\right)=1-\mathbb{P}_n\left(N_{\rm mis}(k)\geq 1\right)\geq 1-\mathbb{E}_n\left[N_{\rm mis}(k)\right]=1-o_{\scriptscriptstyle\mathbb{P}}(1). \tag{34}$$

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 $D_i^{(\rm er)}\leq D_i$ and $D_i^{(\rm er)}=D_i(1+o_{\scriptscriptstyle\mathbb{P}}(1))$, conditionally on the degrees

$$a(k)\mid(D_i)_{i\geq 1}=\frac{1}{kN_k}\sum_{i:D_i^{(\rm er)}=k}\sum_{j\in\mathcal{N}_i}D_j^{(\rm er)}=\frac{1}{k}\mathbb{E}_n\left[\sum_{j\in\mathcal{N}_{V_k}}D_j^{(\rm er)}\right]=(1+o_{\scriptscriptstyle\mathbb{P}}(1))\mathbb{E}_n\left[D_{\mathcal{N}_{V_k}(U)}\right], \tag{35}$$

where $V_k$ denotes a uniformly chosen vertex of degree $k$, and $\mathcal{N}_{V_k}(U)$ is a uniformly chosen neighbor of vertex $V_k$. With high probability, we can couple the degrees in the neighborhood of a uniformly chosen vertex of degree $k$ to i.i.d copies of $D_n^*$. Then,

$$a(k)\mid(D_i)_{i\geq 1}=(1+o_{\scriptscriptstyle\mathbb{P}}(1))\mathbb{E}_n\left[D_n^*\right]=(1+o_{\scriptscriptstyle\mathbb{P}}(1))L_n^{-1}\sum_{i\in[n]}D_i^2. \tag{36}$$

Note that this expression is independent of $k$. Using that for $t$ large

$$\mathbb{P}\left(D^2>t\right)=\mathbb{P}\left(D>\sqrt{t}\right)=\frac{c}{\tau-1}t^{(1-\tau)/2}(1+o(1)) \tag{37}$$

we can again use the Stable Law Central Limit Theorem to conclude that

$$\frac{a(k)}{n^{(3-\tau)/(\tau-1)}}\xrightarrow{d}\frac{1}{\mu}\left(\frac{2c}{(\tau-1)(3-\tau)}\Gamma(\tfrac{5}{2}-\tfrac{1}{2}\tau)\cos\left(\frac{\pi(\tau-1)}{4}\right)\right)^{2/(\tau-1)}\mathcal{S}_{(\tau-1)/2}, \tag{38}$$

where $\mathcal{S}_{(\tau-1)/2}$ is a stable distribution. The fact that (36) is independent of $k$ proves the joint convergence of Remark 2.1. □

## 3.3 Large $k$

Now we study the value of $a(k)$ when $k$ grows large. For $k\gg n^{(\tau-2)/(\tau-1)}$ there exists a specific range of degrees $W_n^k(\varepsilon)$ which gives the largest contribution to $a(k)$. We define

$$W_n^k(\varepsilon)=\left\{u:D_u\in[\varepsilon n/k,n/(\varepsilon k)]\right\}, \tag{39}$$

and we write

$$a(k)=a(k,W_n^k(\varepsilon))+a(k,\bar{W}_n^k(\varepsilon)), \tag{40}$$

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

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

$$\limsup_{n\to\infty}\frac{\mathbb{E}\left[a(k,\bar{W}_n^k(\varepsilon))\right]}{(n/k)^{3-\tau}}=O\left(\varepsilon^{\kappa}\right). \tag{41}$$

**Proposition 3.2** (Major contributions)**.**

$$\frac{a(k,W_n^k(\varepsilon))}{(n/k)^{3-\tau}}\xrightarrow{\mathbb{P}}c\mu^{2-\tau}\int_{\varepsilon}^{1/\varepsilon}x^{1-\tau}(1-{\rm e}^{-x}){\rm d}x \tag{42}$$

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,

$$\frac{a(k,\bar{W}_n^k(\varepsilon))}{(n/k)^{3-\tau}} = O_{\mathbb{P}}\left(\varepsilon^{\kappa}\right). \tag{43}$$

Combining this with Proposition 3.2 results in

$$\frac{a(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau}\int_{\varepsilon}^{1/\varepsilon} x^{1-\tau}(1-\mathrm{e}^{-x})\mathrm{d}x + O_{\mathbb{P}}\left(\varepsilon^{\kappa}\right). \tag{44}$$

Taking the limit of $\varepsilon \to 0$ then proves the theorem. □

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 $a(k, W_n^k(\varepsilon))$ when we condition on the degree sequence.

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

$$\mathbb{E}_n\left[a(k,W_n^k(\varepsilon))\right] = \frac{1}{k}\sum_{u\in W_n^k(\varepsilon)} D_u(1-\mathrm{e}^{-D_uk/L_n})(1+o_{\mathbb{P}}(1)). \tag{45}$$

*Proof.* It suffices to prove the lemma under the event $J_n$ from (26), since $\mathbb{P}(J_n) \to 1$, so that we may assume that $L_n = \mu n(1+o(1))$. Let $\hat{X}_{ij}$ denote the number of edges between $i$ and $j$ in the erased configuration model, so that $\hat{X}_{ij} \in \{0,1\}$. If $w$ is a uniformly chosen vertex such that $D_w^{(\mathrm{er})} = k$, by (1)

$$\mathbb{E}_n\left[a(k,W_n^k(\varepsilon))\right] = \frac{1}{kN_k}\sum_{v:D_v^{(\mathrm{er})}=k}\sum_{u\in W_n^k(\varepsilon)} \mathbb{E}_n\left[D_u^{(\mathrm{er})}\mathbb{1}_{\{u\leftrightarrow v\}}\right] = \frac{1}{k}\sum_{u\in W_n^k(\varepsilon)} D_u^{(\mathrm{er})}\mathbb{P}_n(\hat{X}_{uw}=1). \tag{46}$$

By [22, Eq. (4.9)]

$$\mathbb{P}_n\left(\hat{X}_{uw}=1\right) = 1-\mathrm{e}^{D_uD_w/L_n} + O\left(\frac{D_w^2D_u + D_u^2D_w}{L_n^2}\right) = (1-\mathrm{e}^{D_uD_w/L_n})(1+o_{\mathbb{P}}(1)), \tag{47}$$

where the last step follows because $D_u \in n/k[\varepsilon, 1/\varepsilon]$ and by (25) $D_w = k(1+o_{\mathbb{P}}(1))$. Using that $D_u^{(\mathrm{er})} = D_u(1+o_{\mathbb{P}}(1))$ and $D_w = k(1+o_{\mathbb{P}}(1))$ we can write (46) as

$$\begin{aligned}\mathbb{E}_n\left[a(k,W_n^k(\varepsilon))\right] &= \frac{1}{k}\sum_{u\in W_n^k(\varepsilon)} D_u(1-\mathrm{e}^{-D_uk/L_n}\mathrm{e}^{o_{\mathbb{P}}(D_u/L_n)})(1+o_{\mathbb{P}}(1))\\ &= \frac{1}{k}\sum_{u\in W_n^k(\varepsilon)} D_u(1-\mathrm{e}^{-D_uk/L_n})(1+o_{\mathbb{P}}(1))\end{aligned} \tag{48}$$

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

### 3.3.2 Convergence of conditional expectation

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

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

$$\frac{\mathbb{E}_n\left[a(k,W_n^k(\varepsilon))\right]}{n^{3-\tau}k^{\tau-3}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau}\int_{\varepsilon}^{1/\varepsilon} x^{1-\tau}(1-\mathrm{e}^{-x})\mathrm{d}x. \tag{49}$$

*Proof.* Define the random measure

$$M^{(n)}[a,b] = \frac{1}{\mu^{1-\tau}n^{2-\tau}k^{\tau-1}} \sum_{u\in[n]} \mathbb{1}_{\{D_u\in[a,b]\mu n/k\}}. \tag{50}$$

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,

$$\begin{aligned} M^{(n)}[a,b] &= \frac{1}{\mu^{1-\tau}n^{2-\tau}k^{\tau-1}} |\{u : D_u \in [a,b]\mu n/k\}| \xrightarrow{\mathbb{P}} \frac{1}{(\mu n)^{1-\tau}k^{\tau-1}} \mathbb{P}\left(D \in [a,b]\mu n/k\right) \\ &= \frac{1}{(\mu n)^{1-\tau}k^{\tau-1}} \int_{a\mu n/k}^{b\mu n/k} cx^{-\tau}\mathrm{d}x = \int_a^b cy^{-\tau}\mathrm{d}y =: \lambda[a,b], \end{aligned} \tag{51}$$

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

$$\begin{aligned} \mathbb{E}_n\left[a(k, W_n^k(\varepsilon))\right] &= \frac{\sum_{u\in W_n^k(\varepsilon)} D_u(1-\mathrm{e}^{-D_u k/L_n})}{k}(1+o_{\mathbb{P}}(1)) \\ &= \frac{\mu n}{k}\frac{\sum_{u\in W_n^k(\varepsilon)} \frac{D_u k}{\mu n}(1-\mathrm{e}^{-D_u k/L_n})}{k}(1+o_{\mathbb{P}}(1)) \\ &= \frac{\mu^{2-\tau}n^{3-\tau}}{k^{3-\tau}} \int_\varepsilon^{1/\varepsilon} t(1-\mathrm{e}^{-t})\mathrm{d}M^{(n)}(t)(1+o_{\mathbb{P}}(1)). \end{aligned} \tag{52}$$

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

$$\left| t(1-\mathrm{e}^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t\in B_i\}} \right| < \eta/\lambda([\varepsilon,1/\varepsilon]), \tag{53}$$

for all $t \in [\varepsilon, 1/\varepsilon]$. 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,

$$\begin{aligned} \Big| \int_\varepsilon^{1/\varepsilon} t(1-\mathrm{e}^{-t})\mathrm{d}M^{(n)}(t) - \int_\varepsilon^{1/\varepsilon} t(1-\mathrm{e}^{-t})\mathrm{d}\lambda(t) \Big| &\leq \Big| \int_\varepsilon^{1/\varepsilon} t(1-\mathrm{e}^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t\in B_i\}}\mathrm{d}M^{(n)}(t) \Big| \\ &\quad + \Big| \int_\varepsilon^{1/\varepsilon} t(1-\mathrm{e}^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t\in B_i\}}\mathrm{d}\lambda(t) \Big| \\ &\quad + \Big| \sum_{i=1}^m b_i(M^{(n)}(B_i) - \lambda(B_i)) \Big| \\ &\leq \eta M^{(n)}([\varepsilon,1/\varepsilon])/\lambda([\varepsilon,1/\varepsilon]) + \eta + o_{\mathbb{P}}(\eta). \end{aligned} \tag{54}$$

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

$$\int_\varepsilon^{1/\varepsilon} t(1-\mathrm{e}^{-t})\mathrm{d}M^{(n)}(t) \xrightarrow{\mathbb{P}} \int_\varepsilon^{1/\varepsilon} t(1-\mathrm{e}^{-t})\mathrm{d}\lambda(t) = c\int_\varepsilon^{1/\varepsilon} x^{1-\tau}(1-\mathrm{e}^{-x})\mathrm{d}x, \tag{55}$$

which proves the lemma. □

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

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

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

$$\frac{\mathrm{Var}_n\left(a(k, W_k^k(\varepsilon))\right)}{\mathbb{E}_n\left[a(k, W_n^k(\varepsilon))\right]^2} \xrightarrow{\mathbb{P}} 0. \tag{56}$$

*Proof.* Again, it suffices to prove the lemma under the event $J_n$ from (26). We denote $\mathcal{S}_k = \{i \in [n] : D_i^{(\mathrm{er})} = k\}$. We write the variance of $a(k, W_n^k(\varepsilon))$ as

$$\begin{aligned}\mathrm{Var}_n\left(a(k, W_n^k(\varepsilon))\right) &= \frac{1}{k^2 N_k^2} \sum_{i,j \in \mathcal{S}_k} \sum_{u,v \in W_n^k(\varepsilon)} D_u^{(\mathrm{er})} D_w^{(\mathrm{er})} \\ &\quad \times \left(\mathbb{P}_n\left(X_{iu} = X_{jv} = 1\right) - \mathbb{P}_n\left(X_{iu} = 1\right)\mathbb{P}_n\left(X_{jv} = 1\right)\right) \\ &= \frac{(1 + o_{\mathbb{P}}(1))}{k^2 N_k^2} \sum_{i,j \in \mathcal{S}_k} \sum_{u,v \in W_n^k(\varepsilon)} D_u D_w \\ &\quad \times \left(\mathbb{P}_n\left(X_{iu} = X_{jv} = 1\right) - \mathbb{P}_n\left(X_{iu} = 1\right)\mathbb{P}_n\left(X_{jv} = 1\right)\right). \end{aligned} \tag{57}$$

Equation (57) 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

$$\mathbb{P}_n\left(X_{iu} = X_{jv} = 0\right) = \mathbb{P}_n\left(X_{iu} = 0\right)\mathbb{P}_n\left(X_{jv} = 0 \mid X_{iu} = 0\right). \tag{58}$$

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 $j$, as well as the degree of vertex $v$. Since the expected number of half-edges from $i$ that pair to vertex $j$ is $O(D_i D_j / L_n)$ [17], the new degree of vertex $j$ is $D_j(1 + o_{\mathbb{P}}(1))$, and a similar statement holds for vertex $v$. Thus, by (47)

$$\mathbb{P}_n\left(X_{iu} = X_{jv} = 0\right) = \mathrm{e}^{-D_i D_u / L_n} \mathrm{e}^{-D_j D_v / L_n}(1 + o_{\mathbb{P}}(1)). \tag{59}$$

This results in

$$\begin{aligned}\mathbb{P}_n\left(X_{iu} = X_{jv} = 1\right) &= 1 - \mathbb{P}_n\left(X_{iu} = 0\right) - \mathbb{P}_n\left(X_{jv} = 0\right) + \mathbb{P}_n\left(X_{iu} = X_{jv} = 0\right) \\ &= 1 + (-\mathrm{e}^{-\frac{D_u k}{L_n}} - \mathrm{e}^{-\frac{D_v k}{L_n}} + \mathrm{e}^{-\frac{D_u k}{L_n} - \frac{D_v k}{L_n}})(1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)})) \\ &= (1 - \mathrm{e}^{-D_u k / L_n})(1 - \mathrm{e}^{-D_v k / L_n})(1 + o_{\mathbb{P}}(1)), \end{aligned} \tag{60}$$

because $D_u k = \Theta(n)$ and $D_v k = \Theta(n)$. Therefore

$$\begin{aligned} V^{(4)}(k) &= \frac{1}{N_k^2 k^2} \sum_{i,j \in \mathcal{S}_k} \sum_{u,v \in W_n^k(\varepsilon)} D_u D_v (1 - \mathrm{e}^{-D_u k / L_n})(1 - \mathrm{e}^{-D_v k / L_n})(1 + o_{\mathbb{P}}(1)) \\ &\quad - D_u D_v (1 - \mathrm{e}^{-D_u k / L_n})(1 - \mathrm{e}^{-D_u k / L_n})(1 + o_{\mathbb{P}}(1)) \\ &= \sum_{u,v \in W_n^k(\varepsilon)} o_{\mathbb{P}}\left(k^{-2} D_u D_v (1 - \mathrm{e}^{-D_u k / L_n})(1 - \mathrm{e}^{-D_v k / L_n})\right) = o_{\mathbb{P}}\left(\mathbb{E}_n\left[a(k, W_n^k(\varepsilon))\right]^2\right), \end{aligned} \tag{61}$$

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 (57) as

$$D_u^2\left(\mathbb{P}_n\left(X_{iu} = 1\right) - \mathbb{P}_n\left(X_{iu} = 1\right)^2\right) \leq D_u^2. \tag{62}$$

Thus, $V^{(2)}$, can be bounded as

$$V^{(2)} \leq \frac{1}{k^2 N_k^2} \sum_{i \in \mathcal{S}_k} \sum_{u: D_u \in W_n^k(\varepsilon)} D_u^2 = \frac{1}{k^2 N_k} \sum_{u: D_u \in W_n^k(\varepsilon)} D_u^2 = O\left(\frac{n^2}{k^4 N_k}\right)\left|W_n^k(\varepsilon)\right|. \tag{63}$$

Since the degrees are i.i.d. samples from (2), $\left|W_n^k(\varepsilon)\right|$ is distributed as a Binomial$(n, C(n/k)^{1-\tau})$ for some constant $C$. Therefore,

$$\left|W_n^k(\varepsilon)\right| = O_{\mathbb{P}}\left(n\,(n/k)^{1-\tau}\right). \tag{64}$$

Then, using that $N_k \geq 1$,

$$V^{(2)} = O_{\mathbb{P}}\left(n^{4-\tau}k^{\tau-5}\right), \tag{65}$$

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

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

$$\frac{a(k, W_n^k(\varepsilon))}{\mathbb{E}_n\left[a(k, W_n^k(\varepsilon))\right]} \xrightarrow{\mathbb{P}} 1. \tag{66}$$

Combining this with Lemmas 3.3 and 3.4 yields

$$\frac{a(k, W_n^k(\varepsilon))}{n^{3-\tau}k^{\tau-3}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau}\int_{\varepsilon}^{1/\varepsilon} x^{1-\tau}(1-\mathrm{e}^{-x})\mathrm{d}x. \tag{67}$$

□

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

In this section, we prove Proposition 3.1 and show that the contribution to $a(k)$ outside of the major contributing regimes as described in (39) is negligible.

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

$$\begin{aligned}\mathbb{E}\left[a(k, \bar{W}_n^k(\varepsilon))\right] &= \mathbb{E}\left[\mathbb{E}_n\left[a(k, \bar{W}_n^k(\varepsilon))\right]\right] \leq \frac{n}{k}\mathbb{E}\left[D\min\left(1, \frac{kD}{L_n}\right)\mathbb{1}_{\left\{D\in\bar{W}_n^k(\varepsilon)\right\}}\right] \\ &= \frac{n}{k}\int_{x\in\bar{W}_n^k(\varepsilon)} x^{1-\tau}\min\left(1, \frac{kx}{\mu n}\right)\mathrm{d}x. \end{aligned} \tag{68}$$

For ease of notation, we assume that $\mu = 1$ in the rest of this section. We have to show that the contribution to (68) from vertices $u$ such that $D_u < \varepsilon n/k$ or $D_u > n/(\varepsilon k)$ is small. First, we study the contribution to (68) for $D_u < \varepsilon n/k$. We can bound this contribution by taking the second term of the minimum, which bounds the contribution as

$$\int_0^{\varepsilon n/k} x^{2-\tau}\mathrm{d}x = \frac{\varepsilon^{3-\tau}}{\tau-3}(k/n)^{\tau-3}. \tag{69}$$

Then, we study the contribution for $D_u > n/(k\varepsilon)$. This contribution can be bounded very similarly by taking 1 for the minimum in (68)

$$\frac{n}{k}\int_{n/(\varepsilon k)}^{\infty} x^{1-\tau}\mathrm{d}x = \frac{\varepsilon^{\tau-2}}{\tau-2}(k/n)^{\tau-3}. \tag{70}$$

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

# 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 Small $k$

First, we show how to prove Theorem 2.2(i). Similar to (25), the degree of a vertex with weight $h$, $D_h$, satisfies $D_h = h(1 + o_{\mathbb{P}}(1))$ when $h \gg 1$ [38], so that $D_h \ll n^{(\tau-2)/(\tau-1)}$ implies $h \ll n^{(\tau-2)/(\tau-1)}$ with high probability. 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. Let $u$ be a randomly chosen vertex of degree $k \ll n^{(\tau-2)/(\tau-1)}$. Then,

$$a(k) \mid (h_i)_{i\in[n]} = \frac{1}{k}\sum_{i\in[n]} D_i \mathbb{P}_n (i \leftrightarrow u) = (1 + o_{\mathbb{P}}(1))\frac{1}{k}\sum_{i\in[n]} h_i \frac{h_i k}{\mu n} = (1 + o_{\mathbb{P}}(1))\sum_{i\in[n]} \frac{h_i^2}{\mu n}, \tag{71}$$

which is equivalent to (36) because the weights are also sampled from (2). This proves Theorem 2.2(i).

For the hyperbolic random graph, we use that the degree of a vertex $u$ is distributed as a Poisson random variable with parameter $t(u)$, so that $D_u = t_u(1 + o_{\mathbb{P}}(1))$ when $t(u) \gg 1$. Since the $t(u)$s are sampled from (19), the largest $t(u)$ is of order $n^{1/(\tau-1)}$ with high probability. Let $u$ again be a randomly chosen vertex of degree $k \ll n^{(\tau-2)/(\tau-1)}$. Then, similarly as for the rank-1 inhomogeneous random graph, by (18) we obtain with $\zeta = \pi/(2\nu)$,

$$a(k) = \frac{1}{k}\sum_{i\in[n]} D_i \mathbb{P}_n (i \leftrightarrow u) = (1 + o_{\mathbb{P}}(1))\frac{1}{k}\sum_{i\in[n]} t(i)\frac{t(i)t(u)}{\zeta n} = (1 + o_{\mathbb{P}}(1))\sum_{i\in[n]} \frac{t(i)^2}{\zeta n}. \tag{72}$$

Combining this with the fact that the $t(i)$ are sampled from the power-law distribution (19), this proves Theorem 2.3(i) (which is Theorem 2.2(i) where $\mu$ is replaced by $\zeta$ and $c/(\tau-1)$ by $(\tau-1)/2$).

## 4.2 Large $k$

Similarly to (39), we define for the rank-1 inhomogeneous random graph

$$W_n^{k,\text{HVM}}(\varepsilon) = \{u : h_u \in [\varepsilon\mu n/k, \mu n/(\varepsilon k)]\}, \tag{73}$$

and for the hyperbolic random graph

$$W_n^{k,\text{HRG}} = \{u : t(u) \in [\varepsilon\zeta n/k, \zeta n/(\varepsilon k)]\} \tag{74}$$

with $t(u)$ as in (17) and $\zeta = \pi/(2\nu)$. Then it is easy to show that Proposition 3.1 also holds for the rank-1 inhomogeneous random graph and the hyperbolic random graph, with (73) or (74) instead of $W_n^k(\varepsilon)$. For the rank-1 inhomogeneous random graph, we use that $\mathbb{P}_n (X_{ij} = 1) = \min(h_i h_j/(\mu n), 1)$. Because the weights are sampled from (2), (68) also holds for the rank-1 inhomogeneous random graph, so that Proposition 3.1 indeed holds for the rank-1 inhomogeneous random graph. In the hyperbolic random graph, the $t(u)$ variables are sampled form a distribution similar to (2) (apart from constants), and by (18) the connection probabilities are $\min(t(u)t(v)/(\zeta n), 1)$. Then, as for the rank-1 inhomogeneous random graph (68) also holds for the hyperbolic random graph, apart from a multiplicative constant, and from there we can follow the rest of the proof of Proposition 3.1.

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

$$\frac{a(k, W_n^{k\text{HVM}}(\varepsilon))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau}\int_{\varepsilon}^{1/\varepsilon} x^{1-\tau}\min(x, 1)\mathrm{d}x. \tag{75}$$

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 $1 - \mathrm{e}^{-D_i D_j/L_n}$ by $\min(h_i h_j/\mu n, 1)$. Note that for the rank-1 inhomogeneous random graph the contribution to (57) from 4 different vertices is 0, because the edge probabilities in the rank-1 inhomogeneous random graph conditioned on the weights are independent. Then (75) follows from these lemmas. This then shows similarly to (44) that

$$\frac{a(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_0^\infty x^{1-\tau} \min(x,1)\mathrm{d}x = \frac{c\mu^{2-\tau}}{(3-\tau)(\tau-2)}. \tag{76}$$

which proves Theorem 2.2(ii).

For the hyperbolic random graph, similar arguments hold. Here, we can prove an analogous proposition to Proposition 3.2 which states that

$$\frac{a(k, W_n^{k,\text{HRG}}(\varepsilon))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} \frac{(\tau-1)^2}{2}\left(\frac{\pi}{2\nu}\right)^{2-\tau} \int_\varepsilon^{1/\varepsilon} x^{1-\tau} \min(x,1)\mathrm{d}x. \tag{77}$$

As stated before, connection probabilities in the hyperbolic random graph between two vertices with uniform radial coordinate are similar as in the rank-1 inhomogeneous random graph, with weights $t(u)$. Therefore, Lemmas 3.3-3.5 also hold for the hyperbolic random graph, replacing the connection probability $1 - \mathrm{e}^{-D_iD_j/(\mu n)}$ of the erased configuration model by $\min(t(i)t(j)/(\zeta n), 1)$ and replacing the constant $c$ from (2) by its equivalent constant for the hyperbolic model of $(\tau-1)^2/2$ (see (19)). Again, similar steps that lead to (44) then show that

$$\frac{a(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} \frac{(\tau-1)^2}{2}\left(\frac{\pi}{2\nu}\right)^{2-\tau} \int_0^\infty x^{1-\tau} \min(x,1)\mathrm{d}x = \frac{(\tau-1)^2}{2(3-\tau)(\tau-2)}\left(\frac{\pi}{2\nu}\right)^{2-\tau}, \tag{78}$$

which proves Theorem 2.3(ii).

**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 Leeuwaarden for their useful comments.

# References

[1] R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.

[2] A.-L. Barabási. *Network Science*. Cambridge University Press, 2016.

[3] A. Barrat, M. Barthlemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.*, 101(11):3747–3752, 2004.

[4] S. Bhamidi, S. Dhara, R. van der Hofstad, and S. Sen. Universality for critical heavy-tailed network models: metric structure of maximal components. arXiv:1703.07145, 2017.

[5] S. Bhamidi, S. Sen, and X. Wang. Continuum limit of critical inhomogeneous random graphs. *Probab. Theory Related Fields*, pages 1–77, 2016.

[6] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. Universality for first passage percolation on sparse random graphs. *Ann. Probab.*, 45(4):2568–2630, 2017.

[7] M. Bode, N. Fountoulakis, and T. Mller. On the largest component of a hyperbolic model of complex networks. *Electron. J. Combin.*, 22(3):P3–24, 2015.

[8] M. Bode, N. Fountoulakis, and T. Mller. The probability of connectivity in a hyperbolic model of complex networks. *Random Structures & Algorithms*, 49(1):65–94, 2016.

[9] M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Phys. Rev. E*, 68:036112, 2003.

[10] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.*, 90:028701, 2003.

[11] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.

[12] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, 124(6):1377–1397, 2006.

[13] E. Candellero and N. Fountoulakis. *Clustering and the Hyperbolic Geometry of Complex Networks*, pages 1–12. Springer International Publishing, 2014.

[14] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, 71:027103, 2005.

[15] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, **99**(25):15879–15882 (electronic), 2002.

[16] P. Colomer-de Simon and M. Boguñá. Clustering of random scale-free networks. *Phys. Rev. E*, 86:026120, 2012.

[17] S. N. Dorogovtsev. Clustering of correlated networks. *Phys. Rev. E*, 69:027104, Feb 2004.

[18] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*, volume 29, pages 251–262. ACM, 1999.

[19] T. Friedrich and A. Krohmer. Cliques in hyperbolic random graphs. In *INFOCOM proceedings 2015*, pages 1544–1552. IEEE, 2015.

[20] L. Gugelmann, K. Panagiotou, and U. Peter. Random hyperbolic graphs: degree sequence and clustering. In *ICALP proceedings 2012, Part II*, pages 573–585. Springer, Berlin, Heidelberg, 2012.

[21] R. van der Hofstad. *Random Graphs and Complex Networks Vol. 1.* Cambridge University Press, 2017.

[22] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures & Algorithms*, 27(1):76–123, 2005.

[23] R. van der Hofstad, A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and C. Stegehuis. Local clustering in scale-free networks with hidden variables. *Phys. Rev. E*, 95:022307, 2017.

[24] R. van der Hofstad, J. S. H. van Leeuwaarden, and C. Stegehuis. Optimal subgraph structures in scale-free networks. -, 2017.

[25] P. van der Hoorn and N. Litvak. *Upper Bounds for Number of Removed Edges in the Erased Configuration Model*, pages 54–65. Springer International Publishing, Cham, 2015.

[26] S. Janson and M. J. Luczak. A simple solution to the k-core problem. *Random Structures & Algorithms*, 30(1-2):50–62, 2007.

[27] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[28] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná. Hyperbolic geometry of complex networks. *Phys. Rev. E*, 82(3):036106, 2010.

[29] J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, 2014. Date of access: 14/03/2017.

[30] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: correlation profile of the internet. *Phys. A*, 333:529 – 540, 2004.

[31] M. Mayo, A. Abdelzaher, and P. Ghosh. Long-range degree correlations in complex networks. *Computational Social Networks*, 2(1):4, 2015.

[32] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89(20):208701, 2002.

[33] M. Ostilli. Fluctuation analysis in complex networks modeled by hidden-variable models: Necessity of a large cutoff in hidden-variable models. *Phys. Rev. E*, 89:022807, 2014.

[34] J. Park and M. E. J. Newman. Origin of degree correlations in the internet and other networks. *Phys. Rev. E*, 68(2):026112, 2003.

[35] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87:258701, 2001.

[36] E. Ravasz and A.-L. Barabási. Hierarchical organization in complex networks. *Phys. Rev. E*, 67:026112, 2003.

[37] M. A. Serrano and M. Boguñá. Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.*, 97(8):088701, 2006.

[38] C. Stegehuis, R. van der Hofstad, J. S. H. van Leeuwaarden, and A. J. E. M. Janssen. Clustering spectrum of hierarchical scale-free networks. arXiv:1706.01727, 2017.

[39] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs. *J. Stat. Phys.*, 133(1):169–202, 2008.

[40] A. Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E*, 67:056104, 2003.

[41] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the internet. *Phys. Rev. E*, 65(6):066130, 2002.

[42] W. Whitt. *Stochastic-Process Limits*. Springer, New York, 2006.

[43] D. Yao, P. van der Hoorn, and N. Litvak. Average nearest neighbor degrees in scale-free networks. arXiv:1704.05707v2, 2017.